ATHABASCA UNIVERSITY

ARTIFICIAL INTELLIGENCE MAY SOON EMULATE PICASSO AND TOLKIEN—IS PORTER NEXT? AN OPTIMIZATION OF ORGANIZATION STRATEGIC DECISION-MAKING WITH AI TECHNIQUES

BY

SUNNY BHASIN

A DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF BUSINESS ADMINISTRATION

FACULTY OF BUSINESS

ATHABASCA, ALBERTA, CANADA

APRIL, 2025

Approval

Athabasca University
Faculty of Graduate Studies

Approval of Dissertation

The undersigned certify that they have read the dissertation entitled

ARTIFICIAL INTELLIGENCE MAY SOON EMULATE PICASSO AND TOLKIEN—IS PORTER NEXT? AN OPTIMIZATION OF ORGANIZATION STRATEGIC DECISION-MAKING WITH AI TECHNIQUES

Submitted by:

Sunny Bhasin

In partial fulfillment of the requirements for the degree of

Doctor of Business Administration

The examination committee certifies that the dissertation
and the oral examination is approved

Supervisors:
Dr. Oli Mihalache
Athabasca University

Dr. Stella George
Athabasca University

Committee Members:
Dr. Mihail Cocosila
Athabasca University

External Examiner:
Dr. Oleksiy Osiyevskyy
University of Calgary

Past Committee Members:
Dr. Mihnea Moldoveanu, University of Toronto
Dr. Joe Cox, Athabasca University

April 22, 2025

1 University Drive, Athabasca, AB, T9S 3A3 Canada
Toll-free (CAN/U.S.) 1.800.788.9041 (6821)
fgs@athabascau.ca | fgs.athabascau.ca | athabascau.ca

Dedication

For my son, in whom I see virtually all of my traits and flaws—and countless opportunities to help him maximize the former while mitigating the latter. Papa is sorry for all the time away. This is all for you.

For my wife, this would not have been possible without your support and acceptance. You once told me that you always wanted to marry a doctor—you probably didn't specify what kind, so this will have to do for now.

For my parents, I hope you're proud of me today.

Acknowledgement

I would like to thank all who have directly or indirectly left indelible marks on me through this journey of learning called life. Thank you to all my classmates and friends in DBA, MBA, Engineering, and all others with whom I've crossed paths for a rhyme, a time, a reason or a season. A special shout out to friends who altered my course for the better in so many ways—you know who you are.

Gane Olsen, my biology teacher in grade 10. With his unique teaching and evaluating style, Mr. Olsen helped me truly realize and believe in my academic potential for the first time. Teachers change the world every day through the gifts they endow on their students. Mr. Olsen, I thank you.

Roger Martin, professor and former Dean of the Rotman School of Management, who taught me about strategy and strategic decision-making in new and innovative ways, and helped lay the roots into an area of interest and inquiry that have stayed with me ever since. Roger, I thank you.

My current and past doctoral supervisors: Oli, Stella, Mihail, Mihnea and Joe. This would not be possible without your guidance, support, and encouragement. I thank you.

I would like to thank the energy of the Universe, which to me is God. We take from it, contribute to it, live and breathe within it, are infinitely small yet immensely powerful specks of it. While it is formless, it is everywhere we can and cannot see. It is everything. I live to return to your divine infinity at the end of this life.

Abstract

This dissertation explores improving strategic decision-making, an intractable and often abstract type of optimization problem, via the application of artificial intelligence (AI) techniques. The latest AI techniques based on deep learning (DL) and artificial neural networks (ANNs) have powerful capabilities including generating stories *ad Tolkien*, art *ad Picasso*, and hold great promise against more complicated problem classes. However, it is still unclear as to how and how well, these techniques can help improve strategic decision-making. I propose several novel management models and approaches to deploy AI techniques as tools against the problem of organization strategy, account-for and mitigate trade-offs, and achieve more optimal outcomes. A combination of inductive, deductive and abductive reasoning approaches were employed to advance two sequential studies, with the results from the first informing the second.

Study one is a review of nearly 500 pieces of peer-reviewed literature, books, databases, empirical case studies and other sources from 1945 to 2025. A set of seven largely uncaptured trade-off dimensions exist, with relevance in both an organization strategy and computational context: accuracy, explainability, fairness, privacy, reliability, security, and speed; and, twenty-one pairwise trade-offs between them. These trade-offs persist despite the advent of several AI techniques examined herein, including Large Language Models (LLMs), Generative Adversarial Networks (GANs), Retrieval-Augmented Generation (RAG), Mixture of Experts (MoE) and GFlowNetworks (GFNs). AI techniques such as Language Models (LMs) and their requisite training also have not addressed important trade-offs such as accuracy vs. speed. Contributions to the literature and theory, include: 1) a comprehensive examination of AI techniques within broad organization strategy contexts spanning industries and functional areas, 2) explication and

extensive exploration of trade-off dimensions and pairwise trade-offs between them that have relevance within both organization strategy and computational optimizations, 3) finding that these trade-off dimensions persist despite advancements in and performance of AI techniques, and 4) finding that different AI techniques are better suited for different problems/goals. Contributions to the practice, include: 1) a categorization of trade-offs by industry and functional area, a common way that strategists and consultants delineate their expertise and problem domains, and 2) a multi-level mapping that organizations can use to choose AI technique(s) for their specific problem/activity.

Study two takes the above, and developed and introduced several management models and approaches for improving strategic decision-making and outcomes through the deployment of AI techniques in these organization processes. Simulations were used to assess AI technique performance against different trade-offs, yielding the following results: RAG was the most effective technique, while LLMs had the widest applicability. Combinations of AI techniques, such as LLM+RAG, outperform individual techniques with respect to both colloquial attributes, e.g. parameters and cost, and the set of seven trade-off dimensions. MoE-based algorithmic architectures exhibit gains in the broadest set of desired outcomes while mitigating unfavourable trade-off dimensions. Contributions to the literature and theory, include: 1) a unifying multi-step approach to maximize fitness of a production function using the NK Model and strategic fit *with* computational resources, and 2) proposition that *AI Technique* is the most accurate way to refer to anything AI. Contributions to the practice, include: 1) a conversion from colloquial attributes such as number of parameters, to the set of seven trade-offs dimensions, and 2) a hybrid process of humans and AI techniques, that organizations can utilize to maximize outcomes while minimizing unintended trade-offs, modelled to improve the accuracy vs. speed trade-off.

This research study has the potential to enhance organization strategic decision-making and outcomes, via computational approaches such as AI techniques. Improving management models and processes to address issues that have thus far contributed to suboptimal outcomes, may lead to increased economic value to organizations, their stakeholders and customers, the economy, and society at large.

Keywords: organization strategy, artificial intelligence, trade-offs

Preface

I have for many years been intrigued by and motivated to study strategy formulation, problem solving, and strategic decision-making. This led to my pursuit of a Master of Business Administration degree, and now a Doctorate in Business Administration. My computer engineering background animates and perhaps equips me to explore the intersection and ever-increasing convergence of computational systems and the abstract realm of organization strategy and other intractable problems within organizations. While developments in AI techniques and their operationalization and application to organization problems has been around for decades, this pervasiveness of this convergence along with accessibility, investment and interest, has perhaps never been greater.

I believe this research study is very timely, and presents opportunities to publish and contribute to academic research and practitioner applications. Portions of my literature review and ensuing research study were accepted at three international conferences between 2020 and 2022. I have also shared and presented portions to other practitioner audiences, including professional workplaces.

I hold degrees in Biological Sciences, Computer Engineering and Master of Business Administration. Prior to pursuing doctoral research, I spent the beginning of my career within industry as a software and systems engineer with International Business Machines (www.ibm.com), then with a strategy consultancy as a management consultant, followed by a role as a technology deal lead within private equity, then as a strategy development leader within Canadian telecommunications companies. I have also co-founded several software and hardware companies, and am inventor or co-inventor on over 20 patents globally.

Contents

List of Tables

List of Figures

List of Abbreviations

AHP	Analytic Hierarchy Process
AI	Artificial Intelligence
ANN	Artificial Neural Network
AGI	Artificial General Intelligence
ASI	Artificial Super Intelligence
CPU	Central Processing Unit
DAG	Directed Acyclic Graph
DL	Deep Learning
DNN	Deep Neural Network
DNA	Deoxyribonucleic Acid
DDSS	Distributed Decision Support System
DI	Distributed Intelligence
DS	Distributed System
DSDSS	Distributed Strategic Decision Support System
DSS	Decision Support System
ES	Expert System
GAI	Generative Artificial Intelligence
GAN	Generative Adversarial Network
GFN	Generative Flow Network
GPT	Generative Pre-trained Transformer
GPU	Graphics Processing Unit
LM	Language Model (of varying size, e.g. large or small)

LLM	Large Language Model
MC	Monte Carlo or Monte Carlo Simulation
MCMC	Markov Chain Monte Carlo or Markov Chain Monte Carlo Simulation
ML	Machine Learning
MLA	Machine Learning Algorithm
MoE	Mixture of Experts
NLP	Natural Language Processing
NLU	Natural Language Understanding
P	Polynomial Time
PEER	Parameter Efficient Expert Retrieval
NP	Non-Deterministic Polynomial Time
RAG	Retrieval-Augmented Generation
RAI	Responsible AI
RL	Reinforcement Learning
SVC	Support-Vector Clustering
SLM	Small Language Model
SVM	Support-Vector Machine
XAI	Explainable AI

Chapter 1. Introduction

Organization strategy formulation and strategic problem solving are complicated processes, and improvements to these have potentially wide and deep implications from economic and societal perspectives. Improvements can occur in many ways, and using AI techniques for this is one novel approach. AI techniques are interesting tools because strategic decisions are types of optimization problems, e.g. maximization:minimization, which computational approaches are adept at solving. This study explores the application of AI techniques to strategic decision-making, and contemplates driving these technologies and systems into what has until recently, largely been the domain of humans, and our ability to make decisions, easily perform abstractions, recognize unconditioned patterns, create language, patterns, images, and so on. In a way, the challenge may be seen as how one solves strategic problems, rather than a particular problem. While AI techniques based on deep learning and artificial neural networks have now achieved natural language processing and understanding to a level where they may soon be tasked to enter the arena of organization *strategy processes*, it is unclear how well current AI techniques can improve strategic decision-making (Chui at el., 2018; Moldoveanu, 2016, 2019a; Ng, 2016; Reeves & Moldoveanu, 2017). Several lines of inquiry emerge: are any gains that have been seen via different systems and algorithmic approaches, achieved at only computational, storage or network costs? Can a set of approaches or models be created that help organizations maximize their intended outcomes through more comprehensive capture or mitigation of such trade-offs? This study will explore these concepts, and whether/how more robust and comprehensive management models can be developed to address these issues for more optimal solutioning.

While the conversation seems to have leapfrogged automation and its impact on workforces (McKinsey Global Institute, 2017), and relatively rudimentary tasks like improved predicters or recommenders, broad viability against intractable problems remains an under-achieved aspiration. As AI techniques are deployed against ever-widening problem types and contexts, the potential positive implications of this research study are significant for organizations across industries and geographies.

Target audiences for this study include researchers, practitioners, organizations and teams, public and private policy makers, legislators, strategists, and even the general public—virtually anyone working in organization strategy formulation, strategic problem solving and decision-making, or economics—and considering AI techniques as tools to improve.

Background and Context

Strategy Problems are Intractable Optimization Problems

Why and how are some organizations able to outperform their competitors within the same competitive environment or landscape? Porter gives us his foundational contribution of intra-organization fit as the key ingredient and driver for superior competitive performance (1996), which stems from a system of activities, eventually leading to relative competitive out-performance and differentiation. There are inherent differences in organizations' actions, sequences of actions and intra-organization structures. Organizations that are able to operate via superior fit are able to create and enhance strategies and inimitable competitive differentiation. This fit from top-to-bottom, bottom-to-top and across the organization seems to be something more than operational choices, and is more internal to the organization. I suggest that organizations with superior fit and performance seem to be almost wired differently—consider an analogy of an organization *super* aligned so as to resemble a biological organism operating as a unified whole.

Kaufmann's work on biological adaptation on rugged landscapes (with Levin, 1987; 1988) and their NK Model, portrays organisms exhibiting natural biological adaptation on competitive landscapes of varying ruggedness. On such landscapes, organisms undertake an adaptive walk and undergo genetic adaptation, as they adapt to their competitive landscapes and/or environment. This adaptive walk towards a more fit genetic makeup, i.e. higher peaks on the landscape, is a combinatorial optimization process, where N represents the length of a genetic peptide chain, and K reflects the number of possible interactions between these peptides based on the level of landscape ruggedness. A *successful* adaptive walk is predicated on and ultimately leads to increasingly more fit combinations of genetic blocks in the chain. I deduce that organisms that do not adapt vis-à-vis the landscape and other competing organisms in the area, fail to achieve more fit genetic combinations and therefore, fail to reach higher peaks on the landscape.

The NK Model is a vector of organism traits (Kauffman, 1988), and I posit that it serves as a helpful area of study for other optimization problems, including those in organization and management study. It serves as a possible unifying explanation and conceptual platform, upon which organization fit can be understood through an adaptation lens, and perhaps vice versa. The notion of biological adaptation as a complex optimization problem and process, bridges with Porter's fit (1996), which is an holistic intra-organization alignment across activity sets towards a more optimal configuration to out-perform competitors.

In mathematical or computational terms, the complexity of strategy problems renders them undividable to solve as a sequence of simpler parts, or sub-problems with sub-solutions (Levinthal, 2011; Moldoveanu & Bauer, 2004; Moldoveanu, 2009, 2016; Rivkin, 2000; Weinberger, 1996). While approximations can be made with varying degrees of accuracy using various techniques, strategy problems cannot be *deterministically* solved in parts. Similarly, Porter (1996) argues that a dominant or winning strategy is a system of activates and *not*, in fact, a collection of parts. If we accept this, is there a path forward for computational resources, such as advanced AI techniques, deployed against strategy problems? And if so, can they stand alone, or would some human involvement be necessary or beneficial?

Organization strategy is a unified articulation of what the organization seeks to achieve, to create or enhance a long-term sustainable and inimitable competitive advantage through differentiation (Porter, 1996), with optionality to address how it will achieve same and where it will operate (Lafley & Martin, 2013; Porter, 1996). It is more an aspirational intent, and certainly not organization effectiveness (Lafley & Martin, 2013; Porter, 1996) or a plan (Martin, 2021). And, I consider *strategic decision-making* as the act and associated processes of making decisions in furtherance of achieving the organization strategy.

Addressing the Complexity of Strategy with Computation

In the same seminal article where they introduce the NK Model, Kauffman and Levin (1987) also make a novel connection between biological evolutionary processes with algorithms used for optimization. They make the finding that adaptive walks, and other optimization problems, have solution spaces that grow faster than polynomially with increasing input variables, i.e. the time to solve grows at least exponentially as a function of input variables.

Using the NK Model in the context of organization and management science, Kauffman's competitive landscape exhibits characteristics of a correlated rugged landscape, with cross and longitudinal interconnectedness between organizations, and between organizations and the landscape. This is of interest because the biological mechanisms observed and theorized by Kauffman and Levin via the NK Model, hold applicability within organization strategic decision-making—sets of decisions as vectors of more or less intra-organization fit against a competitive landscape.

Taking concepts from the study of complex systems, as a bridge between Porter's fit (Porter, 1996) and the NK Model (Kauffman & Levin, 1987), McKelvey (1999) supports the notion that complexity is both a cause and a consequence. In his inquiry of whether complexity theory is useful for explaining competitive selection within industry populations, McKelvey posits that multicoevolutionary complexity at the firm level internalizes natural selection processes, thereby turning interdependencies between intra-firm and inter-firm parts or activities, into segments of Porter's value chain (Porter, 1985, as cited in McKelvey, 1999). This holistic view and its connection to Porter's empirical work helps understand distinctive aspects of organization strategy and why its complexity may be particularly well-suited to more optimally address via computational algorithms.

Organization strategic problems are defined as the general desire to go from a current situation to some other state, and are often abstract problems that lack clear and precise articulation (Moldoveanu & Leclerc, 2015). Moldoveanu and Leclerc assert that business problems are at best, imprecise descriptions of situations or predicaments. Hence, the act of trying to solve an imprecisely worded or suboptimally articulated problem seems futile, without some sort of translation language or lexicon. Different from frameworks and other seemingly similar strategists' tools which impose rigidity and fit upon the problem, Moldoveanu and Leclerc (2015) propose types of flexible problem articulation languages, to allow for the improved articulation and capturing of the true problem-to-solve, without simplifying, altering or otherwise moving away from the original problem(s).

Organization problems are at their core, optimization problems (Moldoveanu & Leclerc, 2015). Therefore, I suggest that deploying probabilistic-based inferential AI technologies against such problems for improved outcomes, is an optimization problem. Optimization is something computational resources and systems are particularly adept at solving, provided we are starting from a defined, well-articulated, specific and solvable problem. This is critical to finding an efficient fitness-fit solution. Strategists seek to achieve Porter's intra-organizational fit and inter-organizational differentiation via myriad of thinking, frameworks, approaches, and executable actions (inferred and expanded from M. Moldoveanu, personal communication, February 4, 2022). Computational resources can and do play a facilitating role to varying degrees.

What is Artificial Intelligence?

> Defining AI is not easy; in fact, there is no generally accepted definition of the concept (Sheikh et al., 2023, p. 15).

What exactly is *artificial intelligence* and how do we bring it *out of* a discussion in computer science into the realm of organization and management science? Sheikh et al. (2023) comment that AI is very difficult to clearly define because even human intelligence, which AI is *supposed to* imitate, simulate or emulate, is not fully understood. And moreover, a lack of clear interface between intelligence research in the human and artificial domains means understanding of both is independently and disjointedly evolving. That said, a definition is important to this study. I adopt a widely contributed-to definition of an *AI system* from the Organization for Economic Cooperation and Development (OECD):

> An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment. (OECD, 2024).

AI systems are being applied to more and more classes and types of problems within all areas including healthcare (Adadi & Berrada, 2018; Vollmer et al., 2020), research and academia (Arrieta et al., 2019; Gerke et al., 2020), public policy (Abowd & Schmutte, 2018; Foulds et al., 2019), environment (Roscher et al., 2020), and business (Bhatt et al., 2020; Pessach & Shmueli, 2022).

The conceptual beginnings of machine learning (ML) and AI date back over 70 years (McCarthy, 1956; Smith et al., 2006; Turing, 1950), or perhaps even earlier (Sheikh et al., 2023), and has evolved from simpler rule-based deterministic inferential systems such as fuzzy sets, decision support, and expert systems (Li, 2000; Liao, 2005a, 2005b; Zadeh, 1965, 1975, 1976). However, its more recent high-accessibility has been spurred on largely by the rapid advancements in computational processing power, virtually limitless storage for virtually zero incremental cost (Agrawal, 2018), and I suggest, distributed or cloud-based infrastructure and applications, and step-function increases in network speeds. Its application in business problems such as analytics, predictions or recommendations, has been ever-expanding and increasingly economical (Agrawal, 2018; Brynjolfsson & McAfee, 2014; Lauterback & Bonime-Blanc, 2016).

Gaps and Ongoing Debates

I've introduced above, the inherent complexity of strategy problems, and human judgment is a critical part of organization strategic processes (Agrawal et al., 2018a, 2018b). Human judgment and the various considerations and risk metrics we are able to evaluate and balance, poses algorithmic challenges for the way computers and AI techniques operate. AI techniques as they currently exist, and as powerful as they are, have limitations. As seen in Adadi and Berrada (2018), Agrawal et al. (2018a, 2018b), Arrieta et al. (2019), Bettis and Hu (2017), Mittelstadt et al. (2018), Siau and Wang (2018) and others, one of the most significant such limits is being able to emulate judgment or prediction with a high enough degree of accuracy and efficacy such that humans willingly defer to and trust these technologies without intervening or overriding it, even in the face of a counterintuitive computationally-derived direction. This suggests a reasonable likelihood of, at most, an incomplete *hand-off* to computational systems and/or AI techniques.

Other issues arising from prospective gains in dimensions such as accuracy, include the loss of explainability, privacy and security. These issues present challenges in not only computationally finding a, or *the*, correct answer, but trusting it enough to allow for more limited *or no* human intervention, the latter being an unavoidable characteristic of current deep computing systems, where explainability is lost (Adadi & Berrada, 2018; Amann et al., 2020; Arrieta et al., 2019; Gunning & Aha, 2019); London, 2019; Piano, 2020).

Another more practical issue is how to actually get AI techniques to compute answers to strategic problems in a reasonable timeframe—in this context, compute would be the total time including, e.g. training of an AI model. The total compute time starts to rise exponentially as the amount of information grows, and high-fidelity strategic decision-making usually requires a lot of dynamic information. This also directionally hints at the need for an iterative process and collaboration between humans and computational systems. Humans are particularly adept at flexible reasoning, but machines are less so—even the most advanced deep learning techniques are, by their very nature, less elastic in their computational logic than humans.

The literature also illuminates gaps that organizations and managers grapple with and make compromises on while selecting, implementing and even evaluating results. Moldoveanu (2020, 2021) has posited that the observed suboptimal performance and results from such touted initiatives are largely due to organizations being unable or unwilling to take into account other important considerations in their quest for gains in dimensions like accuracy or speed. Building upon some of his research (2016, 2019a, 2019b, 2019c, 2019d, 2020, 2021), this study seeks to develop and propose management models that organizations can incorporate into their strategic decision-making and problem-solving processes to maximize their outcomes, through more comprehensive inclusion and mitigation of unintended trade-offs.

Finally, the literature is unclear as to whether, how, and if so, which AI techniques have the highest utility and efficacy in helping organizations improve their strategic decision-making and strategy formulation processes (Ng, 2016). That's not to say that gains have not been had against certain classes of problems, for example against classification or recommendation, or deep learning against pattern identification and generation, and so on. However, these are less complicated classes of problems than those presented in the context of organization strategy. Moreover, the evidence suggests that any prospective outcome gains via the deployment of these technologies comes with unaccounted-for trade-offs. These are in addition to known and quantifiable, satisfiable and addressable inputs such a capital/financial or computational cost. The concept of a set of multi-dimension trade-offs was first conceived by one of my former doctoral co-supervisors, Mihnea Moldoveanu[1]. The concept of *no free lunch* (Abowd & Schmutte, 2018; Dunjko & Briegel, 2017; Holzinger, 2018; Wolpert & Macready, 1997; Papernot et al., 2016; Xu & Liu, 2021; Xu et al., 2021) becomes important in an analysis of AI techniques deployed against complex problems. The literature, including those cited above, provides support for the notion that there are trade-offs that need to be better understood and addressed, and that appear to be not completely captured in such organization processes, ultimately contributing to suboptimal outcomes (see Tables 7 to 9 below for more examples).

Research Questions

A review of the literature indicates a gap in comprehensive coverage and understanding of trade-offs in general, and specifically, in the application of AI techniques in the organization strategy domain. To better understand these trade-offs, how they impact organization outcomes, and how the deployment of computational resources can help organization maximize their outcomes, I propose the following research questions:

[1] Through the course of several conversations and correspondences on my dissertation research area, he shared with me his research and existing empirical studies that demonstrate that a set of trade-offs dimensions factor into AI technology deployments, and that these appear to be often uncaptured. In the conception phase, my specific contribution to Moldoveanu's research on hypothesized trade-offs was to help add clarity around some of the proposed trade-off dimensions while working with Moldoveanu on ensuring the proposed trade-offs were valid in the context of AI technologies applied to organization strategic decision-making. More specifically, that any set of proposed dimensions was mutually exclusive and collectively exhaustive, that they were discrete and scalars, and that potential 1-to-1 and 1-to-many relationships were indeed applicable to, and could be analyzed to answer my specific research questions.

RQ1: How can organizations use AI techniques to improve strategic decision-making?

RQ2: What trade-offs dimensions are relevant in the application of AI techniques in organization strategy processes?

RQ3: How can organizations improve strategic decision-making processes to maximize outcomes while minimizing the corresponding trade-offs?

RQ1 frames and bounds this research study, and serves as my primary research question, while RQ2 and RQ3 are secondary questions. These questions are addressed via two sequential studies, where the results of the first inform the second. Collectively, these three research questions provided guidance in the exploration, funneling, and synthesis of concepts from two very large domains, and allow for specific and useable outputs.

Significance of this Research

This research study has the potential to enhance complex, traditionally human-led, strategic decision-making processes via computational tools such as AI techniques. More holistic frameworks allow for bridging together of concepts from the domains of management and computer sciences. This study seeks to examine computational and process approaches to address issues that have thus far contributed to suboptimal outcomes when computational tools have been applied in these organization strategy contexts.

Why is a study of AI techniques important within the context of organization strategy and strategic decision-making? There is much research and commentary on the importance of these technologies in this context, and perhaps more importantly, why broadly speaking the results achieved thus far have been less than optimal (Acemoglu & Restrepo, 2018; Brynjolfsson & McAfee, 2014, 2017; Chui at el., 2018; Domingos, 2015; Moldoveanu, 2007, 2016, 2019a; Reeves & Moldoveanu, 2017; Ng, 2016). There is obvious economic value to organizations, their stakeholders and customers, and the economy at large, when organization actions can lead to more optimal performance and outcomes. Beyond this type of efficiency optimization, more robust selection, deployment, and utilization of AI techniques—along with optimized outcomes—presents nascent opportunities to improve strategic positioning. This AI fitness landscape may be the competitive landscape upon and against which winning organizations will create, develop and enhance their competitive advantages.

The potential for AI techniques to move beyond predictions, recommendations, and even more advanced generative capabilities, and play a role in more complex organization functions is a possibility that should be considered. It is critically important considering the observed results thus far, for organizations to fully understand and incorporate a more comprehensive analysis set of the potential gains, losses and trade-offs.

While it may seem that an inquiry about AI and its application to business and organization most impacts and aids the formulators of these strategies—which it no doubt does—it also has the potential to significantly alter how users and consumers of products and services react and interact with businesses and organizations. These consumers include customers in the business context, organization stakeholders, and society at large in the public policy context. Domingos (2015) asserts that studying machine learning and AI more broadly, should be of interest and intrigue to virtually everyone because you cannot escape their reach and impact, nor should you want to. Consider systems leveraging AI techniques to autonomously make strategic decisions such as resource/personnel allocation or management, geographic/market expansion or contraction, or product development/pricing. These are just some applications of AI technique deployment which already exist or are being contemplated. And with the advent and ubiquitous proliferation and access to generative AI techniques by all manner of organizations and their customers, this type of research will be critical to not only contribute to the value-capture from the utilization of these techniques, but help avoid disadvantageous trade-off dimensions that still persist in these new and emerging contexts.

Study Scope and Limitations

Research and development in the AI domain and underlying technologies, are advancing and expanding at an immense pace, and are one of the largest investment areas globally (Williams, 2021). Within this vast space, I limited my analysis to a small set of AI techniques: Large Language Models (LLMs), Generative Adversarial Networks (GANs), Retrieval-Augmented Generation (RAG), Mixture of Experts (MoE), and GFlowNetworks (GFNs). This set of AI techniques is not exhaustive, and was selected because of the current relative commonness of the deployment of these techniques in industry, along with some baseline understanding within *non-technical* strategist practitioner groups of how these techniques operate. Other computational approaches such as genetic algorithms are discussed herein in the limited context of a maximization (see Ch. 4), but techniques such as linear programming were deselected because they are not generally considered to be self-refining ML-based approaches. Lastly, findings I make with respect to the performance of different AI techniques against trade-off dimensions are based on simulations and not real-world implementations. As such, actual deployments may yield different results.

My research questions will prompt exploration into trade-off dimensions and the pairwise trade-offs between these dimensions, thereby expanding the research into this accelerating, yet under-theorized area. Trade-off dimension-interaction is a relatively new area without the existence of comprehensive study in the available literature. I will limit my analysis to 1-to-1 pairs of trade-off dimensions, and leave more complex multi-/n-wise manifold interactions to subsequent studies.

This theoretical study relies on existing primary, secondary and tertiary literature[2] and sources. As such, there is an inherent risk and limitation that important information or knowledge may not be captured because it is not publicly accessible by the author or is otherwise not incorporated. It is assumed that any primary research conducted and utilized for this study has undergone an ethical review and approval process, especially in case studies which involved individual and group participants. This study was undertaken with an objectivist ontology and positivist epistemology.

Brynjolfsson and McAfee (2014) and Domingos (2015) and others have articulated that there is incalculable progress and benefits to be had from AI. Within the literature I reviewed specific to my study, there are threads highlighting risks and issues involved in AI adoption, such as the nature of work and impact on employment, governance, income distribution, GDP and so

---

[2] https://library.carleton.ca/guides/help/primary-secondary-and-tertiary-sources, https://en.wikipedia.org/wiki/Tertiary_source, https://guides.library.cornell.edu/sources/primary, https://guides.library.cornell.edu/sources/secondary, https://guides.library.cornell.edu/sources/tertiary; all retrieved Mar 17, 2025.

on (Acemoglu, D., & Restrepo, P., 2018; Brynjolfsson & McAfee, 2014; Canhoto & Clear, 2019; Lauterbach & Bonime-Blanc, 2016; Magoulas & Swoyer, 2020; Meek et al., 2016; Ojanpera et al., 2018). However, the case has been made that the gains outweigh the risks and that mitigating steps can be taken through public, organizational and management policies, and individual decision making within society (Brynjolfsson & McAfee, 2014; Ojanpera et al., 2018). The intended outputs of this study may help address some of these risks and issues.

Outline of Dissertation Chapters

This dissertation consists of seven chapters (visualized in Figure 1), the first being this introduction. The topics and concepts I cover in this study are wide-ranging and not simple, so the introduction was used to establish a baseline for subsequent chapters. Above, I introduced my main lines of enquiry, and intended objectives of this research study. I introduce the nature of strategy as a particularly complex and intractable optimization problem. I also made a preliminary connection between complex problems and forms of competitive optimization observed in nature and biological contexts.

Chapter two focuses on a comprehensive chronological review of ML and AI concepts in literature from the 1940s to now, highlighting key events in the history of AI and knowledge engineering, and brings this exploration to present day by exploring the impacts of generative technologies and models. While it may be argued that the huge technical leaps observed in the past twenty or so years is sufficient for this analysis, I felt it was important to undertake a more comprehensive exploration, especially for audiences not directly involved in or versed with computer science or technical fields—this is after all, a doctoral dissertation in business administration, and this study is focused on the application of these techniques in organization contexts. Grounded in literature spanning management science and computational optimizations, I explicated a set of trade-off dimensions and a matrix of pairwise trade-offs between them. I then categorize the trade-off literature in the trade-off matrix, and by industry and functional area, and undertake an extensive analysis of key themes in the trade-off literature. I also offer an assessment of whether current generative AI techniques address the trade-offs, and whether any should be selected as candidates for later analysis, e.g. in chapter four, where I start to examine the application of these techniques in strategic decision-making.

Chapter three contains my research paradigm including ontological, epistemological and philosophical positions, and the methodology employed for this study. Using trade-off constructs from the literature review, I fully describe the search criteria, expressions and sources, and finally, limitations of this study.

In chapter four, I develop and propose models and approaches to maximizing fitness of a production function. After establishing foundational mathematical constructs and grounding in literature, a synthesis of concepts from evolutionary biology, organization and management science, economics, mathematics and computer science takes places. The literature and my analysis supports not only interconnectedness and proposed interoperability of the above concepts, support is established for optimization of organization strategy problems via computation. The chapter contains a multi-step model and approach on how to solve such optimization problems, including rationalized steps where AI technique interventions should take place.

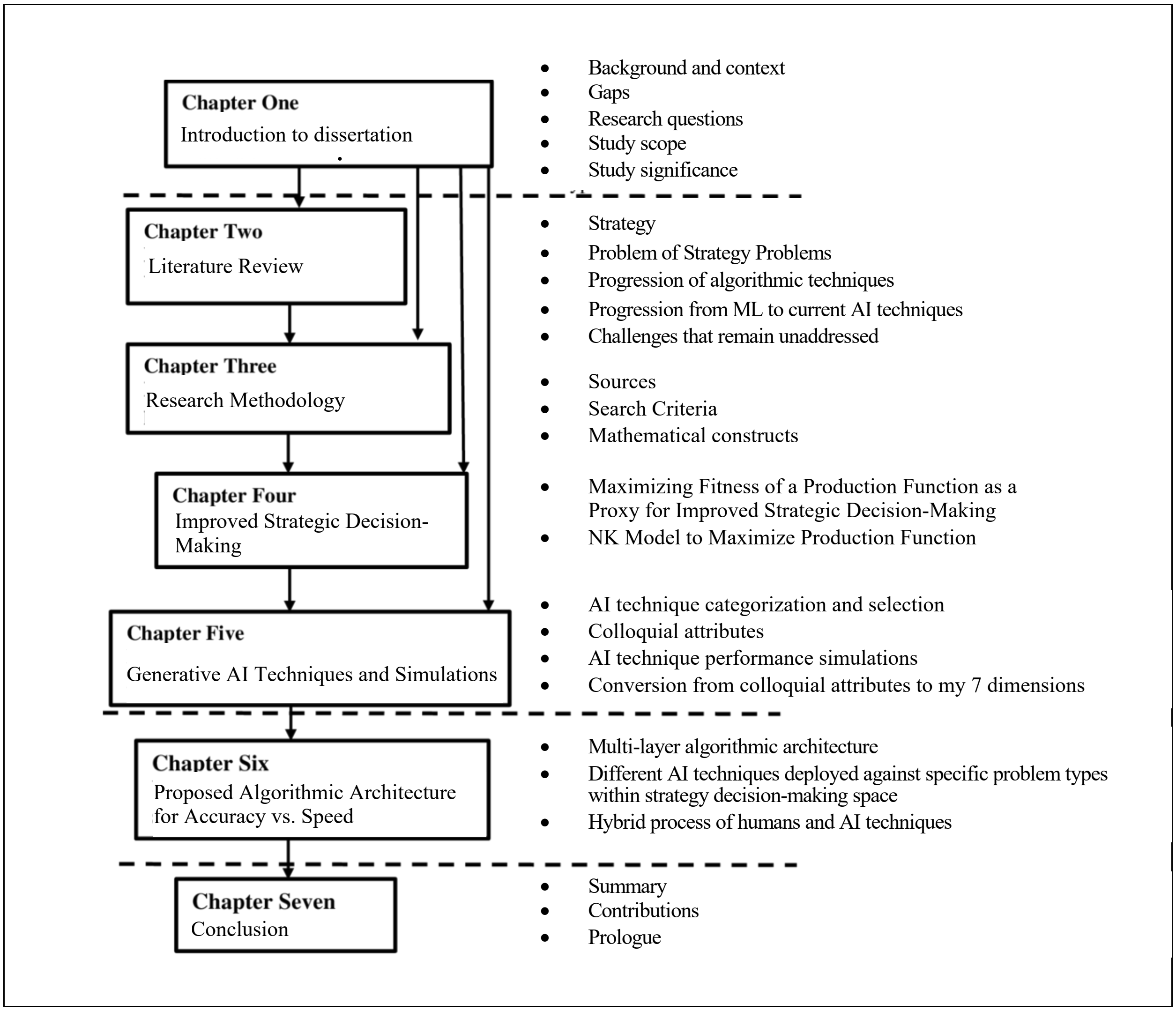
Chapter five begins with basic constructions of concave down functions, with identification of zones of interest to apply AI techniques. I then develop a three-level categorization of organization problems/goals against which different AI techniques should be deployed: 1) trade-off tolerance, 2) problem context, 3) and purpose/efficacy of technique. I next explore widely referenced colloquial attributes of AI techniques, such as parameters of language models. Not satisfied with these attributes, I suggest a conversion and mapping from these colloquial attributes to my set of trade-off dimensions. I then undertake several simulations and analyze the results: 1) individual AI techniques against colloquial attributes, 2) combinations of AI techniques against colloquial attributes, 3) individual AI techniques against my set of trade-off dimensions, and 4) combinations of AI techniques against my set of trade-off dimensions. I conclude this chapter with a discussion of my results.

In chapter six, I develop an approach to optimize algorithmic architecture for the accuracy vs. speed trade-off, culminating in an algorithmic architecture of AI techniques deployed within an organization strategy development process.

I conclude this dissertation in chapter seven, wherein I present my contributions to the theory and literature, and practice of applying AI techniques to organization strategy problems. I also highlight limitations of this study, and implications for future research. My final words are an epilogue to this dissertation.

Figure 1

Dissertation Roadmap (adapted from Fazelpour*, 2016)*



Chapter 2. Review of the Literature

Introduction

This chapter provides a comprehensive review of nearly 500 works from 1945 to 2025, including papers, articles, books, other published literature and some unpublished manuscripts. I will first focus on literature about strategic decision-making, and then move into technical works on the topics of machine learning and AI technologies. With respect to the latter, I will proceed in a predominantly chronological order from conceptual scientific roots, the path from non-self-refining deterministic rule-based inferential approaches and algorithms to self-refining probabilistic inferential approaches, then progressing towards more mainstream application to business problems, and current deep-learning and generative techniques. Along the way, I will weave together an analysis of the application of various technologies and techniques to more complex problem types such as human judgment and decision making, challenges faced, the unique challenges posed by strategic problems, why and how organization strategy can be improved via computational resources, and the trade-offs that organizations and managers explicitly or implicitly make in this context. I will also touch on issues related to the use of AI techniques to emulate complex human-brain functions, and considerations such as the future of work and the ethics of machines making determinations.

The journey to what is today's ubiquitous term and noun *AI*, started in the 1950s with the seminal works of people like Vannevar Bush (as cited in Smith et al., 2006), Alan Turing (1950), John McCarthy (1956), and importantly was continued by Lotfi A. Zadeh[3] (1965, 1975, 1976)

[3] https://scholar.google.ca/citations?user=S6H-0RAAAAAJ&hl=en&oi=ao, retrieved Feb 10, 2019.

wherein he created and published a model, and indeed new language, enabling computers to understand non-discrete or imprecise values for the first time—like humans use to communicate. Deterministic rule-based inferential systems, procedures or algorithms using fuzzy logic or expert systems are a precursor, contributor and alterative to probabilistic-based inferential systems and approaches such as neural networks. The former is expected to yield the same output for the same input—based on a set of programmed rules—whereas the latter answers problems of classification or other optimizations using probabilities and self-refining rules based on existing outputs and training data. This categorization around rule-based versus probabilistic-based systems helps to motivate some of the directions I take, and I'm also inspired by concepts from evolutionary biology, complexity theory, computer science, and management science.

While AI used as a noun, now seems to be everywhere, and in everything (Domingos, 2015; Gunning, 2017; Meek et al., 2016; Namaki, 2018), the journey of reaching the promised land of what AI techniques can do for organizations and their decision making, has been fraught with disappointments and failed to meet the promised vision (Chui at el., 2018; Moldoveanu, 2020; Moldoveanu, 2021; Ng, 2016; Reeves & Moldoveanu, 2017). The issues faced and less-than-expected results seen by organizations are attributable to myriad of challenges including but not limited to the following related concepts: abstract, ambiguous or imprecise formulation of the problem (Bhasin 2020a, 2020b, 2020c; Moldoveanu, 2009; Moldoveanu & Leclerc, 2015), a disconnect between the true problem and the problem being articulated, and a lack of understanding of how mathematical or computational resources can and should be brought to bear. Extending some of existing thinking on the detrimental contributions of uncaptured trade-offs (Moldoveanu, 2009, 2016, 2019a, 2019b; Reeves & Moldoveanu, 2017; others), one can hypothesize that such omissions may be occurring at any or all steps usually undertaken as part of a strategic decision-making process, including problem articulation, systems design, data gathering, analysis, and evaluation of choices.

As part of this research study, I've explicated and examined a set of scalar trade-off dimensions and the trade-offs between them. I then proceeded to clarify and build upon others' primary research into the trade-offs in a comprehensive manner rather than focus on a single or small group of trade-offs, as is the case with most of the existing literature in the context of a particular industry or functional eras. There does not appear to be existing literature which comprehensively examines *all* of the relevant trade-off dimensions in the context of organization strategic decision-making and employing computational resources against same.

Initial Observations

Algorithms, technologies and techniques foundational to current AI techniques have been around since the early-to-mid 1990s (Hao, 2019; Liao, 2005a, 2005b; Martinez-Lopez & Casillas, 2013; Oke, 2008; Rouhani, 2012). More rudimentary rule-based statistical approaches such as fuzzy sets or classification, date back to decades earlier. The literature and technical developments have trended towards increased specialized and an empirical nature—this is attributable to the blurring of lines between research and practice, and in my view greater emphasis in the application and commercialization of AI techniques, compared to the core algorithms and technologies. This is also an artifact of shifts in funding from public sources to industry, the latter also being a substantial beneficiary of these advances (Domingos, 2015). Another cause or by-product—it's not clear which came first—of this blurring of lines is that leading academics are sharing their time and skills between their scholarly commitments and commercial pursuits within industry, for example as seen with pioneers such as Geoffrey

Hinton[4]. He's not alone and the vast financial rewards for all involved is further fueling this economic model (Abreu & Grinevich, 2013; Kinnunen et al., 2018).

AI techniques are thus far unable to fully take on the challenge of abstract and intractable classes of organization problems (Moldoveanu, 2016, 2019a). I agree with Moldoveanu and Leclerc (2015), that part of the issue is the imprecise articulation of the problem itself. Business problems—broadly defined as going from the status quo A to some desired B—they argue, are more akin to situations and predicaments and are not tightly worded discretely solvable problems. I accept this and their prescriptive approach to improve problem solving through improved problem definition and articulation. However, *the problem of* problem articulation *is* solvable, either through their flexons or by other approaches (Moldoveanu & Leclerc, 2015; Moldoveanu, 2016, 2019a). The notion of organizations being unable to gain satisfactory outcomes, e.g. in the quest for accuracy or speed, is widely supported in the literature (Arrieta et al., 2019; Bellamy et al., 2018; Choraś et al., 2020; Holzinger et al., 2018; Liu & Vicente, 2021; Liu et al., 2020; Manheim & Kaplan, 2019; Mehrabi et al., 2019; Moldoveanu, 2020; Moldoveanu, 2021; Papernot et al., 2016; Piano, 2020; Reeves & Moldoveanu, 2017; Rudin, 2019; The Royal Society, 2019; Tolan et al., 2019; others). In many cases and notwithstanding losses in an *opposing* trade-off dimensions, even the primary goal of a gain in speed or accuracy may fall short of expectations. Suboptimal outcomes and trade-offs, while real, do not preclude the possibility that a different approach to strategic decision-making may allow computational resources to make a significant contribution to traditionally human-led or human-only approaches.

[4] https://en.wikipedia.org/wiki/Geoffrey_Hinton, retrieved Apr 10, 2019.

While the nature of this research study and questions asked may motivate some sort of examination of organization strategy and grounded views on what it is and isn't, I do not intend to examine in detail the issue of trying to establish a thesis on an/the optimal models for strategy formulation outside the context of my specific research focus, i.e. from a more organization behaviour or other lens within management study. My focus is in developing approaches and models to improve outcomes detrimentally affected by trade-offs. Also, and acknowledging that there may not be unanimous consensus amongst researchers and practitioners on the definition of strategy or strategy formulation (Hambrick & Frederickson, 2005; Porter, 1979, 1991, 1996), it is not considered a critical problem that has not been examined, for which supported answers do not exist, nor something this study turns on. Finally, I've alluded to above and remain open the possibility that a different approach to strategic problem articulation and solutioning may lead to efficacy gains via computational recourses.

Search Criteria

A theoretical study was undertaken, wherein the principal sources of data are existing works of scholars and practitioners. I introduce my search criteria here, and more details appear in chapter three.

An additive inclusive keyword search using ANDs and ORs was employed, and no specific criteria was created to exclude any particular article or piece of literature (no NOTs or XORs), so long as that it was determined to be relevant to the topic and research question, peer-reviewed or from a credible source, and in English or translated into English. Literature was deemed to be credible if it was peer-reviewed or from an academic source, and purely practitioner literature was kept to a minimum. Several online databases were searched, including

Google Scholar, EBSCO, arXiv.org, Athabasca University Library, University of Toronto Library, and Harvard Business School Library. The keyword index searches included, for example: 1) strategy AND artificial intelligence, 2) strategy formulation AND AI, 3) AI literature review, 4) expert systems, 5) limitations of AI, 6) human judgment AND AI, 7) artificial intelligence. In all cases, Boolean operators were utilized along with both acronym and fully spelled out words, for example: {"AI" or "artificial intelligence" or "ML" or "machine learning"} and "strategy ", {"AI" or "artificial intelligence" or "ML" or "machine learning"} and "judgement", and so on.

The reviewed literature covers a broad array of relevant topics and the research questions of this study, but like all literature reviews, it is a work in progress and should not be considered completely exhaustive. From an initial search set of over 4,000 pieces of literature, additional Boolean search parameters and keywords were used to review the abstracts of over 1500, eventually reaching a total of nearly 500 works that form the primary focus of this review, including over 20 existing literature reviews covering relevant topic areas, and several dozen case studies. I begin with a grounded analysis of organization strategy and strategic decision-making.

Organization Strategy

What is Strategy and How is it Formulated?

> The essence of strategy formulation is coping with competition.
> (Porter, 1979, p. 137)

Delving into any discussion about organization strategy relies heavily on the foundational works of Porter (1979, 1980, 1985 with Millar, 1991, 1996), which have spawned a large volume of literature and discussion, and practitioner learning and adoption. Porter (1985, 1996) has provided a series of empirically-found axioms, which have stood the test of time, pan-industry application and technological advancements. He creates an important distinction between operational effectiveness and strategy (1996)—the former is not the latter—and lays the cornerstone of intra-rival competitive advantage as being predicated on either performing similar activities at a lower cost, or via differentiation. A differentiated product or service offering leads to greater customer value and corresponding willingness-to-pay (1985). However, to outperform a rival the company must not only establish a difference in price or value, it must also be able to preserve it. The sustained delivery of greater value or comparable value at a lower cost, or both, underpins competitive advantage (Porter, 1996). The entire set of activities which allow an organization to create and sustain a competitive advantage is collectively, its strategy. And, this strategic position is sustained through making trade-offs, or deselecting, other *available* positions thereby continually reinforcing the outperforming strategic position. Organizations make trade-offs when certain activities are incompatible with other outperformance-generating ones, and the creation or enhancements of the former necessitates less of the latter (Porter, 1996). These findings ultimately lead us to Porter's unifying concept of strategy:

> Fit drives both competitive advantage and sustainability.
> Positioning choices determine not only which activities a company will perform and how it will configure individual activities but also how activities relate to one another. While operational effectiveness is about achieving excellence in individual activities, or functions, strategy is about *combining* activities. ...strategy involves a whole system of activities, not a collection of parts. ... competitive advantage comes from the way its

> activities fit and reinforce one another. Fit locks out imitators by creating a chain that is as strong as its *strongest* link (Porter, 1996, pp. 44-46).
>
> ... the most valuable fit is strategy-specific because it enhances a position's uniqueness and amplifies trade-offs (Porter, 1996, p. 46).

According to Porter (1996), fit is the critical ingredient that allows organizations to develop and executive superior strategy, thereby creating and sustaining competitive advantage over their rivals, which leads to relative outperformance. The three not mutually exclusive types of fit in Porter (1996) are: 1) simple consistency between activities and strategy, 2) activity reinforcement, and 3) effort optimization, enable organizations to create and executive a system of activities which together cultivate competitive advantage. More fundamentally from the perspective of a strategist or strategy-formulating team, fit can be in thought and action. Going further, fit can be amongst activities, people, groups of people, departments, technologies, and systems. Simplifying and putting things together, superior results vis-à-vis rivals relies on superior execution of differentiated activities, which rely on fit, which relies on a unified strategy. In some ways, this is similar to the more broad-ranging concept of the core competencies of the corporation from Prahalad and Hamel (1990), which encapsulates concepts such as operational integration, learning, and communication. Where Porter (1996) posits repeated superior execution leading to differentiation, Prahalad and Hamel (1990) submit that core competencies enhance and grow with proper utilization. This begs the question of how and where does the strategy underpinning fit, come from?

For this, we also look to Porter (1979, 1980), wherein he roadmaps strategy formulation as an organization undertaking an assessment of an/its industry's competitive forces and their underlying causes, i.e. barriers to entry, substitutes, competitors, then identification of organization strengths and weaknesses, followed by a corresponding organization "posture vis-à-vis the underlying causes of each force" (1979, p. 143). The organization may position itself vis-à-vis these forces to defend against competitors or entrants, improve their position via moves or re-balance, or move before others to capture an opportunity. Connecting with Porter (1996), these sets of positioning actions must be aligned across and up/down the organization.

The importance of information technology in and for strategy (Porter & Millar, 1985) continues with Morton (1986), which gives an insightful indication of the complicated nature and interplay of strategy and information technology. Drawing upon Simon (1960, as cited in Morton, 1986), Morton creates a direct connection between the "rational actor" organizational model, with effective use of IT within the context of business strategies. While Morton's thematic focus may be on IT strategy formulation, many of his insights and conclusions are applicable to general strategy formulation. Of particular interest is Morton's fourth premise (1986), in which he highlights the difficulty and non-scientific nature of strategy formulation, making "good strategic management" (p. 5) a rare observance, and his balanced yet thoughtful definition of strategy formulation:

> …strategy formulation as being how to create an appropriate mission, and to position the organization to accomplish this mission in light of the reality of its internal strengths and weaknesses, its customers and the external environment, Strategy formulation is concerned about the desired positioning of the firm and how to get there. (Morton, 1986, p. 9)

While this definition certainly has tones of Porter's Five Forces Framework (Porter, 1979), it is on its own useful in the current analysis. Morton concludes that strategic problems cannot be answered via formulaic techniques, and suggests the use of frameworks and methodologies to generate ideas, and build consensus. In his discussions of the future, Morton succinctly incorporates emerging ideas such as the reduction in hardware/software costs as facilitators to rapid change and technology integration, the development of heuristics and the capture of human judgment via expert systems, thereby allowing for the transition from data to information to knowledge.

Mintzberg (1987a) creates a valuable distinction between planning strategy *versus* his preferred crafting process. Like a clay potter, he posits that skilled strategic managers *form* and formulate strategy like the potter does clay, simultaneously balancing and aligning organizational capabilities with opportunities—the past, present, and the future. Importantly, Mintzberg states that strategy can be the output of a deliberate process of thinking and analysis, or it may emerge out of an “evolving situation” (p. 68). In the context of this study, this separation of mind and hands can be viewed as a systemic need to have feedback and ability to alter course based on new data. In other works, Mintzberg (1987b, 1987c) tells us that strategy is a combination of five “Ps” as opposed to a single monolithic definition: plan, ploy, pattern, position, and perspective; and each of them in themselves can be an organizational strategy. Mintzberg (1987c) discusses the need for organization to have strategies—a good strategy sets the direction but in order to outperform, the organization must execute. A good strategy aids in focusing and aligning efforts and integrates business parts and units, while defining the organization and providing consistency. This is relevant in context of topics such as improving predictive accuracy and the challenges in using algorithms or systems to replace or enhance human judgment (Agrawal et al., 2018a; Agrawal, 2018; Bettis & Hu, 2017; Brezillon, 1999; Oke, 2008).

Another view of strategy is presented by Eisenhardt (1999), especially within rapidly changing organizations and industries, as rapid and effective strategic decision-making. She submits that effective decision-making firms and leaders exhibit the traits of speed, quality and support and that these are not in conflict nor mutually exclusive. Effective strategic firms and decision makers create strategy through four approaches: 1) building collective intuition to see opportunities and threats faster, 2) stimulating quick conflict, 3) maintaining pace, and 4) defusing politics. These insights are of interest to this paper because these key approaches of these subjects may be modelled and improved via the application of specific AI techniques. Similarly, Nasi (1999) proposes a process of strategic decision-making via three distinct models: management process, game-playing, and leadership. He highlights some of the issues with strategy-making, including past and present data and information, the need for facts *and* values, analysis and synthesis, and the more complicated phenomena of logic vs intuition. Moving from strategy-making as a management process towards leadership, the processes and challenges become more abstract and difficult to model and design for a computational approach.

Given the wealth of scholarly and practitioner work and articles on the subject of strategy, it may be easier to reach a consensus definition of this complicated subject by first defining what it is not. Hambrick and Frederickson (2001) state that strategy is not piecemeal nor specific to any particular business unit nor department, nor is it a target. It is not a goal, statement, rigid, forever, nor immediate. Also most interestingly, they state that it is not internal or inward-facing, nor a catch-all—strategy is a set of outward-facing “integrated, mutually reinforcing set of [five] choices forming a coherent whole” (2001, p. 54) which are guided by the organization’s mission and objectives. They perform a detailed construction of strategy using their strategy diamond and lay out the five elements of a sound strategy: specific arenas (where will we play?), vehicles (how will we get there?), differentiators (why us?), staging (what’s our execution plan?), economic logic (how will we benefit?). Expanding Hambrick and Frederickson

(2001) and supported by Moldoveanu (2009; with Leclerc, 2015), much academic and empirical research seems to indicate that most intra-organizational practitioners are neither defining, formulating, nor executing strategy correctly. After 30 years of looking at the subject, "guidance on what actually constitutes a strategy is missing" (Hambrick & Frederickson, 2001, p. 48). This observation is important in context of not only strategy formulation as an independent science, art or set of skills, but even more so if we consider the possibility of AI techniques being able to improve strategic decision-making and achieve better and more useable outcomes.

The above works support my earlier suggestion that organization strategy, while being an old and widely studied area spanning several decades, still lacks complete consensus. Strategy may be some or all of the above, and its definition may be a live issue to the extent it is. Like many areas of organization science, it's a non-unanimous and non-homogeneous quasi-fuzzy picture largely based on observations and longitudinal qualitative research of inter-organization competitive performance across geographies and industries. However, this does not derail nor cause this research study to change course. The concept of how *better or best* to solve strategic problems does present opportunities for improvement via algorithmic and/or hybrid approaches.

For this study I will largely adopt Porter's widely accepted definition of fit and organizational characteristics as the key drivers of strategic differentiation and competitive outperformance (1996). Organization problems such as strategy formulation are at their core model-based optimization problems (Moldoveanu & Leclerc, 2015). Therefore, I will consider probabilistic-based inferential AI technologies applied to strategic decision-making as an optimization problem. Computational systems are particularly adept at solving optimization problems in the general sense. As a unique class of optimization problem, strategy presents other challenges for algorithmic systems, which I will examine in detail below.

Problem Complexity Classes

Organization strategy is a type of problem (Moldoveanu, 2009; Moldoveanu & Leclerc, 2015), and one way to classify problems is by their complexity. In the context of this study, complexity is an important attribute in that any model or approach to improve organization strategic decision-making processes, cannot ignore complexity. In a computational context, the concept of complexity goes to measures such as time-to-solve, and considerations of system design and resources such as processing chips, memory, network, and architectures. One way to examine and express problem complexity is in terms of Polynomial (P) and Nondeterministic Polynomial (NP) time classes, which respectively, refer to the time required for a computer to solve a problem where time is a function of the size of the input (Baker et al., 1975; Cook, 1971, 2000). These are important mathematical and computer science concepts applicable to this study, and particularly in the context of utilizing computational resources against types of problems to improve outcomes. This brief introduction to complexity concepts such as P and NP will help readers who may be unfamiliar with these, and a more detailed overview is contained in Appendix B.

Polynomial Time class problems, or P-class for short, have solutions or solution sets which at most grow in polynomial time as the problem variables increase, e.g. constant, linear or logarithmic time as a function of input variables. Examples of P-class problems include arithmetic, sorting or generalized Checkers (i.e. a defined board size such as 8x8).

Nondeterministic Polynomial Time class problems, or NP-class for short, have solutions or solution sets which grow at greater than polynomial time, e.g. exponentially or higher, as the input variables increase. Examples include non-generalized Sudoku or non-generalized Chess, i.e. a non-defined or NxN board (Wigderson, 2006).

It is important to note that within both P-class and NP-class problems, there is a spectrum in terms of complexity and the measures *time-to-solve* and *time-to-verify*. This consideration will be helpful as I proceed into my analysis of problem-solving approaches and model development. Moldoveanu (M. Moldoveanu, personal communication, September 18, 2022) argues that there is evidence to suggest that managers prescribe to selective problem scenarios when evaluating problem complexity and planning scenarios, e.g. best case, average, expected, worst case. However, 'there is no evidence that suggests that strategy researchers think it possible that managers could do this' (M. Moldoveanu, personal communication, September 18, 2022). In other words, managers simplify or otherwise reduce the strategy problem at hand.

Most strategy problems are NP-class and often NP-hard (Levinthal, 2011; Moldoveanu & Bauer, 2004; Moldoveanu, 2009, 2016; Rivkin, 2000; Weinberger, 1996), meaning they cannot be divided or solved via a sequence of simpler, say P-hard, sub-problems with sub-solutions. While approximations can be made with varying degrees of accuracy using various techniques, the nature of NP problems means that they cannot be *deterministically* solved in P-parts. Similarly, Porter (1996) argues that a dominant or winning strategy is a system of activates and *not*, in fact, a collection of parts. If we accept this, is there a path forward for computational resources, such as advanced AI techniques, deployed against NP-class problems? And if so, can they stand alone, or would some human involvement be necessary or beneficial?

The Problem of Strategy Problems

Moldoveanu (2009) posited that managers prefer to conceptualize, engage with and *try to* solve P-class canonical easy problems in polynomial time, over NP-class hard problems which take greater than polynomial time, or in a *worst* case, exponential time. The calculus of P vs. NP is a factor of the problem's variables, constraints, and number of operations required to reach a solution. I introduced P and NP in my Introduction (Chapter 1) and provide more detail in Appendix B; for a more detailed analysis of the mathematical and computational challenge of P vs. NP and whether P = NP or P ≠ NP, I point the reader to the voluminous literary record focused on that topic, e.g. Baker et al., (1975), Cook (1971, 2000), Goldreich, (2010), Sipser (1992). While Moldoveanu's binary classification of managerial problems into P or NP may appear simplistic in the context of the myriad of managerial or strategic problems, it achieves its desired goal, which is to highlight observed biases and problem-simplifications undertaken by managers, and how to overcome these. This is an important development in the context of this research study because it directly goes to defining difficult classes of strategic problems, variables and solution selection biases and heuristic reasoning. Critically, Moldoveanu explicitly draws a connection between problem type and complexity, and their algorithmic solutions procedure (2009). Moldoveanu implicitly posits that most if not all *properly constructed* and addressed strategic problems are NP-hard. I draw attention to the term properly constructed, because the implicit or explicit simplification or manipulation of the problem is a major contributor to organizations not achieving anticipated outcomes—simply put, they are asking the wrong questions (Moldoveanu & Leclerc, 2015). This reinforces many of the gaps identified above in terms of how AI techniques are thus far unable to produce satisfactory outcomes against NP-class problems. Moreover, the managerial tendency to simplify or simply ignore key complexity considerations is an important finding for this research study and the research questions.

Continuing down this examination into the nature of organization problems, Moldoveanu and Leclerc (2015) present business problems as unlike math, economic or engineering problems. They posit that these problems are more like 'quandaries, issues, predicaments, and situations that must be defined and articulated before they can be solved' (p. 5). The key challenge, then, is the development or use of language to turn an abstract, fuzzy and *presumably* suboptimally articulated predicament into a crisp defined and precisely articulated problem. This according to Moldoveanu and Leclerc (2015), is a combination of prediction and optimization problems. The prediction is the *What*, i.e. what should we do about X, while the optimization is the *How*, i.e. how do we get from A to B. Enabling superior business problem solving, then, is heavily predicated on superior business problem articulation. Where frameworks often used by strategists and other business problem solvers have limitations and are fixed like computer system hardware, language—actually multiple languages—present an adaptable lens through which problems are viewed and analyzed, in order to turn imprecise situations into precise *solvable* prediction and optimization problems. To that end, Moldoveanu and Leclerc (2015) propose five (5) such languages called *flexons* ("flexible objects for the generation of novel solutions", p. 6), and for which they introduce an associated lexicon.

Like tools in a problem-solvers arsenal chest, the flexons (summarized in Table 1) can be used in series, parallel, any combination or permutation, and repeatedly. Individually and collectively, they allow the problem-solver to achieve prediction and optimization through the redefining and refining of abstract situations or predicaments—the original problem. The use of flexons not only reimages the problem itself, but allows for tangible progress towards solutioning.

Table 1

Summary of the Five (5) Flexons or Problem Articulation Languages proposed by Moldoveanu and Leclerc (2015)

Flexon	Description
Networks Flexon	The Who, forming a nexus of nodes and the connections between them.
Decision Agent Flexon	The selection of the Who (Networks), along with associated decision rights, aligned incentives and measurements, and resources.
System Dynamics Flexon	Define the industry or business as a whole, then allocate effects and impacts of changes to stock and flow of inputs and outputs.
Evolutionary Flexon	Viewing organizations and the people within them as organisms subject to external competition and internal selection. Survival, selection and outperformance depend on competitive fitness, which depends on fit between organism characteristics and external environment. This line of thinking is particularly useful as the reader will see below. Organisms and organizations undertake a series of 'mutation-variation-selection-retention' (pp. 168, 180) processes to offer products and services that will outperform in the competitive landscape.
Information Processing Flexon	"What information is used, how much memory and memory access cost, how costly each computation is, and how efficiently the computational device is at solving certain kinds of problems" (p. 194).

The notion of the Information Processing Flexon opens up the idea of how best to solve a problem, by defining characteristics such as problem complexity, level of accuracy desired, and resources required to get from "*where we are*" to "*where we want to be*" (Moldoveanu & Leclerc, 2015, p. 203). This presents interesting avenues of attack in the context of this study, and as the reader will see, presents possible opportunities in the context of the challenges posed by strategy problems by expanding the toolkit of problem articulation, solutioning by parts, and associated variable lexicons.

The concept of flexons is a means to an end, wherein the essential underlying problem in some of Moldoveanu's recent research is and remains, how to best solve the intractable abstract problems that organizations and managers face. Moldoveanu's research in the directions of problem redefining/refining/translation is present in much of his recent work (2016, 2019a, 2019b, 2020, 2021). The key takeaway of interest to this study is his well-articulated thrust that a key challenge lies in how organizations define problems. An efficient fitness-fit solution can only be found when a problem is first properly articulated. Prediction and optimization can only take place against properly articulated problems. And, relevant powerful attributes of computational systems can only be brought to bear once we've *re-stated* these abstract intractable predicaments into *solvable* precise problems. This is not a simplification of the problem to better suit available means; this is an improvement in specificity and precision, while accounting for required accuracy and available computational or other resources.

I posit that the application of AI techniques to organization strategy is challenging regardless of exactly how the process of strategic decision-making is defined—so long as it involves a process and associated steps, such as: a need to address an organizational problem or goal, data gathering, information assessment, evaluation of alternatives, decisions, execution, and so on.

I will now proceed deeper into the concepts of organization strategy and competitive outperformance. What enables certain organizations to outperform their competitors in the same but ever-changing competitive landscapes? In this analysis, we learn some distinctive aspects of organization strategy and why its complexity, while also creating challenges, may be particularly well-suited to more optimally address via computational resources such as machine learning and other algorithmic approaches. The following section examines a line of research that treats organizations more as biological organisms and less as structures or groupings of the latter. This creates a connection between Porter's fit (1996) within an organization, and how an organization adapts within a competitive landscape.

Strategy Problems are Intractable NK Fitness Landscapes

Porter states that fit is derived and achieved from a system of activities, *qua* precursors to strategy and competitive differentiation (1996). He suggests inherent differences in not only the actions, sequences of actions and intra-organization structures, but almost genetic phenotypic differences. A phenotype is an observable expression of a gene or genetic encoding, resulting from the interaction of an organism with its environment (Wojczynski & Tiwari, 2008). Taken as a system of activities and choices, some organizations exhibit superior fit. Moldoveanu provides examples of this fit from activity sets to demand characteristics, technological and production functions, regulatory frameworks, and an end goal sought by strategists (M. Moldoveanu, personal communication, February 2 & 4, 2022). Because it's a system view and in a way all-encompassing, this suggests that superior-performing organizations are wired differently—like the genetic blueprint or DNA in living organisms. An external observer cannot gain access to this DNA, but can bear witness to an organism's phenotype or externally *observable* characteristics.

So, if this phenotype is analogous to an outsider's observation of another's expressive fit, does it follow that an organization taken as a whole, exhibits other activities that emulate living organisms? Prahalad and Hamel support this notion via their concept of core competencies (1990): core competencies are the underpinning root structure, glue, and innovation engine of an organization. These core competencies don't diminish but rather grow, and enhance with consistent use and nourishment, like rings of a tree. Similar to Porter's fit, Prahalad and Hamel (1990) offer up a clear empirical sequence: core competencies lead to core products, which are grouped together and optimized into business units, that then produce marketable end products.

For its utility as a possible unifying explanation and conceptual platform, I'll now turn my attention to the NK Model first proposed by Kaufman and Levin (1987) and subsequently extensively developed and supported in Kauffman (1988; with Weinberger, 1989; 1990; 1993), and notably for this study, expanded into management science by Weinberger (1990, 1996), Levinthal (1997), Levinthal and Warglien (1999), Rivkin (2000), Rivkin and Siggelkow (2003), Siggelkow and Levinthal (2003), Rivkin and Siggelkow (2007), Ghemawat and Levinthal (2008), and Levinthal (2011). The NK Model is fundamentally a "vector of traits characterizing an organism" (Kauffman, 1988, p. 276), which I will bridge with Porter's fit (1996) and other works to help ground *successful* strategy as the creation and development of intra-organizational fit to enhance inter-organizational performance within a competitive fitness landscape.

The study of evolutionary biological adaptation in Kauffman and Levin (1987) has put down an important foundation for the notion of adaptation as a "complex combinatorial optimization process" (p. 11), and indeed, problem. Organisms undergo biological/genetic adaptation at the gene and peptide level, as they adapt to their landscapes. The adaptive walk of, or towards, a more fit genetic variant is an optimization process. The fitness landscapes are the backdrop and *problem* against which genetic optimization ensues. Visualize an organism beginning as some initial genetic combination—or more accurately, a peptide sequence—on a fitness landscape of some level of ruggedness (or smoothness). Via an adaptive walk of changes within its peptide sequence, the entity genetically moves or "walks" to more fit combinations, towards local or global maxima. N in the NK Model represents the number of amino acids in the peptide (the length of the peptide chain), and K reflects the number of possible interactions between these peptides. So, for the completely smooth landscape case with K=0, the *fitness contribution* of any locus in the chain is independent of all others and there are no interactions. For the fully dependent and most rugged case where K=N-1, all loci in the peptide effect all others, thereby rendering the fitness landscape completely correlated. In the case of two (2) possible amino acids (dimensions) for each locus in the chain, we have 2^N possible genotypes or combinations of traits. Therefore, for a peptide of length fifty (50) with twenty (20) possible dimensions, we get $20^{50} = 1.13 \times 10^{65}$ possible combinations. Kauffman and Levin (1987) connect the above biological evolutionary process with non-biological optimization problems such as the TSP in computer science, and spin glasses in physics, supporting the notion of a general theory and model for entities on landscapes, and that improvement by local mutation and selection is important in each. Key to this study is their finding that adaptive walks and other optimization problems are NP-class and therefore, the solution space grows faster than polynomially with increasing N.

This area of study into biological or evolutionary phenomena has been expanded into organizational and management study. More recent work has simulated the NK Model onto organization design (Ghemawat & Levinthal, 2008; Levinthal & Warglien, 1999; Rivkin, 2000;

Rivkin & Siggelkow, 2003; Siggelkow & Levinthal, 2003). Moreover, the connection between Kauffman's general model and organization strategy problems begins to emerge, with the latter acting as what I'll currently describe as a relatively correlated rugged landscape, with cross and longitudinal interconnectedness. Correlated in this context applies to actions and reactions between organizational actors within a competitive fitness landscape, and also changes in the landscape itself. I posit that in the context of organization strategy, there is no such thing as a completely uncorrelated or random fitness landscape, and any randomness will be temporary and time-bound due to natural competitive forces such as information flow and competitor imitation. It's also worth highlighting that even as the understanding has deepened and expanded, I did not find any substantive literary disagreement or divergences in terms of the validity of the NK Model and its application to areas such as my research. This lends support to the robustness of the notion of adaptive walks on fitness landscapes as optimization problems.

Bridging Fit with the NK Model

Having introduced the connection between concepts of other complex systems and this research study (Cohen, 1999), where complexity is both a cause and a consequence (McKelvey, 1999), I'll now bridge Porter's fit and Kauffman's NK Model more concretely. McKelvey (1999) posits that multicoevolutionary complexity at the firm level internalizes natural selection processes, thereby turning interdependencies between intra-firm and inter-firm parts or activities, into segments of Porter's value chain. Let's take a walk through the progression of this area of study from the seminal works of the late 1980s focused on biological adaptation, to organization and management theory.

I'll start with a consideration of landscape characteristics and how these may impact entities *adaptively walking* on it. In his analysis, Kauffman (1988) draws out an important distinction between adaptive walks on rugged landscaped compared to smoother landscapes. While adaptive evolution occurs on both types of landscapes, the rate of evolution via fitter genetic variation rapidly slows on rugged landscapes, or where the organism makes a genetic "long jump" to a distant landscape. On smoother fitness landscapes, this rate of successive adaptation slows more gradually. Because this is an optimization problem with a goal of achieving an optimal state such that no further adaptive step offers a more fit option, and we've seen from Kauffman (with Rivkin, 1987; 1993) that as K increases, the ruggedness of the landscape increases, Kauffman and Weinberger (1989) have said that a more rugged landscape increases the frequency of local optima, reachable via shorter walks. Weinberger (1990) focuses on properties of the landscapes themselves via a suggested framework for their mathematical treatment, and expands its applicability to other disciplines outside of evolutionary biology. In the context of computer science, the fitness landscape becomes a set of allowable configurations within an optimization problem. This work further supports the use of evolutionary strategies in developing new methods of solving combinatorial optimization problems (Papadimitriou, 1977). Weinberger's key conclusion relevant to this study is the finding that fitness landscapes autocorrelate similarly to a decaying exponential, and how their model can be used to "tune" combinatorial optimization algorithms to locate optima such as the ones I will be considering. Measurements of the autocorrelation function are a means of classifying different types of landscapes, i.e. problems, and act as a prelude to a theory of which optimization strategies work best on which types of landscapes. Weinberger (1996) continues his analysis of rugged landscapes but this time focused on whether they are P or NP, and why. He asks why P ≠ NP for different fitness landscapes, and suggests that qualitatively similar fitness landscapes can differ considerably, answerable in polynomial time if the problem is efficient, or be NP-complete (if inefficient). We've seen that K can be "tuned" from K=0 (no interaction between sites) to K=N-1 (complete interaction between sites). Weinberger (1996) proves that the NK problem with adjacent neighborhoods is in P-class of problems, solvable in polynomial time with N, and it's NP-class with random neighborhoods for $K \geq 3$. In other words, a tuned correlated landscape with $K<3$ is in P, while an uncorrelated landscape with $K \geq 3$ is NP/intractable and yields superior or higher optima.

Levinthal (1997) brings the NK Model into management science, by proposing a model focused on interrelationship between organizational level change processes and population selection forces, based on the premise that organizational fitness and form are interactive, and collectively make up the rugged landscape. The landscape is then the space of alternative organizational forms. This interrelationship means that organization form and strategy interact, and the fitness contribution of form effects the topology of the landscape (Kauffman, 1993). And, interaction effects suggest that similar landscapes may contain populations of diverse forms because the contribution of each may be different. The contribution of an organization, due to its original form and current place on the landscape, may be different than another's residing in a similar location. As local search and adaptation continue, a reduced set of dominant forms remain at local peaks. These are then subjected to selection forces to create a distribution of forms, eventually leading to one dominant form within the population on that landscape. Hence, organizational self-adaptation leads to selection or de-selection.

Levinthal and Warglien (1999) highlight the importance of organizational design as a tool to tune landscape design by examining how the former influences fitness landscapes and corresponding behaviors. Their thrust relevant to this study is, as strategy is formulated and self-reinforces direction as prescribed by Porter (1996), the landscape topography is sensitive to and indeed, reacts to these actions. A robust strategy shines light on a few select *greater-than-local* activity peaks within the fitness landscape. In this way, a firm enables itself to tune its landscape, thereby allowing for a coherent competitive strategy of lowers costs or greater differentiation, or both, to transform a rugged landscape of many local peaks into a smooth surface with a small number of potentially global peaks. Levinthal and Warglien (1999) propose the use of powerful and accurate analytical representations of the fitness landscape, as a tool to reduce landscape dimensionality and complexity, then guide in direction. This area of study and analysis into the manipulation of interdependencies, is another powerful parameter organizations can consider to optimize strategy and outcomes.

Rivkin (2000) asserts the widely supported notion that complexity in strategy deters imitation (Porter, 1980, 1996). He proposes a model capturing two critical parameters, a *new* N and K of complexity: 1) N, the number of elements or decisions comprising the strategy and 2) K, the degree of interaction among the elements or decisions (N). This complexity not only deters imitation, but it renders search for an optimal strategy intractable—a NP-class problem. In this case, imitation can only be attempted through heuristics or learning, both of which are also hindered by complexity. As in Porter (1996), the selection and deselection of activities and decisions that comprise an inimitable competitive embodiment, here work to deter imitation, because a heuristic or learning of this itself becomes insurmountable. The organization developing this strategy *de novo*, through the two parameters highlighted by Rivkin (2000) renders its own strategy intractable. Similar to Porter's (1996) articulation of tight system-wide fit as a competitive differentiation, Rivkin (2000) posits that a set of imitable activities taken together become inimitable. Of interest here, Rivkin goes further in dealing with the intractable nature of complex strategies, suggesting that managers forsake global optimization, and rather satisfice or search for *good enough* sets of decisions and activities (March & Simon, 1958, as cited in Rivkin, 2000; Moldoveanu, 2016, 2019a). In this way, strategic decision-making can be viewed as an upward-seeking adaptive search for payoff peaks in a "very-high-dimensional decision space" of combinations of choices (Wright, 1931, 1932; Rivkin, 2000, p. 827). Rivkin poses a kind of catch-22 gauntlet wherein global optimization (complex winning strategies) is intractable in the very interactive systems that are able to create same, which can in-effect render them unattainable. Iteration is ineffective due to compounding constraints and penalties for mistakes. Rivkin proposes an escape: a wholesale, but risky, reconfiguration or "long jump". While Rivkin (2000) mentions the potential for transformations that convert an NP-complete problem into P-complete, he reaches the same conclusion as Weinberger (1996) with respect to algorithm computational time: K=0 is P-hard and optimizable in polynomial time (time complexity is a linear function of N), but for K>2, the strategy decision problem is technically intractable, and becomes literally intractable for high N and, even with computational speed increases, the time needed for each algorithmic step is high. Rivkin's model helps explain the asymmetry in adoption and transmission of "winning strategies" that may be in plain sight, present for a prolonged period, but at best have only piece-meal copycats.

Rivkin and Siggelkow (2003) continue the study of organization design, and examine how and why elements of organizational design depend on one another. Their key finding is that success is predicated on organizations being able to conduct broad searches for good sets of decisions, and stabilize around these good sets once discovered, thereby balancing search and stability. This reinforcing but seemingly counter-active activity set leads to interdependencies among the design elements. In effect, the goal of attainting a high spot on the landscape leads to a higher payoff only if such a spot if found *and then* stabilized-upon.

Siggelkow and Levinthal (2003) takes yet another approach to how organizations can locate, find and adapt for highest payoff. As in Porter (1989, 1996), organizations can gain competitive advantage through internally consistent configurations and sets of activities, *and* aligned with the requirements of their current environment or landscape. This is not a trivial problem in an environment of changing peaks, that require alteration of activities. Their suggestion is to not only balance the interdependence between activity configurations and environmental requirements, but at times suspend or strengthen these to search for new optima. This reinforces the findings in Rivkin and Siggelkow (2003) of superior performance through balancing exploration and stability, but here it's the effective sequential deployment of different strategies that leads to superior payoffs.

Ghemawat and Levinthal (2008) present an analysis using Kauffman's NK Model to demonstrate whether optimal strategy is a tightly executed set of activities as in Porter (1996), or a longitudinal sequence based on a set of capabilities, or is it an evolving process. Their key findings support the notion that constrained *a priori* strategy-making is important for achieving fitness for higher peaks on the landscape, subsequent to finding myriad of local peaks via local search, with a target to achieve a peak at or close to the global optimum. While the connection to Porter's fit (1996) is clear, this also brings back shades of Mintzberg (1987a, 1987c) and Hambrick and Frederickson (2001), wherein those scholars have argued that strategy may be a process or an output, and an outward-facing integrated reinforcing set of choices, respectively. All of these concepts point to an *a priori* inception.

How can Fit and NK be Used?

Strategists seek to achieve Porter's intra-organizational fit and inter-organizational differentiation by thinking, analyzing, doing, and aligning execution, or combinations thereof. The thinking can take place with human brains-intellect-experience-judgment, or with microchips-memory-network, or combinations thereof with varying levels of algorithmic and system architectural sophistication. It is generally accepted that most strategy problems are in NP and perhaps NP-hard (Levinthal, 2011; Moldoveanu & Bauer, 2004; Rivkin, 2000; Weinberger, 1996), which means that they cannot be articulated or deterministically solved as a set of simpler, say P-hard, sub-problems with sub-solutions. In other words, it is not a combination of the parts. Porter also argues that a [dominant] strategy is a system of activates and not, in fact, a collection of parts (1996). So, is this the end of an algorithmic approach to optimization or maximization? Perhaps an answer lies in choosing a level of accuracy that renders the problem solvable, yet retain sufficient reliability (Moldoveanu, 2016, 2019a; Rivkin, 2000). This may mean local maxima/peaks of sufficient height, rather than a global maximum.

It is unclear whether intractable NP-hard problems can in fact be *transformed* into algorithmically solvable P-hard sub-problems, and then presumably recomposed into the whole, at least in terms of an efficacious utility function and evaluation. Moldoveanu (2009, 2016, 2019a) posits that NP-hardness does not preclude the use of computation and algorithms so solve such problems, and proposes the use of non-deterministic, intelligently stochastic systems configured and deployed on optimal architectures of humans and machines of theoretical computer science. He prescribes the art of the solution as the formulation of the problem and design of the algorithm and system to solve for the solution—in effect tuning the problem, the landscape, and then computing the solution (2019a):

> *Intelligent Artificiality* shows how the languages and disciplines of theoretical computer science, 'artificial intelligence' and computational complexity theory can be used to devise a set of heuristics, blueprints and procedures that can help strategic managers formulate problems, evaluate their difficulty, define 'good enough solutions' and optimize the ways in which they will solve them *in advance of attempting to solve them*. (Moldoveanu, 2019a, p. 2)

Moldoveanu introduces both a framework for the analysis of strategic problems in computational terms, and a set of prescriptions for strategic problem formulation and problem solving relative to which deviations and counter-productive moves can be specified and measured. He suggests (M. Moldoveanu, personal communication, February 4, 2022) that strategists should think more like algorithm designers to design, develop and deploy computable algorithms on the same machines they themselves use to solve big harry problems or achieve big hair audacious goals (Collins & Porras, 1994). Moldoveanu does not necessarily contradict others like Rivkin (2000), but proposes a subtle recombinant approach, whereby strategy may be formulated via automation-by-parts. For this, we will need to examine the myriad of algorithmic approaches and solutions computer scientists have developed for organization strategic problems and goals. This can be viewed as a form of instantiation—to borrow a software development term—of automation-by-parts of a strategist's production function. But is it really and if so, to what degree and at what *total* cost? Better understanding needs to take place in terms of what humans require when interacting with "decision-making, explanation-generating machines" (M. Moldoveanu, personal communication, February 4, 2022).

It's important to note that this type of automation-by-parts poses several issues in terms of the optimizations considered via the NK Model and Porter's fit (1996). Furthermore, they would seem to render such strategies-by-parts as easier to imitate, thereby making the organization less competitively distinct within the landscape. Even if automatable-by-parts models for strategy formulating can be found, and are found to be effective, does this come at only computational or other easily addressable costs?

The literature contains examples of trade-offs that occur when computational machines are used to solve even relatively simple problems, far simpler than NP-hard strategy types. I now undertake an expansive exploration of the roots of what we now call AI, from its conceptual beginnings to rule-based deterministic approaches, to some of the most current AI techniques.

An Abridged Chronology from AI's Roots to the Present[5]

Background and Roots

Academics in engineering and computer science have been developing deterministic rule-based inferential approaches such as fuzzy sets[6], fuzzy logic[7] and expert systems[8] for several decades. The actual term and/or field of artificial intelligence is believed to have been coined by

[5] This chronology is considered abridged because: 1) The voluminous literary record since the 1940s cannot be covered in its entirety; 2) Gaps during "AI Winter" periods, e.g. in the 1970s and late 80s to early 90s (https://en.wikipedia.org/wiki/AI_winter, retrieved April 1, 2019); and, 3) Gaps in publicly available literature due to research becoming more privately funded by and for the benefit of corporate or for-profit organizations.

[6] https://en.wikipedia.org/wiki/Fuzzy_set, retrieved Feb 10, 2019.

[7] https://searchenterpriseai.techtarget.com/definition/fuzzy-logic, retrieved Feb 10, 2019; https://en.wikipedia.org/wiki/Fuzzy_logic, retrieved Feb 10, 2019.

[8] https://searchenterpriseai.techtarget.com/definition/expert-system, retrieved Feb 10, 2019; https://en.wikipedia.org/wiki/Expert_system, retrieved Feb 10, 2019.

John McCarthy and came from a workshop at Dartmouth College in 1956[9], Vannevar Bush first published mention of a system which augments human knowledge and understanding in 1945 (as cited in Smith et al., 2006). Then Turing (1950) described the concept of intelligent machines being able to simulate humans and perform intelligent functions such as playing chess (Smith et al., 2006). While Zadeh wrote his seminal works more than a decade later from the mid-1960s, and focused on far more technical and mathematical issues, it can be argued that his work was a crucial step to lay the ground work for machines being able to compute non-discrete information. Without minimizing the ground-breaking work by scholars before him, in my view Zadeh's critical yet under-appreciated (Negnevitsky, 1997); Zadeh, 2015) development of fuzzy sets (1965) and then fuzzy logic (1975), enabled the application of computer and logic concepts to human-type problems in human-type language. These were foundational steps to be able to get to what is now known as machine learning, algorithmic intelligence or artificial intelligence. Starting with these pioneering works is important, to set the stage for considering and applying AI techniques to strategic decision-making.

In short, f*uzzy things*, unlike *crisp things* (Zadeh, 1996), are approximations of a true value. The latter can be anything with a precisely defined value, such as a number or absolute truth (represented as a 1 or 100%), while the former would include the concepts of probabilities, likelihoods, and imprecise values. While the precise value of a number may be *x*, what is the value for *beautiful* or *large* or *tall*? Fuzzy things, sets and logic allow us to represent the imprecise concepts of language and words in ways computers can understand and compute, and

---

[9] https://en.wikipedia.org/wiki/Dartmouth_workshop, http://www-formal.stanford.edu/jmc/slides/dartmouth/dartmouth/node1.html, both retrieved Feb 10, 2019.

allows for linguistic measures, e.g. *not likely*, *likely*, *very likely*, and so on (Zadeh, 1976). This was the beginning of the "mathematization of human reasoning" (Zadeh, 1975, p. 425).

While Zadeh was developing a new non-discrete language, Meinhart (1966) proposed an insightful convergence of machine simulation of human processes such as thinking and cognition, with management thought. Meinhart identifies the controversy and difficulty in a machine or system being able to "think" (1966, p. 295). Juxtaposing Meinhart with Agrawal et al. (2018a, 2018b), Agrawal (2018), and Ng (2016), it's perhaps not totally unexpected that the needle on what AI techniques can do and not do in the context of abstract problem classes, hasn't moved nearly as fast and as far as advances in classification, prediction and other simpler use-cases.

From Fuzzy Sets to First Signs of Algorithmic Intelligence

While there has been asymmetric capability advancement over the past more than 70 years, there is no question around the broadening of utility, applicability, and study of these algorithmic concepts, e.g. thousands of publications within engineering, computer science and mathematics (Zimmermann, 1993). Since the late 1990s and early 2000s, a clearer connection became visible between the concepts of predictive statistics, analytics and business decisions. As presented in Agrawal (2018) and Brynjolfsson and McAfee (2014), drivers of this were ever-faster and more powerful computer hardware, which helped drive down the cost of computational processing, making the use of these analytic tools more accessible to now, virtually all *hypercomputing* taking place in the public/private clouds of *hyperscalers* such as Google Cloud, Microsoft Azure, and Amazon Web Services to name a few. Zimmermann provides a comprehensive set of models and approaches on how we moved away from classical (crisp) methods and information to fuzzy sets, and into more meaningful uses in "practical decision making" (1993, p. viii), to data-based decision support systems and knowledge-based expert systems.

To help connect with more current capabilities, the move away from discrete values allows mathematical logic to analyze and provide meaningful answers to conceptually complex humanistic questions: nomenclatures such as, e.g. *always yes*, *mostly yes*, or *sometimes yes* allow for the creation of comparable linguistic sets. A garden can be described as beautiful along with other characteristics such as its location, composition, or size, and then compared to a new undefined garden, to try and predict whether this new garden is beautiful or not. Fuzzy logic applied to fuzzy things to produce outputs allows computers to *compute with words*, and is considered one of the most important contributions in the field, as indicated by Zadeh's over 281,000 citations[10]. The progression from this to probabilistic inferential approaches such as machine learning, allows systems to learn from outputs and data sets, and construct linguistic characteristics from known data to try and compute or predict something which is unknown, thereby self-learning and self-refining the rules by which tasks such as classification takes place.

My chronological survey is aided by coverage and analyses provided in Negnevitsky (1997), Luger and Stubblefield (1998), Brynjolfsson and McAfee (2014) and others. Negnevitsky's comprehensive review achieves a valuable grouping of relevant topics and advances into logical and rationalized time-based periods, such as 1943-1956 which he terms the dark ages, and the early 1970s to mid 1980s, and so on (1997). While asymmetrically spaced,

---

[10] https://scholar.google.ca/citations?user=S6H-0RAAAAAJ&hl=en&oi=ao, retrieved Feb 22, 2025.

there is a coupling and general correlation between technological advancement and the passage of time. Notably, Negnevitsky suggests several "unfulfilled promises" (p. 5) in the 1960s and 1970s eventually led the important realization that restrictions had to be placed around the problem types intelligent machines were tasked with solving. This seems to foreshadow notions of limitations of AI techniques. He further posits a paradigm "shift from general-purpose, knowledge-sparse weak methods to domain-specific, knowledge-intensive techniques" (p. 7), seemingly keys to current deep learning systems, loosely laying the groundwork for complementary solutions composed of different technologies, each designed and uniquely equipped to handle different aspects of the complexity posed by strategic problems.

Building upon Negnevitsky (1997), in Table 2, I've added available information pertaining to the year certain key events occurred, for example, LISP development in 1958 (McCarthy, 1958). I've also continued his approach beyond the late 1990s to present time, by providing a list of important events. These include developments within AI technologies and techniques, but more importantly, I also include events relating to suboptimal outcomes and trade-offs. It is helpful to be able to see the sequence and continuity from the birth of this area to the present.

Table 2

An overview of Key Events in the History of AI and Knowledge Engineering, adapted and expanded from Negnevitsky (1997)

Milestone (Time Period)	**Key Events**
Birth of AI as a Concept (1943 - 1956)	McCulloch and Pitts, *A Logical Calculus of the Ideas Immanent in Nervous Activity*, 1943

Milestone (Time Period)	**Key Events**
	Bush, V, 1945
	Turing A., *Computing Machinery and Intelligence*, 1950
	Neumann J. V., The Electronic Numerical Integrator and Calculator project, 1944-1946
	Shannon C., *Programming a Computer for Playing Chess*, 1950
	The Dartmouth College summer workshop on machine intelligence, artificial neural nets and automata theory, 1956
Beginnings of Algorithmic Intelligence (1956 - the late 1960s)	McCarthy J., LISP, 1958
	Newell and Simon, The General Problem Solver (GPR) project, 1959
	Newell and Simon, Human Problem Solving, 1972
	Minsky M., A Framework for Representing Knowledge, 1975
Failures of Algorithmic Intelligence (the late 1960s-the early 1970s)	Cook, The Complexity of Theorem Proving Procedures, 1971
	Karp, Reducibility Among Combinatorial Problems, 1972
	The Lighthill report, 1971
First Rule-Based Deterministic Systems (early 1970s - mid-1980s)	DENDRAL (Feigenbaum, Buchaman and Leberberg, Stanford University), 1965
	MYCIN (Feigenbaum and Shortliffe, Stanford University), 1972
	PROSPECTOR (The Stanford Research Institute), 1970s
	PROLOG - a logic programming language (Colmerauer, Roussel and Kowalski, France), 1972
	EMYCIN (Stanford University), 1979
	Donald Waterman, A Guide to Expert Systems, 1986

Milestone (Time Period)	Key Events
Computing with Words (1960s – 1990s)	Zadeh L., Fuzzy Sets, 1965
	Zadeh L., Fuzzy Algorithms, 1969
	Medical Diagnostic Decision Support (MDDS)
	Mandani, Application of Fuzzy Logic to Approximate Reasoning Using Linguistic Synthesis, 1977
	Sugeno, Fuzzy Theory, 1983
	Sendal Subway System (Hitachi, Japan), 1986
	Negoita C., Expert Systems and Fuzzy Systems, 1985
	Japanese "fuzzy" consumer products (dishwashers, washing machines, air conditioners, television sets, copiers), 1990
	The First IEEE International Conference on Fuzzy Systems, 1992
	Kosko B., Neural Networks and Fuzzy Systems, 1992
	Kosko B., Fuzzy Thinking, 1993
	Cox E., The Fuzzy Systems Handbook, 1994
	Yager and Zadeh, Fuzzy Sets, Neural Networks and Soft Computing, 1994
	Fuzzy Logic, MATLAB Application Toolbox (The MathWorks, Inc.), 1994
	Kosko B., Fuzzy Engineering, 1996
	Zadeh L., Computing with Words - A Paradigm Shift, 1996
Artificial Neural Networks (1965 – 1990s)	Hopfield, Neural Networks and Physical Systems with Emergent Collective Computational Abilities, 1982
	Kohonen, Self-Organized Formation of Topologically Correct Feature Maps, 1982
	Rumelhart and McClelland, Parallel Distributed Processing,

Milestone (Time Period)	Key Events
	1986
	The First IEEE International Conference on Neural Networks, 1987
	Neural Network, MATLAB Application Toolbox (The MathWorks, Inc.), 1992
	Haykin S., Neural Networks, 1994
Probabilistic Inferential Systems and Self-Refining Artificial Intelligence Systems (1990s - Present)	Clinical Decision Support Systems (CDSS)
	COMPAS, 1996
	Deep Blue (IBM), 1996
	Watson (IBM), 2011
	AlphaGo (DeepMind, Google), 2014
	Brynjolfsson and McAfee, The second machine age: Work, progress, and prosperity in a time of brilliant technologies, 2014
	Domingos, The master algorithm: how the quest for the ultimate learning machine will remake our world, 2015
	Others
New Unifying Models addressing Trade-Offs and Issues (2000s to Present)	Gunning (DARPA), Explainable Artificial Intelligence (XAI), 2017
	Bellamy et al. AIF360, 2018
	Arrieta et al. (2019), Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI, 2019
Conversational and Generative Technologies (Mid 2010s to Present)	Generative Adversarial Networks (GANs) (Goodfellow et al.), 2014

Milestone (Time Period)	Key Events
	GPT-1 (OpenAI), 2018
	Retrieval-Augmented Generation (RAG) (Lewis et al.), 2020
	DALL-E (OpenAI), 2021
	LangChain (Harrison Chas), 2022
	ChatGPT (OpenAI), 2022
	Midjourney (Midjourney). 2022
	NotionAI (Notion Labs),2023
	Auto-GPT (Toran Richards), 2023
	DeepSeek (DeepSeek), 2023

The Shift from Math and Science to Business Applications

The period from the late 1990s into the 21st century witnessed a remarkable increase in both the actual and perceived development, mention, and application of machine learning and other AI technologies (Brynjolfsson & McAfee, 2014; Domingos, 2015; Negnevitsky, 1997; others). Without going into details here, I am of the view that the widespread use of the term AI is mostly inaccurate—what they are describing are predominantly long-existent rule-based predictive statistical models and applications. I raise this only as a point of correctness. For the purposes of this study, it's sufficient to just delineate rule-based non-self-refining approaches and algorithms, versus probabilistic-based self-refining approaches algorithms and systems.

To that end, several literary works have looked at specific applications and approaches of the relevant concepts of business intelligence (BI), ML, MLAs, expert systems (ES), and various AI techniques in business contexts. Others have also provided more comprehensive surveys of these types of applications and systems across time, technology, or problem type. Table 3 below summarizes many important developments, with some specific examples. This view helps to contextualize *first-moves*, and creates a picture of approaches building upon previous ones.

Table 3

Chronological Survey of Specific Applications of AI Techniques Applied to Business Problems and Contexts

BI/ML/AI/Approach	Application or Business Context (based on Abstract)	Researchers / Entity (Year)
Data Mining	Inform the information systems personnel about the role of ML techniques in business data mining.	Bose and Mahapatra (2001)
Prediction	Predicting customer value from perspectives of product attractiveness and marketing strategy.	Chan and Cho (2010)
Artificial Neural Networks	Application of ANN for stock market predictions: A review of literature.	Dase and Pawar (2010)
Prediction	Predicting customer value from perspectives of product attractiveness and marketing strategy.	Chan et al. (2011)
Decision Support	DSS for integrating manufacturing and product design into the reconfiguration of supply chain networks.	Kristianto et al. (2012)
Fuzzy Expert System	Design of fuzzy ES for microarray data classification using a novel Genetic Swarm Algorithm.	Kumar et al. (2012)
Business Intelligence, Knowledge Management	Review of the umbrella concept of BI. Undertaking to develop or create such	Rouhani et al. (2012)

BI/ML/AI/Approach	Application or Business Context (based on Abstract)	Researchers / Entity (Year)
	intelligence. Classify different definitions and approaches in BI into 56 groupings and three approaches: managerial, technical and enabler. Study the use of technologies and processes to improve overall decision making as part of strategy formulation.	
Prediction, ES, DSS, Hybrid	AI-based systems applied in industrial marketing: An historical overview, current and future insights.	Martinez-Lopez and Casillas (2013)
Decision Support System	DSS for market-driven product positioning and design.	Lei and Moon (2015)
Prediction, Expert System	Rule-based ES for predicting consumer preference in new product development.	Yang et al. (2016)
Fuzzy Logic	Max-min fuzzy programming approach for compromise farming: a case study.	Kumar and Roy (2018)
Machine Learning, Machine Learning Algorithms	Value-based pricing in digital platforms: A machine learning approach to signaling beyond core product attributes in cross-platform settings Data-Driven Analytics Leveraging Artificial Intelligence in the Era of COVID-19: An Insightful Review of Recent Developments	Christen et al. (2022) Majeed and Hwang (2022)
Artificial Neural Networks, Big Data, Deep Learning	Automatic Business Location Selection through Particle Swarm Optimization and Neural Network	Hinton (2012, 2015) Luo et al. (2022) Patriarca et al. (2022)

BI/ML/AI/Approach	Application or Business Context (based on Abstract)	Researchers / Entity (Year)
	Democratizing business intelligence and machine learning for air traffic management safety	
Large Language Models, Generative Models	Natural language conversational engines, generative tools to create images, video, text via user-prompts and other inputs, contextual and meta inputs to predict and generate optimal outputs	Goodfellow et al. 2014 OpenAI, 2018 OpenAI, 2022

The above table illustrates the significant and diverse work undertaken in various organizational contexts. As the use of these technologies continued to be developed and applied to very diverse problems or goals, the application set kept, and keeps, expanding. Commonly observed examples include: learning and playing board games, determining what is contained within radiology images and video with greater accuracy than humans alone, 2D and 3D facial image construction, vast generative use-cases, and so on.

Also of interest in reviewing the above literature, is that most if not all of the core ML algorithms and AI technique building blocks have been around for many years or even decades. This is quite remarkable given how frequent we hear about "a new AI" as if something truly new has been created. What this tells us is that the application of these techniques seems to lag the capabilities of the underlying technologies themselves. If this were not the case—if the technologies were lagging industry or other application needs—then I suggest we'd expect to see far more developments and publications in the literature and theory, and less so in their uses or applications.

Advent of Generative Technologies and Models

On Nov 30, 2022, an organization called OpenAI (openai.com) publicly launched and offered easy website-based access to their conversational platform ChatGPT[11]. The value-add basis of ChatGPT was two-fold: 1) it could conversationally interact with humans using natural language processing (NLP) and natural language understanding (NLU), and 2) the amount of data upon which OpenAI had trained ChatGPT's underlying Large Language Model(s) increased the response quality and rate. As a simplified explanation, ChatGPT is a LLM-based chatbot that enables users to steer and refine a conversation towards a desired length, format, style, level of detail, and language, via widely available and free-of-charge access[12].

While the concept of LLMs is not new and traces its roots to MIT in the 1960s and perhaps earlier[13], ChatGPT's ability to undertake real-time conversational prose on seemingly endless topics via a prompt, was new and seemed to make this technology immeasurably more and more-easily accessible. The highly touted launch of ChatGPT, and the seemingly countless other companies and products that have emerged since then, seem to have driven a sea change in terms of access and utilization of conversational and generative technologies and models. However, given that much of the underlying approaches and technologies, e.g. deep learning (DL), artificial neural networks (ANNs), GANs, have been around for some time, what changed? What drove such rapid innovation with seemingly substantive speed gains?

11 https://openai.com/chatgpt/, retrieved Dec 10, 2022.
12 https://en.wikipedia.org/wiki/ChatGPT, retrieved Nov 7, 2023.
13 https://www.dataversity.net/a-brief-history-of-large-language-models/, retrieved Jul 20, 2024.

The Prompt Revolution, or Evolution?

Even though these concepts are not new, the advent and wide accessibility of conversational and generative technologies has been described by some as revolutionary (Haleem et al., 2023; García-Peñalvo, 2023; Gordijn & Have, 2023; Kalla et al., 2023; Wardat et al., 2023), and others as more evolutionary (Bektaş et al., 2024; Gordijn & Have, 2023; Wang et al., 2023). I suggest that the introduction of *the input prompt*, that users can literally interact with and input their queries, while interesting, is of less importance and impact than the intense computational resources that have been applied in the training of LLMS and other models and approaches, that augment these models, e.g. GAN and RAG. Advancements by OpenAI and other organizations using LLMs are built upon a foundation of specialized computational processing units of micro-chips and micro-chipsets. Within time-limited sets, LLMs and other ML models have been advanced and trained on immense amounts of literary texts and other data of all kinds and from all available sources[14]. Just a few examples of these training data sets include written prose, voice, video, images, art, software, hardware, and so on. Advancements of AI techniques into areas of visual imagery and motion, as well as sound and speech present appealing opportunities, but not without associated risks of misuse. Pre-existing techniques, such as retrieval augmentation rooted in search, and adversarial networks, have been updated and used in new models, in an effort to incrementally gain in trade-off dimensions such as accuracy and reliability. Products such as ChatGPT demonstrated to vast audiences and users that computers could converse with humans using human language, but in fact, researchers and practitioners

[14] https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research#Internet, retrieved Nov 2, 2024.

have been able to converse with computational systems via APIs and other computational methods long before the emergence of these conversational prompts.

Increases in computational speed and storage[15] are the critical factors that have facilitated the rapid expansion of machine learning models and approaches into the area mentioned above. Immense amounts of data—language and otherwise—has been stored and used to computationally train LLMs and other techniques. Conversational models are now not only able to linguistically process and understand what a user is stating or asking, but also assemble a response that seems to make linguistic or visual *sense*. Vector databases and other constructs at the root of how LLMs operate, help achieve linguistic, grammatical and composition precision. Basically, these storage structures are used to calculate, formulate and decrease local distance between query word strings, word weightings, and response word strings. I point this out because, while the response *may* be accurate or inaccurate, the output *will* likely appear to makes sense due to the correct stringing together of words, or sticking together of pixels, sounds, or other forms of output. Kalla et al. (2023) support the notion that the revolution is more about how we interact with these technologies, than perhaps the technologies themselves. I caution the reader to not equate precision or sense-making, with correctness.

Weaknesses of Conversational Chatbots

Currently and as architected for the most part, there are several weaknesses of conversational chatbots and other *only* LLM-based tools such as ChatGPT. To begin with, they are trained on existing data or information which by definition has a finite timespan of existence.

---

[15] https://developer.nvidia.com/blog/nvidia-blackwell-platform-sets-new-llm-inference-records-in-mlperf-inference-v4-1/, retrieved Aug 31, 2024.

While the models can no doubt be trained and retrained to expand the set of training data, and even with the advancement of computational, storage and network infrastructure, there is a real and measurable latency each time model(s) must be retrained. As the amount of data grows in magnitude, the time to train increases. In other words, any increase in training data set, while keeping other variables such as CPU and memory consistent, leads to an increase in time.

In addition to the issue of data-to-speed, LLMs on their own cannot formulate an *original answer*. For example, if a user asks or requests a response outside the timeframe of the training data set or on a topic that is not contained within same, the system is unable to respond. Despite this weakness, many LLM-based tools have been algorithmically augmented and programmed to deliver responses where none may exist. This has been termed by many (Ji et al., 2022), as forms of hallucination. Early users of LLM-based conversational tools tested this by literally lying to the system and were lied to in response. I posit that while this may be a form of hallucination; these models simply do not know better or more accurately what a response should be. These are not truth engines, by reference or computation, nor have they been designed to maximize or optimize accuracy of the content in their response. They have been designed to rapidly respond with word strings, or code, or whatever other response is being queried and the LLM is able to respond with, with a response that make linguistic and grammatical sense.

Connecting Strategic Decision-Making Problems to Computational Approaches

Early Learnings from Machine Learning Applied to Strategic Decision-Making

This section will review literature covering AI techniques more directly applied to organization strategic problems. A historical review allows for the consideration of different approaches and techniques and their suitability to strategic decision-making.

Spangler (1991) gives us a particularly interesting and robust analysis and approach into the use of AI and related technologies to strategy formulation. He concludes that the use of AI to strategy formulation is grounded in theory of problem solving itself and not necessarily the type of problem. I think Spangler correctly focuses on the early stages of the strategy formulation process as the nebulous and particularly troublesome steps in the process, in terms of using DSSs, ESs and other forms AI. He proposes several specific reasons why the use of AI technologies in support of organizational strategic planning has been slower than with other uses such as in medicine or accounting, including the less defined nature of strategic problems, the need for communication, and the creative and abstract nature of strategy formulation. Through this, he underscores the notion of humans being particularly adept at undertaking complex processes and making difficult decisions (pp. 149-150). He cites other works which support the conclusions that the bottleneck or unsolved problems are less software and hardware, and more the actual lack of understanding of [strategic] decision making (p. 150). Spangler's survey of the various DSSs and related AI systems to date presents a gap in the area of cognitive modelling, which I will loosely define as essentially the process of problem solving or the cognitive and expert thought processes in strategic planning. Spangler concludes with a series of provocative and profoundly impactful questions, focused on what he asserts as the heart of the issues vis-à-vis AI applied to strategy formulation. Connect with my loose definition, the problem *becomes* about how to solve problems, rather than solve a particular problem.

A general framework and functioning of a distributed strategic decision support System (DSDSS) is described in Pinson et al. (1997), whereby advances in the areas of distributed decision-making and AI are leveraged for improved results. This is an evolution and improvement from the DSS model proposed in Spangler (1991) and elsewhere. At its core are two features: 1) user intervention during solution formation, and 2) specific strategic and domain knowledge are distributed and communicated between agents. Due to the complex and unstructured nature of decision making, Pinson et al. posit that there is no algorithmic solution. They take the reader through a proposed model of an architecture with the existing tools and technologies. While the model is interesting and certainly worthy of inclusion and study, there already seems to be contradictory work in that certain aspects of the decision and strategic decision-making processes have indeed been algorithmized to varying degrees (Agrawal, 2018; Albescu et al., 2008; Brynjolfsson & McAfee, 2014; Rouhani, 2012).

The learnings from Spangler (1991) and Pinson et al. (1997) illustrate that a single solution or approach in applying AI to strategic problems is likely insufficient. The nature of the problem made it difficult for these attempts in the 1990s to try and find alignment between one of the AI approaches and strategy formulation. However, in McKelvey (2000) we find a rather interesting and provocative model to align and optimizing return from strategic decisions as a potential target area for where AI researchers should focus—the *economic rent* as he calls it is (p.1). He successfully puts forward the argument that the critical driver of being able to generate higher economic rent is distributed intelligence (DI) or “speeding up rates of intrafirm change” (p. 1). This is similar to Eisenhardt (1999) in that speed of decision, action and change are differentiating and indeed competitive advantages of certain organizations over others. Building on Hamel and Prahalad (1994), Porter (1996) and Prusak (1996), McKelvey (2000) develops the definition of completive advantage as a firm’s ability to learn, adapt and move into new directions and opportunities compared to its competitors. Moreover, human capital and the knowledge contained with it must be shared and available via networks throughout the firm. While his comments on computer DI systems is dated and no longer deemed to be factually correct, it is worth noting that he and others did see the connection and potential for computer systems and networks as players within any DI system design and the value it offers through network development and speed. This has implications for subsequent research and approaches such as the series, parallel or hybrid computational systems and models.

More Recent Approaches to Applying ML to Strategic Decision Making

Li (2000) begins what becomes the first of a multi-publication effort to develop and empirically test a hybrid intelligent system for developing marketing strategy. He explores a system designed to align marketing strategy needs with available technologies and approaches such as expert systems, fuzzy logic, decision support systems, and artificial neural networks. Although the core elements of Li's model are not necessarily new, Li attempts to specifically address issues which were driving managerial dissatisfaction with existing systems and approaches. He posits that no individual technology on its own is up to the task, and he proposes a hybrid intelligent system:

> Through integrating the advantages of diverse support techniques and technologies into one hybrid system, a more powerful system than each of its components standing alone can be created, exploiting the synergy of the component parts ... the motivation for developing a hybrid intelligent system is to integrate the powers of different support techniques and technologies, such as ESs, fuzzy logic, ANNs and DSSs to support the development of marketing strategy where the required conditions are not satisfactorily met by individual techniques and technologies. (Li, 2000, p. 398)

Liao (2005a) is a literature review of *Expert System Methodologies and Applications* from 1995 to 2004. As we've seen, expert systems are a subset of AI and have broad applicability where we can model and codify entire operational frameworks. This paper reviews some 166 articles from 78 academic journals and categorizes expert systems into eleven categories. Liao further breaks down the academic literature within each category and highlights key findings and steps forward in the technologies and their applications to various industries from business to agriculture to genetics, as well as functional problems which span industries, like planning, decision support, and strategic management. Many of the functional and industry-specific applications of ESs have implications for the topic of this paper. In particular, I'm interested in neural networks, fuzzy expert systems, and code-based reasoning and modelling. Liao correctly point out that there is overlap between different methodologies and applications (p. 98), and this is a key finding as we consider how a set of AI systems with or without human intervention could be developed and applied to the complex problem of strategy formulation, which simultaneously requires multiple decision types including but not limited to language processing, learning, planning, forecasting, predicting, modeling, judgment and so on. In his second literature review, Liao (2005b) overviews 546 articles from 1995 to 2003 and classifies *Technology Management Methodologies and Applications* into eight categories, of which we're most interested in the categories of artificial intelligence/expert systems, and statistics. This paper provides a macro review of AI/ES within technology management and application and is helpful in considering what if any other methodologies or applications may be useful for the topic at hand. I will extract from this paper, the specific application of different methodologies to specific industries and problem types. The key conclusion is that technology management methodologies typically “develop towards expert orientation” (p. 389), while their applications are towards problem types.

Albescu et al. (2008) provide insights into an approach on bringing together what they call structured data management or Business Intelligence (BI), with unstructured content management or Knowledge Management (KM), to posit that hybrid software, technologies and systems which bring together BI and KM are real technological support for strategic management—called Competitive Intelligence. Of interest to this paper is the explicit inclusion of information and other metrics about competitors, customers and other external factors, and its conversion into relevant and actionable intelligence. In order words, they are really talking about systems and processes which convert information into intelligence or knowledge.

Another literature review focused on AI is presented by Oke (2008). I note that his opening sentence mentions “two decades” (p. 535), but his citations seem to cover the briefer period from 1992 to 2007. Nonetheless, some of the highlights research and findings are of value to this study, as he: 1) provides sixteen categories of AI application, 2) distinguishes specific vs general applications of AI, and 3) analyses and discusses the concept of intelligence. While Oke identifies gaps in his survey of this vast and complication topic, his paper covers a broad set of applicable and important work in the area with 164 citations. Looking deeper into the sixteen categories of AI applications (Becker et al., 2000; Chen & Van Beek, 2001; Hong, 2001; Peng & Zhang, 2007; Singer et al., 2000; Stone et al., 2001; Wang et al., 2007; Zhou et al., 2007; all as cited in Oke, 2008), Oke (2008) provides a view of the various areas of AI as of that timeframe highlighting the relationship amongst the diverse areas of: (1) belief revision, (2) expert systems, (3) artificial life, (4) reasoning, (5) data mining, (6) genetic algorithms, (7) theory of computation, (8) theorem proving, (9) programming, (10) distributed AI, (11) knowledge representation, (12) natural language understanding, (13) machine learning, (14) systems, (15) neural networks, and (16) constraint satisfaction. I don't agree with some of the categories Oke brings forward, because they are simply invalid or not truly a distinct category of AI—we mustn't forget that almost two decades have passed since Oke (2008) and even longer for most of the underlying works he cites. For example, virtually all areas of AI rely on some form of machine learning, and while the latter is encompassed within the former, it doesn't make sense to define ML as a distinct category within AI—ML is used within data mining, genetic algorithms, natural language understanding, neural networks and so on.

Of interest to this paper and topic, are the particular areas in Oke (2008) which cover reasoning, expert systems, knowledge representation, and natural language understanding. Specifically, Oke underscores that a lack of "comparison and integration" (p. 561) exists in the research and application of AI technologies across areas and methods. As we have seen since Oke's work and candid comments on the lack of integration across the AI fields, more modern applications of AI technologies in diverse fields has indeed improved on this amalgamation and uses of multiple AI fields or technologies to gain greater efficacy in the application of AI to business problems. In other words, researchers and organizations are now creating and applying models in sequence and in parallel (Lauterbach & Bonime-Blanc, 2016; Brynjolfsson & McAfee, 2014; Chan et al., 2010; Min, 2010, Mosele, 2018; Namaki, 2018), and more obviously in the form of hybrid systems (Li, 2000; Li & Li, 2009; Li et al., 2011).

A complicated hybrid system composed of multiple software processes, while incorporating human judgment within the context of marketing strategy formulation, is presented in Li and Li (2009). This is conceptually similar to the model proposed in Pinson et al. (1997). Li and Li empirically study six managers using their hybrid approach, resulting in improved "processes and outcomes of strategy formulation" (2009, p. 5557). By bringing together and incorporating various decision support systems (DSS), expert systems (ESs), etc., they attempt to benefit from the strength of each while "avoiding their specific disadvantages" (p. 5557). They hypothesize that they can improve the strategy formulation process through the hybridization of various accepted techniques: "human judgment, AHP, Monte Carlo simulation, and a fuzzy expert system for strategic marketing planning" (p. 5558). While their study involved a small sample of six experienced managers, they measured improvements to the process and outputs of marketing strategy formulation:

> The findings show that the hybrid approach is efficient in improving the speed of strategy development and effective in coupling strategic analysis with human judgement, improving strategy-making performance, helping understand strategic factors, helping strategic thinking, building up strategy-making confidence, and improving the output and quality of strategic decision-making. Written feedback and comments also suggest that the hybrid approach can help improve the process and outcomes of strategic decision-making. (Li & Li, 2009, p. 5562)

These are promising results and indicate that the direction taken by Li and Li (2009) is worth exploring more widely across other types of organizational decisions and strategy formulations, such as staffing, pricing, operations, product development, and others. The specific inclusion of human judgment, intuition and creativity into the hybrid process is worth noting, and directly refutes a popular misperception that machines are "taking over". I would suggest that human knowledge and judgment play a critical role in the development of optimal decisions and strategies, and the ability to algorithmize specific aspects of these important processes allows for the creation and enhancement of competitive advantage in new and exciting ways. In radiology and other areas of medicine, for example, machines and algorithms are already able to reliably scan and propose diagnoses from digital images such as x-rays and CT-scans, and then a trained doctor is able to intervene and make the final diagnosis, via a similar hybrid or augmented machine-human system (Hinton, 2016, Syed & Zoga, 2018; Trajtenberg, 2018).

In a subsequent paper, Li and Li (2010) present, discuss and evaluate the "linking of web-based decision aid techniques and technologies for international marketing planning" (p. 7094). They propose, *inter alia*, a framework for international marketing planning processes using a "Web-enabled hybrid intelligent software system combining Web-based knowledge automation, fuzzy rules and on-line databases for international marketing decision-making" (p. 7094). In a subsequent publication, Li et al. (2011) present a functioning system modelled upon their 2010 proposal: a web-based hybrid knowledge automation system created for the purpose of formulating digital marketing strategies. I need not get into an analysis of their details of system design nor its construction, but the key insight from Li et al. (2011) is the achievement of the application of this hybrid system with various digital marketing models, via computerization, adaptation and extension. Li et al. (2011) found: 1) gained efficiencies and effectiveness, 2) aided decision-making, 3) coupling of "web-enabled knowledge automation with human judgment and creativity" (p. 10612) improved ability to handle uncertainty, 5) greater levels of creativity via alternative generation, and 6) improved planning quality and enhanced confidence and satisfaction. Most if not all of these empirical findings are important for this study, in that they help to address several key issues and points of failure when considering any sort of technologization or automation of complicated tasks such as strategy formulation.

Other proposed models of human-machine collaboration and interaction based on problem type and output sought include Gerbert (2018): more advanced and machine-focused models, which move from the most basic augmentation, to true human-machine collaboration (something like we saw in Li and Li, 2009), to “hyperlearning” where computer systems are trained to operate complete virtual environments.

Lauterbach and Bonime-Blanc (2016) brings the research and technologies to a state where AI is considered ubiquitous and on the verge of being a background player in ‘virtually everything we do and touch’. They lay out an optimistic yet cautionary argument of what different classes of AI are, what they can do, and what leaders need to get their heads around in order to survive or thrive in the new business and competitive reality. The clear delineation of Narrow AI versus Deep AI or AGI is important and helpful in this analysis.

Continuing with this concept of AGI and the apparent need for a connected whole made up of hybrid parts, Bettis and Hu (2017) give us a very insightful and holistic view of many inter-connected concepts such human rationality, decision-making, and AI. As the technology advances, researchers in fields from economics, to psychology, to computer sciences and engineering, seem to be an inflection point—what is the best approach going forward to more accurately model human brain functions such as learning and judgment. While we’ve seen this theme before (Agrawal et al., 2018a; Albescu et al., 2008; Brezillon, 1999; Brynjolfsson & McAfee, 2014), Bettis and Hu take a slightly different approach around their central thrust that behavioural strategy and strategic management can benefit and learn from computational complexity and AI. Thinking of the brain as a form of computer, they posit that “human cognition can be modelled in terms of computation” (p. 2). Problems can be classified into different categories, one simple division being *tractable* vs *intractable*. As we've seen with the abilities and limits of AI, Bettis and Hu (2017) state that these categories can be applied to human cognition and managerial intelligence, and to strategic decision-making within organizations. The implications for this paper are: 1) AI approximates solutions to intractable problems through heuristics, 2) computers can only solve well-structured problems, and 3) human rationality is bounded and computational. Where they take these conclusions is most interesting and notionally aligned with others (Agrawal et al., 2018a; Brezillon, 1999; Mosele, 2018), by connecting the concepts of organizational decision-making, intractable problems and human processing time (p. 14). To understand the nature and complexity of such problems:

> For organizationally important decision making (e.g. strategic decisions) based on rational analytical technologies, the total time required for an organization to organize the decision making process, gather the required information, perform the necessary analytical and judgment tasks to arrive at a logically justifiable analysis (sufficiently rigorous and comprehensive), and make the decision is long enough to preclude the possibility of arriving at a solution before the decision problem has changed in important ways or has become irrelevant (Bettis, 2017, as cited in Bettis & Hu, 2017, p. 15).

This is of value to this study in that a picture seems to be emerging across different types of research, around what are the key problems left to be solved in context of AI techniques and strategy formulation, why these are key next steps, and how they might be solved. Bettis and Hu (2017) conclude by asking if there are problem types computers cannot solve, or does the problem itself end up being altered in order to solve it.

Aristodemou and Tietze (2018) is a literature review on the use and application of AI on intellectual property (IP) data, which they call "patent analytics" or "Intellectual Property

Analytics" (p. 37). It is the study and analysis of data contained within patent and intellectual properties, to be used to try and uncover or create competitive advantage, by analysing and learning what IP others are creating and developing. The end goal of this exercise is to assist with, and improve, decision making. I'm not so interested here on their approach to gather and splice articles in their review, but more so in the application of deep learning and its approaches to this type of problem. I'm also interested in their use of different disciplines such as mathematics, statistics, programming and operations research. Their key findings relevant to this study relate to areas of AI application that render the most IP development, including: knowledge management, technology management, economic value of IP/patents, and the hybrid of extraction and effective management of information. They also point out that these categories are permeable or perhaps overlapping. Yan and Han (2018) also undertook an analysis of qualitative and quantitative patent registration data to uncover competitors' technological strategies, to try and establish a competitive advantage for one's own organization. They underscore the value and power of analysing data using big data analysis and machine learning tools. More prescriptively, they suggest organization move from "descriptive and diagnostic analysis" to "predictive and strategic analysis" (p. 2).

In his opening paragraph called "The Problem", Namaki (2018) excellently connects the current state of AI and where it is rapidly moving to:

> Computing equipment capable of what one may term partial and quasi-intelligent behaviour, commonly referred to as Artificial Intelligence (AI), is assuming a key role in business. The probability is high that this Artificial Intelligence (AI) will lead to a fundamental change in the process of business strategy formulation as much as the very contents of this strategic behaviour. Product and market strategies and the resultant competitive behaviour will, more

> likely than not, be the outcome of those artificial intelligence processes and reiterations. A start is made and one can observe substantial progress in this direction. Who has done it and is there a conceptual framework behind this strategic behaviour? (Namaki, 2018, p. 77)

While Namaki (2018) provides a well-constructed definition of AI and comments on its current strengths and weaknesses vis-à-vis specific strategy formulation in his examples, my analysis here is more interested in his geographically and industry-diverse look at what industry is undertaking through his seven case studies. What is even more interesting is his bold assertion that "recent strategic moves by several corporations Worldwide have been in response to AI events or outcomes" (p. 78). This directly supports and goes to my research question. While we are constantly exposed to readily available applications of AI such as engines of suggestion or rudimentary prediction, some organizations such as The North Atlantic Treaty Organization (NATO) are actively engaged in what Namaki calls "future" instruments for future markets (p. 80). This latter example is the most advanced place in Namaki's new conceptual framework, which separates the environment where the AI work is undertaken from the market where the "function delivering instruments are exchanged" (p. 80). This 2x2 matrix creates a very interesting field for organizations to consider for their internal AI investments and where its outputs will be delivered or exchanged.

Do we seek a Simulator or Emulator?

Recalling the distinctions above between rule-based deterministic systems and probabilistic-based inferential systems, a similar distinction can be made between algorithmic agents that simulate the models and processes of a strategist versus agents that emulate strategists. The 'strategy simulator is intended to replicate the outputs of models of strategic decision-making' (M. Moldoveanu, personal communication, July 12, 2022). In other words, a simulator replicates the outputs of models to produce the same output that a human actor would produce, given the same inputs, variables, rules, approaches and resources. An emulator on the other hand, has the potential to learn, self-refine and indeed, replace the human strategist in part or in whole. So, what we're seeking is in fact an intelligent algorithmic emulator.

We've seen that even contemplating such systems is predicated on yet-to-be realized gains in measures such as speed and judgment (Agrawal et al., 2018a), but also on accuracy. The emulator must not only *learn* and self-refine to a level of performance that a human strategist chooses to defer to the emulator, but in the context of this study, capture the myriad of intangible human judgement and experiential variables that human strategists purport to capture in their analysis and decision making. I note that there is evidence to the contrary, and human strategists also seem to fall short in their ability to fully articulate, analyze, and solve NP-class organization problems (Moldoveanu, 2009, Moldoveanu & Leclerc, 2015). Thus, in both the human and algorithmic context, there are areas of improvement in many facets of the strategy-making process. Human simplification of the problem leading to suboptimal solutions, while problematic, persists because it's undertaken by strategists, managers and teams playing with similar toolkits. Building upon Moldoveanu's concept of trade-off dimensions and the trade-offs between them (2021), I posit that an expanded set of variables/dimensions must be considered and included into the strategy-making process. This allows for an improved and more comprehensive solution, and assessment tool as to the quality of solution. This is also critical as we envision and try to design algorithmic emulators.

Algorithms Applied to Organization Strategy are Fraught with Trade-offs

We've already seen that there are known and unknown gaps in the current capabilities of even the most advanced algorithmic systems and technologies, to truly enter the domain of human decision-making and complex tasks such as strategy formulation. The push towards ever-greater involvement and leveraging of technology has caused organizations to ignore or overlook downside risk or negative impacts of such technologies (Babcock et al., 2019; Canhoto & Clear. 2019; Gerke et al., 2020; Manheim & Kaplan, 2019; Piano, 2020; Szychter et al., 2018; The Royal Society, 2019). According to Moldoveanu (M. Moldoveanu, personal communication, July 11, 2021), a novel model of trade-offs among several dimensions explicated from the literature, shows how 'pairwise or n-wise, trade-offs between/among them shed light on a host of problems/quandaries/predicaments and difficulties that organizations have run into.' The key take-away is that the deployment of AI techniques in the context of organization strategic problems comes at a cost, and there is no free lunch in terms of potential gains in speed or accuracy.

Of particular interest and worth noting is that the explicit or implicit trade-offs made by organizations and managers is in a way, similar to Moldoveanu's finding that 'managers prefer to engage with and solve P-type problems (canonically easy/ier) over NP-type problems (hard ones)' (2009, p. 737). In that analysis, Moldoveanu posits that managers switch or alter—in fact, simplify—the problem statement or representation in favour of their preference for less complex P-hard problems. In this study, the literature shows that managers explicitly or implicitly overlook trade-off dimensions and binary trade-offs in their quest for accuracy or speed, or some other objective function. In fact, Moldoveanu (2009) brings forward the trade-off dimensions of accuracy and reliability, which the trade-off literature has shown to be valid, now widely recognized, and much better understood. These trade-offs are not new and have not come into existence with the deployment of computational approaches against such classes of problems. However, these trade-off dimensions are perhaps more visible, and the issues considered more valid, as the advancement of algorithmic into business problems has accelerated over the past twenty-odd years and more research has taken place into these areas. Moldoveanu (2006) posits that many issues of complexity are simply ignored when, for example, NP problems are reduced to P problems. The simplification and reduction in complexity, *ipso facto*, means a reduction in variables considered. I suggest that variables that are left out include considerations such as trade-off dimensions. For example, in an effort to gain accuracy and as a result of problem simplification, dimensions such as explainability or privacy may not be fully considered and captured into the organization objective function. As the application of algorithms and technologies have become more accessible, pervasive and expanded, the desire for dimensions such as speed and perceived accuracy gain, have become perhaps disproportionate motivations, as the expense of other important considerations that when lost, diminish the overall net benefit to an organization and its stakeholders. In effect, while the strategist's challenge is not new in the literal sense, the notion of a necessary and explicit expansion of the variables considered is more recent. In the context of applications within industries such as healthcare or legal judgements, I would argue that problem simplification is not an option, and certainly not as more and more decisions are turned over to algorithmic *emulators*.

Explicated Trade-off Dimensions and Proposed Trade-offs

The trade-off dimensions explicated from the literature are: Accuracy, Explainability, Fairness, Privacy, Reliability, Security and Speed. These dimensions have relevance in organization decision-making *and* specific meaning in the context of computational optimizations. While some may seem related, each of these stands mutually exclusive from the others as a scalar dimension (Moldoveanu, 2020). Each of these constructs also has a particular meaning in the context of AI technologies deployed to organization strategic problems solving, strategy formulation and decision making. As is apparent in the next section, while there is support for the general definition of the trade-off dimensions, there is no singular definition or complete consensus on any particular dimension.

Definitions of Trade-off Dimensions

Accuracy. The accuracy of a machine learning classification algorithm is one way to measure how often the algorithm classifies a data point correctly. Accuracy is the number of correctly predicted data points out of all the data points. More formally, it is defined as the number of true positives and true negatives divided by the number of true positives, true negatives, false positives, and false negatives.[16] Machine learning model accuracy is the measurement used to determine which model is best at identifying relationships and patterns between variables in a dataset based on the input, or training, data. The better a model can generalize to 'unseen' data, the better predictions and insights it can produce, which in turn

[16] https://deepai.org/machine-learning-glossary-and-terms/accuracy-error-rate, retrieved Oct 21, 2021.

deliver more business value.[17] Moldoveanu (2021) defines accuracy as the degree of precision with which the results match the problem solver's objective.

Explainability. Explainability means that the model is able to explain why and how an AI algorithm arrives at a specific decision while maintaining its accuracy.[18] Arrieta et al. (2019, p. 31) define explainability "as the ability a model has to make its functioning clearer to an audience". According to IBM, explainability "is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms. Explainable AI is used to describe an AI model, its expected impact and potential biases. It helps characterize model accuracy, fairness, transparency and outcomes in AI-powered decision making. Explainable AI is crucial for an organization in building trust and confidence when putting AI models into production. AI explainability also helps an organization adopt a responsible approach to AI development".[19] Explainability "is the concept that a machine learning model and its output can be explained in a way that 'makes sense' to a human being at an acceptable level. Certain classes of algorithms, including more traditional machine learning algorithms, tend to be more readily explainable, while being potentially less performant. Others, such as deep learning systems, while being more performant, remain much harder to explain."[20] "Explainable AI provides methods and techniques to produce explanations about the used AI and the decisions made by it, consequently, helps attaining human trust."[21] Moldoveanu (2021)

---

[17] https://www.datarobot.com/wiki/accuracy/, retrieved Oct 21, 2021.
[18] https://www.ericsson.com/en/reports-and-papers/white-papers/explainable-ai--how-humans-can-trust-ai, retrieved Oct 24, 2021.
[19] https://www.ibm.com/watson/explainable-ai, retrieved Oct 24, 2021.
[20] https://c3.ai/glossary/machine-learning/explainability/, retrieved Oct 24, 2021.
[21] https://www.ericsson.com/en/reports-and-papers/white-papers/explainable-ai--how-humans-can-trust-ai, retrieved Oct 24, 2021.

defines explainability as the degree to which the workings and results of an algorithmic solution, along with the algorithmic procedure, can be explained to all of the stakeholders or people whose welfare is affected by the output, but also to regulators, other executives and the 'public at large'.

Fairness. In the context of decision-making, fairness is the absence of any prejudice or favoritism toward an individual or a group based on their inherent or acquired characteristics. (Mehrabi et al, 2019). Bias and fairness in AI are two sides of the same coin and while there is no universally agreed upon definition for fairness, it can be broadly defined as the absence of prejudice or preference for an individual or group based on their characteristics.[22] Relevant fairness definitions aim to detect and prevent discriminatory (or other) bias with respect to a set of protected attributes, such as gender, race, and disability status. (Foulds et al., 2019). Moldoveanu (2021) defines fairness as the degree of impartiality of the results with respect to reference classes of the individuals whose behaviors generate the data set (e.g. race, gender of subjects) but that are *not* causally meaningful to the specific patterns in the data, in a way that either does not unduly penalize individuals in specific populations or demographic or are consistent with social norms of re-distributive justice and fairness.

Privacy. The basic definition of privacy is having the power to seclude oneself, or information about oneself, in order to limit the influence others can have on our behavior. In the information age, privacy hinges on our ability to control how our data is being stored, modified, and exchanged between different parties (Deane, 2018).[23] Moldoveanu (2021) defines privacy as

[22] https://towardsdatascience.com/understanding-bias-and-fairness-in-ai-systems-6f7fbfe267f3, retrieved Oct 24, 2021.

[23] https://towardsdatascience.com/ai-and-the-future-of-privacy-3d5f6552a7c4, retrieved Oct 24, 2021.

the degree to which the solution guards the identity of the subjects that supply the data used in the solution from being known to others, without the subject's permission, and to a level that is acceptable to the subjects themselves.

Reliability. Moldoveanu (2021) defines reliability as the degree to which the solution can be extrapolated or generalizes across a wide enough range of conditions, other cases and situations, potentially dissimilar to those in which the model was calibrated and tested, to provide substantial value.

Security. AI security refers to tools and techniques that leverage artificial intelligence to autonomously identify and/or respond to potential cyber threats based on similar or previous activity.[24] Moldoveanu (2021) defines security as the degree to which the data, algorithms and predictions are sufficiently immune to attacks that are either random or targeted to provide sufficient assurance to users of the solution and originators of the data that the solution will not be misused.

Speed. Moldoveanu (2021) defines speed as the rate at which an algorithmics solution generates useful information, or the probability that enough useful information will be generated within a certain time window, which is the 'maximum viable time scale' of the organization, given current CPU/GPU, memory and input/output limitations.

Proposed Pairwise Trade-offs

With seven scalar trade-off dimensions, we get *7-choose-2* or twenty-one (21) trade-offs as a series of pairwise 1-to-1 trade-offs, in a matrix below in Table 4. In comparison to the

[24] https://awakesecurity.com/glossary/ai-security/, retrieved Oct 24, 2021.

volume of work and literature which exists on the general concepts of AI and ML technologies applied to organization strategy and decision making, the volume of literature on these hypothesized trade-offs is less wide and deep, with substantial disparity on coverage of specific pairwise trade-offs. For example, trade-offs such as *accuracy vs. explainability* or *accuracy vs. fairness* appear to be much better studied and published than trade-offs such as *reliability vs. fairness* or *fairness vs. security*.

Table 4

Matrix of Seven Trade-off Dimensions and Twenty-One Pairwise Trade-offs, expanded from Moldoveanu (2021)

	Accuracy	**Explainability**	**Fairness**	**Privacy**	**Reliability**	**Security**	**Speed**
Accuracy		Trade-off 1	Trade-off 2	Trade-off 3	Trade-off 4	Trade-off 5	Trade-off 6
Explainability			Trade-off 7	Trade-off 8	Trade-off 9	Trade-off 10	Trade-off 11
Fairness				Trade-off 12	Trade-off 13	Trade-off 14	Trade-off 15
Privacy					Trade-off 16	Trade-off 17	Trade-off 18
Reliability						Trade-off 19	Trade-off 20
Security							Trade-off 21
Speed							

There are also significant gaps in the trade-off literature, with a lack of coverage of several trade-offs: explainability vs. security, explainability vs. speed, fairness vs. reliability, fairness vs. security, fairness vs. speed, privacy vs. reliability, reliability vs. security, and security vs. speed.

This is a growing and evolving field of study, and it's likely that the coverage of specific trade-off and trade-off dimensions will continue to develop through the process of completing this research study. A total of 89 pieces of literature were examined and are categorized below in Tables 5, 6 and 7. The idea of categorizing trade-offs into industries and functional areas, while different from categorization by deployment of different AI techniques, was nonetheless sparked by Chiu et al., (2018). This categorization presents a very classical management consulting approach, which applies learnings from past engagements across industries and functions. My purpose in using an industry and functional area classification is to spur thinking and research into areas where there are clear gaps. We've already seen a disparity in terms of the quantity and quality of research and analysis into the different trade-offs—my testable hypothesis is that, areas with gaps are not because the trade-off is less relevant or important. My current thinking is that this may be because of a lack of industry drivers or a clear articulation of the challenges, which has led to less study, thereby offering gaps and opportunities for the literature to catch up.

Table 5

Literature Categorized Within a Trade-off Matrix

Dimension	**Accuracy D_A**	**Explainability D_E**	**Fairness D_F**	**Privacy D_P**	**Reliability D_R**	**Security D_S**	**Speed D_{SP}**
Accuracy D_A		*TRADE-OFF 1* Adadi & Berrada (2018) Amann et al. (2020) Arrieta et al. (2019) Bhatt et al. (2020)	*TRADE-OFF 2* Arrieta et al. (2019) Bellamy et al. (2018) Choraś et al. (2020) Foulds et al. (2019)	*TRADE-OFF 3* Abowd & Schmutte (2018) Mazeh & Shmueli (2019) Papernot et al. (2016)	*TRADE-OFF 4* Bader & Kaiser (2019) Gerke et al. (2020) Kononenko (2001) Kundu et al. (2021)	*TRADE-OFF 5* Lu & Shen (2015) Papernot et al. (2016) Arrieta et al. (2019) Babcock et al. (2019)	*TRADE-OFF 6* Bansal et al. (2019) Guo (2019) Kononenko (2001) Lalmuanawma et al. (2020) Nguyen et al. (2019)

Dimension	**Accuracy** D_A	**Explainability** D_E	**Fairness** D_F	**Privacy** D_P	**Reliability** D_R	**Security** D_S	**Speed** D_{SP}
		Choraś et al. (2020) Gerke et al. (2020) Goebel et al. (2018) Gunning & Aha (2019) Guo (2019) Hacker et al. (2020) Holzinger (2018) Holzinger et al. (2018) Kuhn & Kacker (2019) London (2019) Piano (2020) Rai (2020) Roscher et al. (2020) Varian (2017) Vollmer et al. (2020) Kononenko (2001) Rudin (2019)	Liu & Vicente (2021) Mehrabi et al. (2019) Oneto & Chiappa (2020) Pessach & Shmueli (2022) Piano (2020) Tolan et al. (2019) Valdivia et al. (2020) Valera et al. (2018) Zliobaite (2015)	Shokri & Shmatikov (2015) Arrieta et al. (2019) Holzinger et al. (2018) Liu et al. (2020) Varian (2017) Xu et al. (2021)	Lalmuanawma et al. (2020) Roscher et al. (2020) Rudin (2019) Vollmer et al. (2020)	Choraś et al. (2020) Szychter et al. (2018) Varian (2017) Wu et al. (2020)	Srivastava et al. (2014) Vollmer et al. (2020)
Explainability D_E			*TRADE-OFF 7* Rai (2020)	*TRADE-OFF 8* Rai (2020) Manheim & Kaplan (2019)	*TRADE-OFF 9* Rudin (2019)	*TRADE-OFF 10*	*TRADE-OFF 11* Guo (2019)
Fairness D_F				*TRADE-OFF 12*	*TRADE-OFF 13*	*TRADE-OFF 14*	*TRADE-OFF 15*

Dimension	**Accuracy D_A**	**Explainability D_E**	**Fairness D_F**	**Privacy D_P**	**Reliability D_R**	**Security D_S**	**Speed D_{SP}**
				Foulds et al. (2019) Manheim & Kaplan (2019)			Mehrabi et al. (2019)
Privacy D_P					*TRADE-OFF 16*	*TRADE-OFF 17* Babcock et al. (2019) Papernot (2017)	*TRADE-OFF 18* Shokri & Shmatikov (2015) Xu et al. (2021)
Reliability D_R						*TRADE-OFF 19*	*TRADE-OFF 20* Kundu et al. (2021) Lalmuanawma et al. (2020)
Security D_S							*TRADE-OFF 21*
Speed D_{SP}							

Trade-offs by Industry Focus and Functional Area

I've already put forward idea that the unequal coverage of trade-offs in the literature may be due to factors and drivers unrelated to the relevance or importance of trade-offs themselves, and may be that certain industry or contextual drivers have not provided sufficiently fertile empirical landscapes. From Table 5 above, it is obvious that of the trade-off literature found and reviewed herein, accuracy vs explainability is the most voluminous. Treating the dimensions on their own, once can see that accuracy seems to be the most studied and written about, and this will be discussed in more detail below where I discuss the themes discovered within the literature. To better understand and visualize the finding that there is unequal study and coverage of the dimensions, the following Tables 6 and 7 categorize the same twenty-one trade-offs, but this time by industry and functional area, respectively. I think these different categorized views present a compelling picture of the current coverage and gaps, and offer researchers opportunities for further study into various fields of study. They also present practitioners and organizations with an interesting picture of where their industry or functional area resides in terms of the current literary coverage into different trade-offs. If a particular organization or intra-organization department/functional area finds itself in a relatively well-covered area, then the hope is that they may be able to improve their probability of a successful outcome, by drawing upon the available research and analysis on relevant trade-off(s). If on the other hand, they find themselves in a less well-covered area, then the challenge becomes more difficult in terms of addressing a less than expected outcomes from an AI deployment, where the reasons now also including the fact that the trade-off(s) they may be grappling with is/are simply not well understood at this time. This study endeavours to move the theory and understanding into all these areas, to the benefit of organization strategic decision-making and outcomes.

Table 6

Literature Categorized by Trade-off and Industry

	Healthcare	Public Policy and Law	Finance and Insurance	ICT and Telecom	Consumer Products	Social Media	Natural and Earth Sciences	Security Services
TRADE-OFF 1 **Accuracy v. Explainability**	Adadi & Berrada (2018) Amann et al. (2020) Bhatt et al. (2020) Gerke et al. (2020) Kuhn & Kacker (2019) London (2019) Vollmer et al. (2020	Adadi & Berrada (2018) Amann et al. (2020) Choraś et al. (2020) Gerke et al. (2020)	Adadi & Berrada (2018) Bhatt et al. (2020) Kuhn & Kacker (2019)	Guo (2019)			Roscher et al. (2020)	Adadi & Berrada (2018)
TRADE-OFF 2 **Accuracy v. Fairness**	Pessach & Shmueli (2022)	Bellamy et al. (2018) Foulds et al. (2019) Liu & Vicente (2021) Pessach & Shmueli (2022)	Bellamy et al. (2018) Liu & Vicente (2021) Pessach & Shmueli (2022)					
TRADE-OFF 3		Abowd & Schmutte (2018)						

	Healthcare	Public Policy and Law	Finance and Insurance	ICT and Telecom	Consumer Products	Social Media	Natural and Earth Sciences	Security Services
Accuracy v. Privacy								
TRADE-OFF 4 **Accuracy v. Reliability**	Lalmuanawma et al. (2020) Gerke et al. (2020) Vollmer et al (2020	Gerke et al. (2020)		Bader & Kaiser (2019)			Roscher et al. (2020)	
TRADE-OFF 5 **Accuracy v. Security**				Wu et al. (2020)				Babcock et al. (2019)
TRADE-OFF 6 **Accuracy v. Speed**	Lalmuanawma et al. (2020) Vollmer et al (2020			Guo (2019)				
TRADE-OFF 7 **Explainability v. Fairness**	Rai (2020)	Rai (2020)						
TRADE-OFF 8 **Explainability v. Privacy**	Rai (2020)	Manheim & Kaplan (2019) Rai (2020)						
TRADE-OFF 9 **Explainability v. Reliability**								

	Healthcare	Public Policy and Law	Finance and Insurance	ICT and Telecom	Consumer Products	Social Media	Natural and Earth Sciences	Security Services
TRADE-OFF 10 **Explainability v. Security**								
TRADE-OFF 11 **Explainability v. Speed**				Guo (2019)				
TRADE-OFF 12 **Fairness v. Privacy**		Foulds et al. (2019) Manheim & Kaplan (2019)						
TRADE-OFF 13 **Fairness v. Reliability**								
TRADE-OFF 14 **Fairness v. Security**								
TRADE-OFF 15 **Fairness v. Speed**								
TRADE-OFF 16 **Privacy v. Reliability**								
TRADE-OFF 17								Babcock et al. (2019)

	Healthcare	Public Policy and Law	Finance and Insurance	ICT and Telecom	Consumer Products	Social Media	Natural and Earth Sciences	Security Services
Privacy v. Security								
TRADE-OFF 18 **Privacy v. Speed**								
TRADE-OFF 19 **Reliability v. Security**								
TRADE-OFF 20 **Reliability v. Speed**	Lalmuanawma et al. (2020)							
TRADE-OFF 21 **Security v. Speed**								

Table 7

Literature Categorized by Trade-off and Functional Area

	Finance and Risk	Human Resources	Sales and Marketing	Manufacturing and Supply Chain	Product Development	Strategy and Corporate Development	Academia
TRADE-OFF 1 **Accuracy v. Explainability**	Bhatt et al. (2020) Guo (2019)				Guo (2019)	Bhatt et al. (2020) Choraś et al. (2020) Guo (2019)	Gerke et al. (2020) Amann et al (2020)

	Finance and Risk	Human Resources	Sales and Marketing	Manufacturing and Supply Chain	Product Development	Strategy and Corporate Development	Academia
							Arrieta et al. (2019) Kuhn & Kacker (2019) Roscher et al. (2020)
TRADE-OFF 2 **Accuracy v. Fairness**	Bellamy et al. (2018) Liu and Vicente (2021) Pessach & Shmueli (2022)	Bellamy et al. (2018) Pessach & Shmueli (2022)	Pessach & Shmueli (2022)			Choraś et al. (2020)	Foulds et al. (2019) Pessach & Shmueli (2022)
TRADE-OFF 3 **Accuracy v. Privacy**			Agrawal & Haritsa (2004)				Agrawal & Haritsa (2004)
TRADE-OFF 4 **Accuracy v. Reliability**						Bader & Kaiser (2019)	Gerke et al. (2020) Roscher et al. (2020)
TRADE-OFF 5 **Accuracy v. Security**	Wu et al. (2020)					Choraś et al. (2020) Wu et al. (2020)	Babcock et al. (2019)
TRADE-OFF 6 **Accuracy v. Speed**	Guo (2019)					Guo (2019)	

	Finance and Risk	Human Resources	Sales and Marketing	Manufacturing and Supply Chain	Product Development	Strategy and Corporate Development	Academia
TRADE-OFF 7 **Explainability v. Fairness**							
TRADE-OFF 8 **Explainability v. Privacy**							
TRADE-OFF 9 **Explainability v. Reliability**							
TRADE-OFF 10 **Explainability v. Security**							
TRADE-OFF 11 **Explainability v. Speed**	Guo (2019)					Guo (2019)	
TRADE-OFF 12 **Fairness v. Privacy**							Foulds et al. (2019)
TRADE-OFF 13 **Fairness v. Reliability**							
TRADE-OFF 14 **Fairness v. Security**							

	Finance and Risk	Human Resources	Sales and Marketing	Manufacturing and Supply Chain	Product Development	Strategy and Corporate Development	Academia
TRADE-OFF 15 **Fairness v. Speed**							
TRADE-OFF 16 **Privacy v. Reliability**							
TRADE-OFF 17 **Privacy v. Security**							Babcock et al. (2019)
TRADE-OFF 18 **Privacy v. Speed**							
TRADE-OFF 19 **Reliability v. Security**							
TRADE-OFF 20 **Reliability v. Speed**							
TRADE-OFF 21 **Security v. Speed**							

Key Themes

Several themes appear in the literature focused on the trade-off dimensions and the corresponding trade-offs. Certainly, the difference in the quantity of work in different area is evident from Tables 5, 6, and 7. The discovered themes vary from analysis on hindrances to greater adoption of certain AI technologies and/or by certain industries, to known issues around trust and reliability, to commentary on lack of consensus around definitions and language, to threats to democracy and society, to work focused on area of improvement. Themes such as validity, reliability, trust and explainability—in its common usage and not as a specific construct in the context of this research study—have varying importance depending on industry and application. For example, trust and reliability mean very different things in the context of a life and death healthcare diagnosis versus the marketing of a product or service, not to take anything away from the latter. Therefore, the thematic clusters and subclusters appearing in the literature are largely driven by the specific industry, application or deployment context. This is an interesting result since the AI techniques themselves are often common across industries and deployments. A MLA for classification does not substantially differ in design regardless of whether it's *fed* movie-viewing preferences or radiology images. That said, the overall architecture may be different, and the entirety of a system of AI techniques applied to a particular problem may have different components.

Accuracy and Speed are Predominant Objectives. The vast majority of the trade-off literature focuses on the trade-off dimension of accuracy vs. one or more of the other dimensions. The most prevalent trade-off is accuracy vs. explainability, followed by accuracy vs. fairness and then accuracy vs. privacy. As a numerical proportion, the accuracy dimension is the focus of 46 articles/studies out of the herein reviewed 89 works (or 52%) focused on trade-offs. It should not be of any surprise that the dimension of accuracy is the most sought-after objective of organizations deploying AI technologies against their strategic problem solving and decision making. Computational resources are often relied upon for their accuracy and speed; for example, one need only think of common computational machines such as calculators or personal computers. While it's expected that additional literature will continue to emerge on the array of trade-offs defined herein, it will be interesting to monitor if and how much the proportional representation of the trade-offs and respective dimensions shifts as the literature continues to grow and evolve.

The quest for gains in the dimension of speed presents unique challenges in the context of this research study and specifically, abstract classes of problems such as strategy and problems requiring human judgment. These types of problems fall into the NP-hard category, which requires exponential computational time for computational machines to solve (Agrawal et al., 2018b; Moldoveanu, 2009, 2016). In human terms, the time needed for an advanced AI system to solve these problems of NP-class complexity means the user will simply run out of time when trying to solve an intractable problem via a deterministic algorithm (Moldoveanu, 2016), rendering it practically unsolvable (Rivkin, 2000, as cited in Moldoveanu, 2016).

Lack of Explainability is a Significant Hindrance to Adoption. Lack of explainability is a hindrance to adoption (Adadi & Berrada, 2018; Amann et al., 2020; Arrieta et al., 2019; Canhoto & Clear, 2019; Choraś et al., 2020; Došilović et al., 2018; Faes et al., 2020; Magoulas & Swoyer, 2020). This can occur for several reasons. One important effect of this is a lack of trust, which is especially relevant in the context of healthcare and mission-critical systems

(Adadi & Berrada, 2018; Kuhn & Kacker, 2019; Liu & Vicente, 2021; Rai, 2020). As an affirming counterexample, Kononenko (2001) suggests support for improved adoption due to increased reliability, accuracy and speed. Within the context of healthcare, several works make the case that trust and reliability are based on an health expert's critical ability to explain rationale, apply causal knowledge, or otherwise mitigate opaqueness. Furthermore, explainability is not a purely technical issue, but involves and invokes numerous other medical, legal, ethical and societal questions (Amann et al., 2020; Kuhn & Kacker, 2019; Rai, 2020). But these points of view are not universal, and London (2019) argues that "a blanket requirement that machine learning systems in medicine be explainable or interpretable is unfounded and potentially harmful" (p. 15).

Adadi and Berrada (2018) focus on the accuracy vs. explainability trade-off, and the lack of transparency as an impediment to greater adoption of AI technologies. They underscore that the lack of transparency and the corresponding lack or loss of trust between user and system as the "*sine qua non* for AI to continue making steady progress without disruption" (p. 52138). Whereas thus far we've seen an unclear path for AI technologies to technically answer important classes of organization problems, Adadi and Berrada (2018) turn our attention to areas where the challenge is less technical and more based on trust and understanding of how these systems compute what they compute, such as within healthcare, life-and-death decisions and disease diagnosis. There is an important development here, coined FAT* academics, which stands for Fairness, Accountability, and Transparency in multiple artificial intelligence, machine learning, computer science, legal, social science, and policy applications. These so-called FAT* academics are primarily focused on promoting and enabling explainability and fairness in algorithmic decision-making systems with social and commercial impact (p. 52139). A challenge pointed out by Adadi and Berrada but widely recognized elsewhere as well, is that lack of a commonly accepted definition of Explainable AI or XAI (Gunning, 2017; Gunning & Aha, 2019; Gunning et al., 2019). A related issue is the interchangeable application of different terminology to the concept and dimension of explainability, such as interpretable, interoperable, understandable, and responsible. Ultimately, their conclusion holds—that the complexity of a ML model or algorithm is directly related to its interpretability (or explainability), and once can argue that enhanced accuracy in the context of this accuracy vs. Explainability trade-off generally means more complexity. So, greater Explainability or interpretability results through algorithmic simplification. But can we achieve the latter while increasing accuracy?

Issues around Fairness and Reliability. Liu and Vicente (2021) discuss the trade-off of accuracy vs. fairness. Specifically, the thrust of their paper is that the application of ML technologies on *real-life* decisions such as credit scoring and criminal justice, there is the potential for the prediction outcomes to be unfair against subjects with sensitive attributes. Rather than adopting a commonly used approach of limiting the unfairness by minimizing fairness dimensions as constraints, which reduces accuracy by reducing the amount of information, Liu and Vincente (2021) propose a novel approach. They introduce a stochastic multi-objective optimization problem which comprehensively defines discrete sensitive attributes such as gender, race, disability; and an algorithm that inequitably treats subjects on any particular pareto front of the objective frontier is deemed unfair. In their own words:

> Instead of looking for a single predictor that satisfies certain fairness constraints, our goal is to directly construct a complete Pareto front between prediction accuracy and fairness, and thus to identify a set of predictors associated with different levels of fairness. We propose a stochastic multi-objective

> optimization framework, and aim at obtaining good approximations of true Pareto fronts. (Liu & Vincente, 2021, p. 2)

While their proposed model holds promise, their solution is a compromise in that it produces "sufficiently accurate Pareto fronts" while ensuring the algorithm itself is not deemed to be biased. In a way they have altered the inherent problem within this trade-off. There is still a gap in terms of practitioners addressing and able to utilize a fully informed model whereby this trade-off can be considered and addressed without altering the overall objective function.

Rai (2020) focuses on the fairness vs explainability and privacy vs explainability trade-offs and how these "black boxes" can be made more transparent "glass boxes" (p. 138). They underscore issues raised by Adadi and Berrada (2018), Choraś et al., 2020; Foulds et al., 2019; Liu and Vicente (2021); Mehrabi et al., 2019; Schmelzer, 2019; Siau and Wang, 2018; The Royal Society, 2019; Tjoa and Guan, 2020; Tolan et al., 2019; US Food and Drug Administration, 2019; Vellido, 2018; and Zhou et al., 2021, where "inscrutable systems" (Rai, 2020, p. 137) cause loss of users' trust within critical fields such as healthcare decisions, or biases leading to prejudicial treatment of certain groups. Their conclusion that different ML algorithms and their application within different AI systems have different inherent complexities and degrees of interpretabilities, the latter term being used interchangeably with explainability (Hall and Gill 2019; Du et al. 2018, as cited in Rai, 2000). They categorize various ML algorithms into interpretable such as decision trees and Bayesian classifiers, versus less-interpretable deep learning algorithms, with the latter being more accurate but less transparent and interpretable. While their conclusion regarding these trade-offs is well-supported, their suggestion that XAI may unmask these so-called black-box AI systems is less clear. The model-specific techniques they put forward appear to reduce complexity, thereby potentially reducing accuracy. The model-agnostic techniques they put forward involve either some sort of human intervention, or a separate treatment of a “perturbed dataset”. In either scenario, there is no clear path to truly mitigation of these trade-offs. Where gains may be made in explainability, there is a potential of loss in accuracy.

Bellamy et al. (2018): Recent years have seen an outpouring of research on fairness and bias in machine learning models. This is not surprising, as fairness is a complex and multi-faceted concept that depends on context and culture. Narayanan described at least 21 mathematical definitions of fairness from the literature (Narayanan, 2018). These are not just theoretical differences in how to measure fairness; different definitions produce entirely different outcomes.

Kundu et al. (2021) focuses on the accuracy vs. reliability and reliability vs. speed trade-offs in the context of systems hardware. While being very technical and clearly written for engineers, this paper does offer some useful and insights for this research study. Specifically, a systems level approach is required to truly examine and address the reliability dimension and only looking at this from an AI system or ML algorithm level is insufficient. Kundu et al. demonstrate that in large-scale systems, hardware issues even at infinitesimal rates can degrade accuracy beyond acceptable limits, thereby rendering the output of limited or no use. This can also be problematic from a security standpoint if hardware vulnerabilities or oversights are used as a point of attack my malicious entities to try and skew or manipulate AI system outputs.

Valdivia et al. (2020) focuses on the accuracy vs. fairness trade-off. They propose a method that, by their own assessment, increased fairness but with a reduction in accuracy. They present an elegant summary of an objective frontier in the context of AI systems and ML algorithms applied to organization decision making:

> A solution is called Pareto optimal if there does not exist another solution that dominates it. Consequently, the set of all Pareto optimal solutions is defined as Pareto front or boundary. Assessing this frontier allows decision makers to select any efficient solution, depending on the worthiness of each objective function. (Valdivia et al., 2020, p. 6)

Of most value for this study is their finding that, for a reduced or even unchanged level of complexity, it's more difficult for the algorithm to increase accuracy than fairness. Put another way, a marginal gain in fairness leads to higher cost in accuracy.

Oneto and Chiappa (2020) examine the accuracy vs. fairness trade-off, with a focus on the ML techniques that seek to specifically protect against unfair treatment of individuals due to biases in the data or model using Causal Bayesian Networks (CBNs). This paper is valuable to this research study as it presents a unified framework that theoretically improves the accuracy vs. fairness trade-off across different settings and fairness criteria.

Explainability, Fairness and Reliability within Mission-Critical Contexts. The deployment of AI technologies is ever-encroaching into mission critical contexts and environment, including but not limited to medical diagnoses, autonomous vehicles, and justice and law. Several trade-off dimensions seen thus far including explainability, fairness and reliability are all indispensably important in mission critical situations where the outcome may be the different between life and death, or someone losing their freedom.

London (2019) examines the trade-off of accuracy vs. explainability within the context of medical decisions. He makes a compelling case for his underlying premise and position, that views held by many in terms of a health expert's critical ability to explain rationale, apply causal knowledge, or otherwise mitigate opaqueness, is at best unnecessary and at worst untrue. London seems to agree with Rudin's position (2019) that explainability is a myth. He is critical of many health decisions taken by so-called experts and argues that many empirical medical findings are epistemic and are in fact more "atheoretical", "associationist", and opaque than the critics of AI and ML realize (p. 15). Without getting unnecessarily involved in his criticisms or human medical decision-making and whether or not those experts act by means put forward by the criticizers like London (2019), there is support for his assertion that technical advancements such as deep learning can and do lead to highly accurate diagnoses (Amann et al., 2020; Chui at el., 2018; Kannan et al., 2020; Kuhn & Kacker, 2019; Kononenko, 2001). Yet, these highly complex systems and algorithms are theory agnostic, and in the problem of radiology, architected to more accurately diagnose tissue malignancies.

Amann et al. (2020) also focus on the accuracy vs. explainability trade-off within healthcare. Their work is another in a long list of contributions by various researchers looking at this trade-off within important life-impacting settings such as healthcare. Of relevance to this study is their submission that explainability is not a purely technical issue, but involves and invokes numerous other medical, legal, ethical and societal questions. They allude to the fact that non-AI or human-based decision making may not be the panacea, and that a collaborative multi-disciplinary approach is required to take advantage of and benefit from medical AI. In addition, they spend considerable effort defining what explainable means for AI systems in the context of healthcare, and how terms such as interpretability and/or transparency—while different—are used synonymously.

Kononenko (2001) examines the role of ML in the context of medical diagnoses, and area fraught with concerns about primarily explainability, but also issues around privacy and security. However, Kononenko work is better classified as a view on the accuracy vs. reliability and accuracy vs. speed trade-offs. He posits, correctly in my opinion, that the potential advances and gains to be had will soon lead to greater deployment and exploitation of these technologies. Critical to this will be the seamless integration of these technologies into existing physicians' instrumentation, or at a minimum not add to their already complicated work. He treats ML tools and technologies as *just another* tool available to physicians, where they retain ultimate control on whether and how, to utilize the information and analysis being provided by these new ML technologies. While Kononenko presents a very pragmatic and potentially supported view of the use of ML technologies within medical diagnosis, there is ultimately little development of concepts to advance solutions to the trade-offs his work fit into.

The theme and issue of ethics runs through much of the literature (Adadi & Berrada, 2018; Amann et al., 2020; Arrieta et al., 2019; Canhoto & Clear, 2019; Greene et al., 2019; Manheim & Kaplan, 2019; Mittelstadt, et al., 2018; Oneto & Chiappa, 2020; Piano, 2020; Rudin, 2019; Russell & Norvig, 1995; Szychter et al., 2018; The Royal Society, 2019; Tjoa & Guan, 2020; Vollmer et al., 2020; Vellido. 2018), especially the works focused on matters or cases involving human subjects or examples of AI techniques impacting peoples' lives. These include heath decisions, inappropriate treatment or biases within certain MLAs, autonomous vehicles, and law and justice matters.

Of interest to this study, Piano (2020) presents an expansive review of literature and concepts focused on the accuracy vs. explainability trade-off, from the premise that the latter dimension is being eroded in favour or speed and efficiency. He examines the ex-ante (underpinning algorithm development) and ex-post (consequences on those impacted or acted-on) impacts of algorithms on several ethical issues, and of interest to this study, puts forward numerous approaches from various literature on how best to incorporate and manage these ethical concerns, i.e. super-parte entities that become *defacto* gate-keeps of standards and approvals. The idea is interesting, but I think ultimately flawed because it directly and explicitly limits two key aspects of learning systems—that the learning Is predicated on the feeding of data sets, and that their accuracy is largely understood to be a *result* and by-product of their complexity, thereby reinforcing the accuracy vs. explainability trade-off. His assertion that a decision-making algorithm will always be a model of the real system, and as such, will always be limited in some way, aligns with Rudin (2019). The ethical aspects of machine learning and AI systems is a very complicated area, and while perhaps it's not as far as the overarching trade-off of explainability, it is becoming an increasingly important issue as these systems and algorithms are handed over critical decision-making powers such as health diagnoses or criminal justice matters. As this research study proceeds, I intend to evaluate the concept of ethics as a stand-alone dimension to see if it should be added to the matrix, or whether it is best captured elsewhere.

Data Privacy. Privacy is an important trade-off dimensions, and traverses many contexts from sensitive health data, to democratic rights and voting, to personal records, preferences and habits (Abowd & Schmutte, 2018; Deane, 2018; Liu et al., 2020; Mazeh & Shmueli, 2019; Shokri & Shmatikov, 2015). Given the interconnected nature of our personal technologies and devices and the mountains of data being mined by companies and applications that *live and work* on these devices like Google, Facebook, LinkedIn, and so on, the safeguarding of privacy is an emerging area of study in the context of AI technologies generally, and within the scope of this research study.

Abowd and Schmutte (2018) make an important contribution focused on the *accuracy vs. privacy* trade-off. They present a model with the objective of maximizing accuracy while minimizing privacy loss, within the context of governmental policy decisions and trying to gain statistical accuracy while mitigating privacy loss to a level acceptable to the data set (the surveyed population), via a so-called Social Welfare Function. Abowd and Schmutte define a positive/increasing concave production function of accuracy vs. privacy loss, and examine an economic model of this trade-off. They suggest an approach for operating where the marginal cost of increased accuracy equals the marginal cost of privacy loss, or worded the other way around, the marginal cost of increased privacy (less privacy loss) provides marginal benefit at the expense of an equal marginal loss of accuracy.

The only shortcoming I will highlight, is that their analysis and mathematical models are limited a production function comprising a closed set of transformation activities that the statistical agency can undertake. This limited problem type offers limited re-applicability of their approach because an underlying hypothesis of this study is that organizations deliberately or otherwise ignore, and therefore do not include, a number of other factors or variables in their AI technology deployments. Hence, the achieved outcomes often fall short of their desired objectives or come with unintended and uncaptured consequences. As an example, Abowd and Schmutte (2018) do not take into account computational resources and how these may factor into gaining accuracy, or how they can be leveraged to potentially mitigate the loss of privacy. It is particularly interesting is that Abowd and Schmutte (2018) hint at and suggest directions for research which I can connect to the directions I pursue in this study (pp. 37-39). While maintain the 1-to-1 trade-off between accuracy and privacy, they introduce several areas of extending the production and preference models. For example, amongst other model expansion ideas within data gathering and dissemination, they introduce a time element into their analysis, whereby population differential privacy preferences may change over time. Aligning with this proposed study, they seem to be laying the groundwork for researchers to incorporate additional variables at various stages of their statistical agency functions, that can be manipulated by stakeholders to enhance the production function, thereby potentially gaining accuracy with less privacy loss.

Another approach to the accuracy vs. privacy trade-off is presented by Mazeh and Shmueli (2019). While not addressing the technical challenge of this trade-off within an/one organization, their proposal attempts to address it via extra/inter-organizational information sharing, access and control. They propose an advancement upon the Open Personal Data Store (de Montjoye et al., 2014, as cited in Mazeh & Shmueli, 2019), to mitigate the privacy concern of a single entity holding users' information, while gaining accuracy through the aggregation of user data, plus user-level control of one's own information. Their proposed approach seeks an optimization by maintaining a baseline level of accuracy while mitigating privacy concerns. A smaller loss in the trade-off dimension of privacy while maintaining accuracy is a net gain in the objective function, but further analysis is required to assess whether this is a maximization of the objective function or whether further improvements can be made in the context of this research study.

Liu et al. (2020) is a review of past and current work focused on the accuracy vs. privacy trade-off, as well as privacy as a standalone dimension, given that "machine learning can act as both friend and foe" (p. 1). On this interesting latter concept, which is very important, and according to Liu et al. largely missing from the current discourse, the accuracy vs. privacy issue falls into three categories: (i) private machine learning, (ii) machine learning aided privacy protection, and (iii) machine learning-based privacy attack and corresponding protection schemes. Most literature thus far looking at this trade-off examines only on privacy concepts during the actual ML process—the computation. Akin to Kundu et al. (2021), Liu at el. take a more holistic approach to their analysis and broaden the discussion to other facets of this trade-off. As seen with explainability and the lack of a common definition, Liu et al. point out a lack of unified metric or notion related to privacy. Of importance to this research study is the notion that ML can play different roles vis-à-vis privacy, and the ever-expanding capabilities of AI and ML technologies requires scholars and practitioners to consider ML as all of: target, protector, and attacker.

Shokri and Shmatikov (2015) focus on the accuracy vs. privacy and privacy vs. speed trade-offs. They highlight the risks of users' personal data being indefinitely stored by 3rd parties for their own disclosed and undisclosed purposes, while medical institutions are prevented from the benefits of these AI systems due to privacy concerns. Their proposed model for multiple parties to share their own datasets with others is interesting, but ultimately does little to address the core trade-offs and may in fact introduce new security, privacy, speed and computational resource concerns through what they call "collaborative deep learning" (p. 1). The sharing of model parameters during the training phase is beneficial, but there is no evidence to indicate that this approach is superior from an accuracy perspective, to a single party having access to all the data. Furthermore, their speed or computational resource baseline comparators are complicated cryptography techniques that are primarily relevant in decentralized and multi-party deep-learning architectures. In my assessment, at best they maintain the status quo accuracy for each participant dataset. “Privacy preserving” is also not really achieved because while they are isolating each dataset, each dataset and the party holding it has its own potential inherent privacy issues. In summary, their solution is almost as accurate, has a marginal theoretical gain in privacy, has reduced speed and higher tangible network cost. So in effect, any gains on the primary trade-offs they focus on, may be lost in performance and computational resources, although I’m not suggesting an equivalency here.

Lack of Consensus Around Definitions and Language, and Misuses. Moving to explainable AI more generally, there is a noteworthy lack of consensus on the term explainable: “The literature clearly asks for an unified concept of explainability” (Arrieta et al., 2019, p. 31). To some, it’s more appropriately termed interpretability (Roscher et al., 2020), while others are unclear and use different words interchangeably (Holzinger et al., 2019) or dismiss the concept altogether (Rudin, 2019). Arrieta et al. (2019, p. 5) provide an interesting amalgamating word: “understandability”. This synonymous use of terminology is particularly problematic because it may not be the same in practice and execution. Adadi and Berrada provide a XAI word cloud (2018, p. 52140) from their linguistic search of other literature, which visualizes the myriad of terms used and the prominence of Interpretable AI *and* Explainable AI as most commonly used words in this growing field of research.

There has been a transition in terms of terminology and practice from interpretable ML to explainable AI since 2017 (Arrieta et al., 2019, pp. 4-5). However, Arrieta et al. also seem to contribute to the lack of consensus in their nine (9) XAI goals (p. 8), some of which are themselves important terms in the context of AI applied to organization strategy: trust, fairness, privacy. The lack of consensus goes beyond the dimension of explainability and impacts other important dimensions such as fairness (Mehrabi et al., 2019). For example, their survey of the literature turns up twenty-three types of biases (p. 4-7), six types of discrimination (p. 10-11), ten definitions of fairness (p. 11-12), and three types of fairness (p. 12). This is remarkable and underscores the challenges posed by some of the trade-off dimensions. How can we even approach the notion of computational machines being able to capture and address these important dimensions if we cannot even clearly define the terms?

Arrieta et al. (2019) summarize prior work with a focus on the dimension of explainability, and undertake an analysis of several trade-offs: accuracy vs. explainability, accuracy vs. fairness, accuracy vs. privacy, and accuracy vs. security. They encompass these together into a new term they call Responsible AI. Of relevance to this study is their proposition of XAI's potential to effectively achieve a solution to the interpretability (explainability) vs. performance (accuracy) trade-off—although we've already seen in other literature that the synonymous usage of the terms explainable and interpretable is itself problematic and some scholars have put forward clear distinctions between these. I disagree with their seemingly causal treatment of complexity and accuracy. The literature is clear that while greater complexity does not in an of itself cause or mean greater accuracy, there is a direct relationship—and indeed a correlation—between greater accuracy stemming from greater complexity. In other words, AI systems which are more complex, for example those that deploy deep learning algorithms, do achieve greater accuracy on tasks requiring such complex architectures. More simple tasks such as differentiation or classification may be solved via less complicated algorithms such as binary trees.

Choraś et al. (2020) considers several trade-offs: accuracy vs. explainability, accuracy vs. fairness, and accuracy vs. security. The thrust of this paper is that there are issues as or perhaps more important than accuracy, namely: security, explainability and fairness. While there is nothing novel about the problem they highlight, nor its importance in building more trustful ML and AI systems and applications, their delineation of explainability and interpretability is an important contribution and helpful. I will quote them in full:

> Interpretability addresses the aspects related to observation of AI system outputs. Interpretability of AI system is higher, if the changes of the systems outputs in result of changing algorithmic parameters are more predictable. In other words, system interpretability is related to the extent to which a human can predict the results of AI systems based on different inputs. On the contrary, explainability is related to the extent to which a human can understand and explain (literally) the internal mechanics of an AI/machine learning system. (Choraś et al., 2020, p. 623)

Roscher et al. (2020) surveys recent ML approaches focused on the accuracy vs. explainability and accuracy vs. reliability trade-offs, in the context of using XAI to help make further advancements in the natural sciences. They differentiate the terms transparency (considers ML approach), interpretability (considers ML model and data), and explainability (considers ML model, data and human involvement). At this stage, of interest to my research is their discerning of related and often-intermixed terminology, and need to consider whether the concepts of transparency and interpretability are encompassed within explainability or not.

Bhatt et al. (2020) presents an empirical study focused on the accuracy vs. explainability trade-off, from the perspective of how organizations utilize explainability and the features of XAI for stakeholder consumption. They found that the majority of XAI's deployment is not for users to gain insights into the data and ML-treatment, but rather for model debugging purposes by system engineers. The key takeaways are helpful: not only is the accuracy vs explainability trade-off a real and well-document limitation, organizations' deployment of XAI does not seem to be currently aligned with advancing this area, and moreover, lacks important foundational elements such as differentiating domain expertise in explainable ML vs expertise in solving the underlying organization problem, and a comprehensive framework for the deployment of explainable ML in terms of why and how.

Gunning and Aha (2019a) is another in a line of significant contributions by David Gunning (2017, Gunning & Aha 2019, Gunning et al., 2019) focused on the topic of explainability. Unlike the ambiguity and lack of consensus we've seen in much of the literature concerning XAI, DARPA (The Defense Advanced Research Projects Agency), a research and development agency of the United States Department of Defense, defines XAI as "AI systems that can explain their rationale to a human user, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future" (p. 44). This a remarkable clear and understandable definition, yet the abundance of research and literature in this area suggests that its application in real world settings falls short of this goal. There is a deliberate intention to call this field explainable AI, and not have it mixed up with literal or semantical alternatives such as interpretable, comprehensible, transparent or others; presumably to capture human understanding and psychology in their XAI models. DARPA hopes to achieve truly understood, trusted, actionable and dependable AI systems through the creation of new or revised ML techniques that produce explainable models. However, the tension between predictive accuracy and explainability remains live and unsolved, and new research challenges are exposed, such as how to achieve this form of user-centric XAI, what do interfaces look like, and how to ensure XAI models meet the test of human psychology and understanding.

Holzinger et al. (2019) urges practitioners to go beyond the issue of explainability, and introduces the concept of causability in the context AI used in medicine: "causability, which is differentiated from explainability in that causability is a property of a person, while explainability is a property of a system" (p. 1). They go further in their differentiation (p. 3):

> Explainability: in a technical sense highlights decision-relevant parts of the used representations of the algorithms and active parts in the algorithmic model, that either contribute to the model accuracy on the training set, or to a specific prediction for one particular observation. It does not refer to an explicit human model.
>
> Causability: as the extent to which an explanation of a statement to a human expert achieves a specified level of causal understanding with effectiveness, efficiency and satisfaction in a specified context of use.

This supports a gap identified in Kannan et al., 2020; Lalmuanawma et al., 2020; and Valera et al., 2018, wherein they highlight the importance of AI systems being both useable and useful, ultimately to the benefit of the human users.

Explainability is a Myth. Rudin (2019) is a highly cited work focused on the trade-offs of accuracy vs. explainability and reliability. Rudin suggests we focus less on the explainability of AI systems and focus rather on the interpretability of the ML models themselves. We've seen the inherent conflict between accuracy and explainability, in that increasing the accuracy of AI systems usually comes from increased complexity, which leads to less explainability. Where XAI, or at least the concept of it, can be general and agnostic to application, Rudin argues that Interpretable AI (IAI) is domain-specific, and Interpretable ML (IML) is in some way "constrained in model form" (p. 1). While much literature underscores the trade-off between accuracy and explainability, Rudin notably argues that the trade-off of accuracy vs. interpretability is a myth, According to Rudin, it's not necessary that a greater complexity means greater accuracy, and that data can be well-structured with meaningful features in order to achieve greater accuracy without the complexity. This is an interesting notion but only holds if such a structure can be had or is known in advance during the modelling stage. Rudin's work is valuable in my analysis and stands against most of the other literature researched thus far which is focused on explainability and uses terms like interpretability either interchangeably, or as Rudin suggests, improperly. Rudin proceeds to caution users about the dangers of explanations and explainability, because by definition "explanations must be wrong" (p. 3). Her compelling argument here is that explanation models, with their own inherent level of fidelity, are models of the black-box model, and not the original model itself. I find Rudin's paper to be insightful but somewhat limited in that she makes anecdotal or ad-hoc arguments against the opposing viewpoint—in support of explainability vis-à-vis black-box models—rather than putting forward well-reasoned computational examples or arguments.

> To our knowledge, all recent review and commentary articles on this topic imply (implicitly or explicitly) that the trade-off between interpretability and accuracy generally occurs. It could be possible that there are application domains where a complete black box is required for a high stakes decision. As of yet, I have not encountered such an application... (Rudin, 2019, p. 3)

Without delving further into supporting or refuting Rudin's arguments, the "chasm" (p. 14) she sees between interpretability and explainability is of significance to this research study, and a topic I intend to specifically re-visit. For now, given the girth of supporting literature on the dimension of explainability, as opposed to the interpretability Rudin puts forward, I will retain the dimension and term, explainability.

Security, Safety and Threats to Organizations and Society. A notable development is the view and treatment of AI technologies as both friend and foe (Babcock et al., 2019; Liu et al., 2020; Szychter et al., 2018). As the former, these systems and technologies help organizations in a wide variety of ways and most often to improve in the dimensions of accuracy or speed. However, it's critical to consider malicious application of these same systems and technologies as foes, which can be deployed in a variety of ways to in some way obstruct or more profoundly damage a technically-exposed system or process. The vulnerable and attacked thing could be anything from a business or organization to and individual's health records, to a voting system, or even a car or train. The degree to which an AI system is deployed as a foe need not be dramatic or apocalyptic—even the blocking of an MLA from performing its computational purpose and providing the user's desired outputs is success for the foe. Szychter et al. (2018) focuses on the trade-off of accuracy vs. security), and see AI potentially playing both a beneficial and malicious role. In their analysis and commentary, they integrate the relatively recent consideration of IoT systems and how this dual perspective of AI can be used to enhance the former's security. In light of these two pieces of literature, the conversation now shifts away from the commonly perceived threats of so-called hi-jacked technologies to inflict harm, and expands to encompass more nuanced, less-apocalyptic and varied objective functions beyond dimensions such as accuracy—and may in fact directly counter these. Consider a malicious AI system altering the workings of an AI system used for medical diagnoses, or manipulating the stock and flow of news or information. This is the frontier of AI as friend and foe.

While the above offers a glimpse into the new frontier of AI as friend and foe, others argue that the large-scale adoption of AI technologies will lead to mass-scale erosion and damage to society. Manheim and Kaplan (2019) make such a case with respect to fairness and democracy. Their contribution focuses on the trade-offs of fairness vs. privacy and privacy vs. explainability in the context of democracy and democratic public institutions. Their premise is that AI tools will be used to "manipulate the preconditions and levers of democracy" (p. 106) and that AI threatens to erode "decisional and informational privacy" (p. 112). They discount any benefit to consumers or otherwise, and warn us about the casualties of AI: privacy, anonymity, autonomy, equality, and the rule of law; they support the notion of enhanced attention and regulation at the national level. The threat to democracy is implicit threat to fairness, as the former is a custodian of the latter. Manheim and Kaplan (2019) is useful for this research study in that it provides some evidence of democratic and institutional issues which should be considered and addressed vis-à-vis the trade-off dimensions of privacy and fairness.

We have seen some examples of the dangers and issues burdening the deployment of AI technologies in the context of organization strategic problem solving and decision making. This has not slowed the pace of development of application of these technologies in these contexts, and many argue that is in fact increasing (Agrawal et al., 2018a, 2018b; Domingos, 2015; Namaki, 2018;). Many including Papernot et al. (2016) argue that trade-offs offer an opportunity for improvement and are not to be feared.

Opportunities for Improvement. Papernot et al. (2016) focus on the trade-offs of accuracy vs. security and accuracy vs. privacy. They support the notion that there is indeed an "opposing relationship between model accuracy and resilience to adversarial manipulation" (p.

1), which is accuracy vs. privacy. They also conclude that there are potentially unavoidable tensions between model complexity—to gain accuracy—and resilience to security and privacy threats. Their discussion points to an assertion that a perfect classifier or ML system, cannot be manipulated or fall prey to an adversarial security or privacy threat. Rather, adversarial manipulation threats must try to take advantage of inaccuracies or prediction loss of ML system. The inference is that a solution to addressing these security and privacy trade-offs is to increase complexity and accuracy. This almost seems counterintuitive in the context of a 1-to-1 trade-off where the gain in one scalar dimension means the reduction of the other. I will revisit their assertion in my research study: whether there exists a model where these trade-offs can become directly correlated rather than oppose each other.

Goebel et al. (2018) is yet another paper which examines the trade-off of accuracy vs. explainability. Considering the distinction provided in Choraś et al. (2020), Goebel et al. (2018) is more interested in interpretability than explainability. To that end, they point out a key aspirational goal of AI and ML solutions is to help humans improve, and the benefits of augmenting human intelligence with AI and ML solutions.

Bansal et al. (2019) focuses on the accuracy vs. speed trade-off, although from a different perspective. In their view, humans and AI technologies form a team, “e.g. AI-advised humans” (p. 2), which has a collective performance. Their incorporate accuracy and speed into their performance metric, and posit that the discussion needs to go beyond south-after gains in accuracy, and that high[er] team performance lies at the optimal intersection of the human and the AI, and ‘knowing how and when to complement one another’ (p. 2). Specifically, it’s essential for humans to know the AI system's error boundary so that decisions can be made on when an override is necessary or not.

Are all Trade-offs Created Equal or Equally Important?

In the analysis of the set of twenty-one trade-offs, I currently hold the opinion that a simple analysis of the quantity of literature focused on any particular trade-off is not an indicator of the importance or relevance of the trade-off. That said, the quantum of work in any one area seems to reflect the level of study and presumably, current understanding of those areas. There may be a multitude of reasons for the disparity of literary coverage and publication between different trade-offs. The most obvious being that, while the terms and growing applications AI and ML and other related technologies have been around for decades, reliable empirical studies into results and outcomes of these deployments is much newer, spanning only the past decade or so. Issues around specific phenomenon like the loss of privacy required to gain accuracy (Abowd & Schmutte, 2018), or issues of fairness and reliability (Mehrabi et al., 2019; Oneto & Chiappa, 2020; Tolan et al., 2019), or the lack of explainability in the results of medical diagnoses (London, 2019; Rashidi et al., 2019; Roscher et al., 2020; Vellido, 2018) or the law (Hacker et al., 2020), are all more recent. As the pace of technology development and specifically, AI techniques, has varied and permeated society and life at different rates (Aristodemou & Tietze, 2018; Brynjolfsson & McAfee, 2014; Domingos, 2015), so does research and understanding into these and issues that may result—some temporary asymmetry seems not unusual.

I've highlighted that 52% of the trade-off literature reviewed thus far, focuses on the accuracy dimension as either the primary objective (D_O) or a corresponding trade-off (D_T). In the context of computational approaches to strategic decisions and the deployment of AI technologies such as machine learning, it's important to consider these tools as components of what may be an organization's *business or commercial* objective. Obvious examples of these may include: revenue growth through enhanced sales, or increased profitability through cost reductions, or a social welfare maximization (Abowd & Schmutte, 2018). Without getting into technical details at this stage, it would be erroneous to draw a direct connection between AI technologies and what I will call, downstream business objectives, which are impacted by many other variables than the outputs of computational tools such as AI technologies.

Therefore, based on my above reasoning and the currently published understanding of the trade-offs, the ever-growing prowess of the AI technologies deployed within industries and against problem classes, I posit and will be proceeding on the basis that *prima facie*, no particular trade-off is or should be treated, more important than another. This does not preclude organizations from selecting or limiting the trade-off dimensions and corresponding trade-offs, that may be more important to capture/maximize, or lessen/mitigate, for their particular strategic goals or problems, e.g. the importance of the dimensions accuracy-explainability-privacy in healthcare coexists, or accuracy-reliability-security in legal contexts, and so on. However, given my hypothesis that trade-off dimensions and trade-offs are explicitly or implicitly ignored or otherwise inadequately captured, I suggest that their analysis must be more comprehensive than the status quo. In other words, I am proposing that all the trade-off dimensions and trade-offs should be considered equality important in the context of this research study focused on strategic decision-making, but within this subset of all organization activities, there may be an unequal actual *or perceived* importance vis-à-vis a particular industry, function, goal or problem. In such cases, an enhanced and expanded trade-off capture should in theory, improve the objective function.

Summary of AI Technologies Applied to Organization Strategy

The conceptual beginnings of what we today call artificial intelligence has been around for over seven decades. The pace of development has been increasing steadily, and the terms machine learning and algorithms are now ubiquitous. Along this journey from the first applications through the “AI winter”, to the present time, the variety, breadth and depth of applications and problems attacked has been expanding. However, gaps still remain and the answer to the ultimate questions of whether AI techniques can reliably enter the domain of strategic decision-making remains an open issue.

If we adopt Porter’s model (1996), then outperformance vis-à-vis competitors is due to superior fit, which is in effect activity chains up and down and across an organization. Strategy formulation is a soup of optimization and prediction problems, in series and sometimes parallel, which are sometimes individually and certainly collectively, in NP and likely NP-hard. The series of papers on the NK Model and fitness landscapes (Kaufmann, 1988, 1990, 1993; Kauffman & Levin, 1987; Kauffman & Weinberger, 1989; Levinthal, 1997, 2011, Levinthal & Warglien, 1999; Rivkin, 2000; Rivkin & Siggelkow, 2003; Siggelkow & Levinthal, 2003; Weinberger, 1996) demonstrate that adaptive walks on fitness more-or-less rugged landscapes are not only NP-hard, but are also prediction and optimization problems in series and parallel. Strategic problems in competitive environments are akin to these NK landscapes, and are therefore NP-class, and often NP-hard.

Being in NP, the time to undertake and solve for a strategy formulation problem grows at greater than polynomial time, e.g. exponential time, as the number of input variables increases. Furthermore, NP-hard problems cannot currently be deterministically broken into faster-to solve

P-hard problems unless they are NP-Complete, in which case they can be verified in polynomial time but may not be solvable. It follows that NP-hard problems, strategic or otherwise, cannot be *deterministically* broken into or solved as P-hard parts.

Approaches thus far have yielded less than expected results, for a multitude of reasons including but not limited to: issues in problem selection and formulation (Moldoveanu, 2009), problem complexity (Agrawal et al, 2018b; Bettis & Hu, 2017; Cohen, 1999; Domingos, 2015; McKelvey, 1999, 2000; Moldoveanu, 2009, 2016; Moldoveanu & Bauer, 2004; Ng, 2016), and management limitations (Moldoveanu, 2009). And at the center of this study, outcome-diminishing side-effects from trade-off dimensions and the pair-wise and n-wise trade-offs between them. Optimization and prediction problems seem to be ideally-suited to machine algorithmic solutions (Brynjolfsson & McAfee, 2014; Domingos, 2015; Moldoveanu, 2019a, 2019b), but being NP-class means that they cannot be solved in polynomial time, thereby reducing computational algorithms' benefits of seemingly simultaneous gains in speed and accuracy—notwithstanding my extensive examination on these specific trade-off dimensions with respect to each other and pairwise with other dimensions.

There has been significant advancement in the underlying technologies and models used in AI technique deployments, and these developments demonstratively allow AI techniques to move from relatively rudimentary tasks such as simply predictions and suggestions, to far more complex problems and goals such as business unit virtualization, deep learning pattern articulation, and other areas of optimization. However, there remains a significant gap between even the latest most advanced AI techniques and system architectures, and their ability to move deeper into intractable problems classes such as strategic decision-making. The latter poses unique complexity challenges for algorithms because it is abstractly defined and typically performed in multi-step processes, such as: problem articulation, data gathering, analysis and selection, and so on. Fundamentally, current AI techniques struggle to create precision and order when faced with the imprecise nature of an organization strategic problem. Is the challenge in the solution, or should we *also* examine the problem differently?

Gaps and Revisiting the Research Questions

Technical Limitations

In his 2016 publication entitled “What AI Can and Can’t Do”, renowned AI practitioner Andrew Ng stresses that the actuality of AI and its current application is far short of common perceptions and the work of science fiction. AI techniques are *beginning* to resemble some human-like capabilities by emulating limited human functions through deep learning and neural networks. While Ng’s assessment that: there is currently no visible path to the higher intelligence capabilities of the human brain, and certainly nothing approached the abstract and intractable nature of strategic problem solving is from a *pre-ChatGPT world*, it largely holds true even today. Ng points to the vast data requirements as a significant shortcoming of current and envisioned AI systems—the training data systems must learn in order to be able to compute a response B from input A. Ng concludes that data is a defensible barrier and valuable asset of organizations, along with the talent to turn A into valuable and useable B—seemingly pointing back to the extensive fit arguments we’ve seen in Porter (1996) and the NK Model literature. We’ve also seen this data barrier echoed literally, through implication, and through the forward-looking models proposed by the research of Agrawal et al. (2018a), Bettis and Hu (2017), Hau

(2019), Ojanpera et al. (2018), and Moldoveanu (2016, 2019a). To this barrier, I would add computational and other resources required to render the data useable.

Improvements in the dimension of speed have been achieved, but this is yet another trade-off dimension that ebbs and flows depending on the system design and problem being attacked. In some instances, gains in accuracy and speed go seem to go hand in hand, but abstractions have been made, and albeit at the loss of another dimension such as explainability. Generally, accuracy vs. speed is an observed trade-off where these respective dimensions are inversely related (Bansal et al., 2019; Guo, 2019; Kononenko, 2001; Lalmuanawma et al., 2020; Nguyen et al., 2019; Srivastava et al., 2014; Vollmer et al., 2020).

The possibility of a gain in one dimension gain without a corresponding trade-off dimensions loss does exist, but the literature points to either a non-accounting of other trade-off(s)—like in the accuracy vs. speed scenario—or the introduction of externalities such improved processing power (Agrawal, 2018), and the natural extension to systems architecture, and necessary accommodating storage or network infrastructure. These are considered forms of abstraction while outside the scope of my analysis, may present possible avenues to address in my proposed approaches. And as I have already suggested, purely algorithmic approaches and solutions may not be the answer.

As shown in this review, strategic decision-making is a complicated and particularly difficult problem to solve from a computational perspective. Human judgment and the various considerations and risk metrics our brains are able to evaluate and balance, is nothing short of a computational marvel, and fundamentally pose algorithmic problems for the way computers and AI techniques operate. As powerful as they are, AI techniques as they currently exist still have limits. The examination of the trade-offs shows us that even where gains in dimensions such as accuracy and reliability are found, they come at the cost of other important considerations such as explainability or privacy. This presents challenges in not only computationally finding a, or the, correct answer, but trusting it enough to allow for more constrained human intervention. Another more practical issue is how to actually get AI techniques to compute answers to strategic or strategy formulation problems in a reasonable not-greater-than polynomial timeframe. Traction has been gained through deep learning and other approaches, but the time required to obtain an answer starts to rise exponentially as the amount of information grows, and high-fidelity strategy formulation usually requires a lot of information which isn't static. Human brains are particularly adept at flexible reasoning, but machines currently are not.

In light of the above unique challenges, there's an incomplete picture of the future of AI in terms of being able to address intractable problem classes such as strategic decision-making, within the realm of currently known or contemplated systems designs**.** The dialogue, and lines of inquiry, should move beyond just algorithms and a technological discussion, and allow for the inclusion, and addressing, of some other issues and considerations brought forward in this literature review, such as human-machine systems, defining reliability thresholds by problem context or class, human judgment and choice structuring, and risk management and mitigation.

The Importance of these Research Questions

> RQ1: *How can organizations use AI techniques to improve strategic decision-making?*
>
> RQ2: What trade-offs dimensions are relevant in the application of AI techniques in organization strategy processes?

RQ3: How can organizations improve strategic decision-making processes to maximize outcomes while minimizing the corresponding trade-offs?

My primary research question is RQ1, with RQ2 and RQ3 being secondary lines of inquiry. Brezillon (1999) presents an insightful view on the importance of context in the various aspects of AI, gathered from two workshops and a conference, with the goal of better understanding and utilizing "cooperative problem solving by a human and a machine" (p. 18). He highlights context vis-à-vis each of "knowledge acquisition, machine learning, communication, and databases and ontologies" (p. 1). He also draws attention to lack of explicit representation of context as being a driver of failures of many knowledge-based systems, primarily because the users of these computer systems are not accounted for. In his discussion on knowledge acquisition, there are parallels with Agrawal et al. (2018a, 2018b) and their detailed consideration of human judgment, in that Brezillon posits that "experts do not report how they reach a decision" (p. 3). He concludes that the context around an expert providing knowledge also predicates the reliance on that knowledge. I intuitively agree with Brezillon's statement that context in communication is a shared space, and each entity within a collaborative partnership should be system-provided ways to "express, explore, recognize, and negotiate their shared context" (p. 7). This also seems to foreshadow later works from several scholars: Agrawal et al. (2018a), wherein Brezillon (1999) cites Maskery et al. (1992) construction of context as an expert in predicting what the user wants or needs to do next. To link this this study, Brezillon himself connects context to problem solving—the context of a computer, machinery or logic system is that of the problem being solved, and all interactions occur within this context. Interestingly, he takes us further into the digital world with "The current context is the physical/real memory, the other contexts are the virtual memory", created dynamically by the system (p. 14). Another key finding he makes is the differentiation of model and representation: “the goal of the former is to give a coherent picture of context that can be used for explaining and predicting by simulation”; whereas the representation of context “is only to account for what is observed” (p. 16). In the context of my research question, the model becomes input and learning for the programmed/coded representation. “Context plays an important role in domains where activities implies reasoning and interpretation” (p. 18). “McCarthy [42] points out that the logical machinery is only a small fraction of the effort involved in building a context-based system” (p. 19). Amalgamating the stances and some key thrusts from Brezillon (1999) with later works from Bettis and Hu (2017), Brynjolfsson and McAfee (2014), and Domingos, (2015), and I don’t believe the importance of these AI technologies and their impact on business and indeed humanity’s future can be overstated.

Innovative Lines of Inquiry and Possible Directions

Some researchers and practitioners have gone beyond the above highlighted AI techniques and their limitations. David Gunning (2017), a program manager at the US Defence Advanced Research Projects Agency (DARPA), focuses on the concept of Explainable AI (XAI). Gunning posits that a leap must take place from the current generation of AI systems to XAI, such that users can understand, trust and manage these technological collaborations. The essential leap Gunning describes is moving from the output of deep learning with some level of predicted accuracy, to an explained model and result which allows the user to fully understand the if/when/why/why not/etc. This new XAI layer consists of an Explainable Model + Explanation Interface (p. 6). While the concept seems novel, it is in fact already under study, development and testing at various public and private institutions. Of key relevance and interest to this paper is the fact that Gunning highlights the use of XAI within an autonomous context (more advanced that analytics) within strategic problems.

Agrawal et al. (2018a) present a series of elegant mathematical models to explain and demonstrate the intersection of machine prediction and human judgement. Like in Acemoglu and Restrepo (2018), Nasi (1999), Ojanpera et al. (2018), Agrawal et al. examine what types of human labour will be replaced vs complemented by technological advances. As AI technologies become better at prediction, Agrawal et al. pose the clever question of *what happens when AI can predict human judgment?* This clever play on words has enormous impact on the topic at hand—perhaps machines cannot replicate or mimic human thinking and decision-making especially when tackling new problems, but can they learn how to predict it?

Wang et al. (2018) is a wide-ranging presentation on different frameworks and studies being undertaken by researchers and academics. Of interest here is the concept of Internet of Minds (IoM). "IoM's key tasks include knowledge acquisition, knowledge representation, knowledge exchange, and knowledge association, aiming to build semantic connectivity among intelligent social entities" (p. 902). This may be a significant development in the context of using AI for strategy formulation in that this framework allows for not only collaborative work between humans and machines but also between machines. IoM is presented as a "cyber-physical-social system" (CPSS) and "a new paradigm for intelligence cooperation" (p. 902). The CPSS presented here includes of technical systems, software/hardware, human beings, social organizations and connected things. This builds upon models and frameworks are being presented to encompass the complexities and difficulties of higher human intelligence functions.

Hao (2019) is a review of twenty-five years and 16,625 papers to figure out where AI is going. She finds three major trends: 1) machine learnings, followed by 2) neural networks, and finally, 3) reinforcement learning. While Hao's methodology is far from exhaustive in that she performed a word-search on the abstracts of these 16,625 papers, it does offer a glimpse of where at least the language of the research was, is and is going.

Generative AI Techniques Do Not Solve the Problem of Strategy Problems

Powerful generative models and techniques, and their abilities with respect to interact conversationally, generate sense-making prose, images, videos or other outputs, still does not address the complex intractable problems which are the subject of this research study. While there is an argument to be made that understanding language or being able to generate images and videos is computationally difficult, these are not considered within the same NP classes of problems for several reasons. First, the computational complexity does not necessarily increase as the type of generative context changes or number of input variables increases. Second, even with computation time increases, the time to solve does not necessarily increase at greater than polynomial time as input variables increase, i.e. NP class of problems necessarily exhibit a greater than polynomial increase in time to solve as input variables increase[25]. Third, these technologies continue to demonstrate relatively low accuracy and as I cautioned above, sense-making should not be confused with truth-telling[26,27]. Finally, while immense resources are

[25] https://www.yahoo.com/tech/sorry-chatgpt-problems-always-too-110000797.html, retrieved Aug 24, 2024.

[26] reddit.com/r/ChatGPT/comments/zt7bjc/chatgpt_is_confidently_writing_polynomial_time/, retrieved Aug 24, 2024.

[27] reddit.com/r/ProgrammerHumor/comments/zliuz1/so_i_asked_chatgpt_to_solve_p_vs_np/, retrieved Aug 24, 2024.

required and have indeed been deployed to create and train generative models, this very observation supports the notion that the trade-offs of speed vs. accuracy/privacy/others have not been computationally or otherwise addressed[28]. Rather, higher power computational and other resources are the key driver of any perceived gains in such trade-offs[29]. Holding computational resources steady, as the amount of training data or other inputs increases, the time to answer will rise—thereby indicating that the underlying trade-offs persist.

The essence of the max:min problem constructs I developed, where I propose the addition of, e.g. N(ε, I) computational resource options, has not been addressed as a variable set to address a trade-off—it has been in a way circumvented. The immense computational and storage resources required to produce results points to a brute-force approach to overcoming at least the trade-off dimension of speed. The trade-offs thus impacted include, speed vs. accuracy, speed vs. explainability, speed vs. reliability, and others. I suggest that the speed trade-off dimension has not been solved, it's notionally being masked or abstracted away.

Chapter Summary

To ground the work, an extensive review of the literature was undertaken, covering peer-reviewed and published articles, books, case studies and other publications from 1945 to 2025. I examined the nature of the problem of organization strategy, how it gets formulated, and its unique complexity characteristics. I then sought a unifying model to address the intractable optimization nature of strategic decision-making. The NK Model from biological evolution

[28] https://www.forbes.com/sites/moorinsights/2024/05/01/anthropic-dethroned-by-gemini-15-pros-1-million-token-context-window/, retrieved Nov 2, 2024.

[29] https://thetechoasis.beehiiv.com/p/openais-strawberry-googles-millionexpert-llm, retrieved Nov 2, 2024.

(Kauffman & Levin, 1987) bridges with Porter's fit (1996), and conceptually allows for the deployment of computational approaches to address strategic decision-making. Next, I examined literature on algorithms, machine learning and other AI technologies, their evolution and application in business contexts, and the recent advent of generative AI techniques.

The literature indicates: 1) the terms machine learning and artificial intelligence conceptually date back to the mid-1950s and have been in broad use in various business contexts for over 30 years; 2) regardless of known issues and risks, the application of AI technologies across domains and use-cases is ever-pervasive and will continue to play an ever-increasing role in society; 3) the pace of development has been exponential in more recent years, catalyzed by decreasing costs of computer hardware, networks and storage; 4) there has been a shift of a broad spectrum of business applications to focus from *if* to *how/when/where*, and to enabling approaches to improve efficacy, value and speed; 5) there are seven trade-off dimensions and twenty-one pairwise trade-offs between these dimensions, and 6) gains in dimensions such as accuracy or speed are achieved via the expenditure of known computational and network costs, and known and unknown non-computational costs such as the trade-off dimensions.

The next chapter will address in more detail my research methodology. This helps bind not only how the literature review was undertaken, but starts to synthesize concepts gained from the review, and the development of mathematical models to address my research questions.

Chapter 3. Research Methodology

Introduction

Based on the nature and aims of my research questions and review of the existing literature, a theoretical study best allowed for the advancement of existing groundings, or potential creation of new theory, frameworks and approaches. This methodology is also aligned with my own objectivist ontological, positivist epistemological and philosophical positions, which I discuss in more detail below. I will conclude this chapter with a discussion of the importance of this research, who will benefit from it, and how.

This study is informed by a comprehensive literature review spanning eight decades, and the findings made thereof. Research sources consisted of primary, secondary and tertiary peer-reviewed literature, empirical case studies, databases and other sources. Using theoretical and empirical literature as data, I intend to develop and propose one or more optimization models, which organizations can use to more optimally deploy AI techniques while more fully capturing the true costs and trade-offs. Empirical case studies will also be analyzed to demonstrate how the decisions taken by organizations and managers made implicit or explicit trade-offs, and lead to suboptimal results.

Research Paradigm: Ontological, Epistemological and Philosophical Positions

All knowledge and its discovery, being subjective (Scotland, 2012), it's imperative to clearly articulate my underlying positions in terms of paradigm, ontology, and epistemology, and how these define and guide my methodology and methods. Drawing upon James' empiricism and rationalism (1909), and then combining Coleridge's Platonist (rationalist) and Aristotelian (empiricist), as cited in Partington (2002), I believe my philosophical position and approach to this research study firmly lands within the realm of logical positivism or logical empiricism.

The philosophical underpinning of this research study is in the scientific paradigm, which comprises, helps define and guide my ontological, epistemological, and methodological approaches and positions (Scotland, 2012). My ontological position is one of objectivism or positivism, and at times may venture into post-positivism. Positivism is based in realism, meaning that the subject of my research, a discoverable reality, exists independent of the researcher or knower (Scotland, 2012; Wilson, 2014). The post-positivism inclination and philosophical position allows me to include concepts such as falsification rather than adopt a purely justificationist approach. My epistemological position is positivist, which is based in objectivism, meaning that I will undertake this research study impartially and treat the subject of my research as an objective reality. A positivist epistemology should support a higher level of reliability within my research, analysis and findings (Wilson, 2014). While my philosophical positions and researcher beliefs are explained above, I am open to the possibility that some of the literature, prior research and researcher, and so on, may have conducted their respective research from a subjectivist ontological basis.

A theoretical study clearly lends itself to an inductive approach wherein advancement or creation of theory can take place, and usually involves qualitative research. However, the application of known or established theoretical frameworks is more aligned with a deductive approach, and usually involves quantitative research. That said, this distinction is less clear than it may seem and given my research topic, it's important for me to be mindful of what Patton underscores:

> As evaluation fieldwork begins, the evaluation may be open to whatever emerges from the data – a discovery or inductive approach. Then, as the inquiry reveals patterns and major dimensions of interest, the evaluator will begin to focus on verifying and elucidating what appears to be emerging – a more deductive approach to data collection and analysis. (Patton, 1991, p. 194, as cited in Wilson, 2014)

While my stated goal for this research study is more aligned with quantitative research and analysis, the data and analysis phases of this study absolutely involve qualitative and perhaps even phenomenological concepts. The latter is an important consideration since I'm intimately relying on others' existing research and written outputs. My review of the extensive literature benefits from the adoption of an abductive reasoning approach. This review has led me to conclude that while some things are known, other things are not. In order to complete my research and answer my research questions, I will make certain conclusions based on what is known. From this abductive reasoning approach, I will proceed to employ deduction and induction as logically makes sense and theoretical bases are available, or can be created.

Constructs and Definitions

A review of the literature has helped explicate a set of seven (7) scalar trade-off dimensions (see Figure 2): *accuracy*, *explainability*, *fairness*, *privacy*, *reliability*, *security* and *speed*. These dimensions are deemed to be commutative or scalar because their order with respect to each other is not important, i.e. accuracy vs. reliability is considered equal to reliability vs. accuracy. Seven trade-off dimensions in turn, result in *7-choose-2* or twenty-one (21) trade-offs as a set of binary 1-to-1 trade-offs (see Table 8). These seven dimensions each have specific definitions in the context of AI technologies and the application of these technologies to organization strategy and decision making. Viewed together as a complete set of trade-offs of scalar dimensions, we get Tables 8 and 9. These tables were revisited on several occasions during the study.

Figure 2

Scalar and Commutative Representation of Seven Trade-Off Dimensions

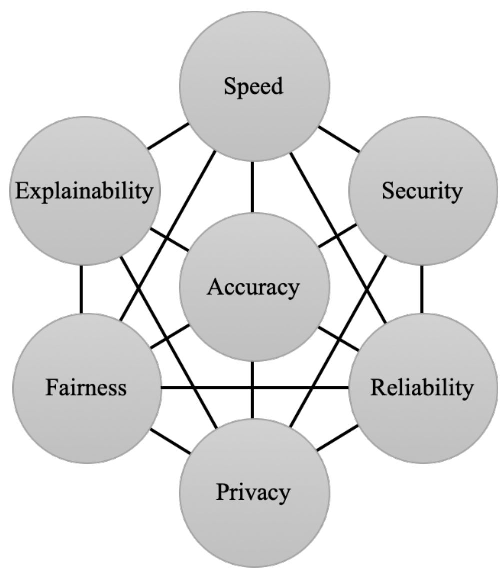


Table 8

List of Seven Scalar Trade-Off Dimensions as Twenty-One 1-to-1 Trade-Offs

Trade-off 1:	Accuracy vs. Explainability	‹—›	Explainability vs. Accuracy
Trade-off 2:	Accuracy vs. Fairness	‹—›	Fairness vs. Accuracy
Trade-off 3:	Accuracy vs. Privacy	‹—›	Privacy vs. Accuracy
Trade-off 4:	Accuracy vs. Reliability	‹—›	Reliability vs. Accuracy

Trade-off 5:	Accuracy vs. Security	‹—›	Security vs. Accuracy
Trade-off 6:	Accuracy vs. Speed	‹—›	Speed vs. Accuracy
Trade-off 7:	Explainability vs. Fairness	‹—›	Fairness vs. Explainability
Trade-off 8:	Explainability vs. Privacy	‹—›	Privacy vs. Explainability
Trade-off 9:	Explainability vs. Reliability	‹—›	Reliability vs. Explainability
Trade-off 10:	Explainability vs. Security	‹—›	Security vs. Explainability
Trade-off 11:	Explainability vs. Speed	‹—›	Speed vs. Explainability
Trade-off 12:	Fairness vs. Privacy	‹—›	Privacy vs. Fairness
Trade-off 13:	Fairness vs. Reliability	‹—›	Reliability vs. Fairness
Trade-off 14:	Fairness vs. Security	‹—›	Security vs. Fairness
Trade-off 15:	Fairness vs. Speed	‹—›	Speed vs. Fairness
Trade-off 16:	Privacy vs. Reliability	‹—›	Reliability vs. Privacy
Trade-off 17:	Privacy vs. Security	‹—›	Security vs. Privacy
Trade-off 18:	Privacy vs. Speed	‹—›	Speed vs. Privacy
Trade-off 19:	Reliability vs. Security	‹—›	Security vs. Reliability
Trade-off 20:	Reliability vs. Speed	‹—›	Speed vs. Reliability
Trade-off 21:	Security vs. Speed	‹—›	Speed vs. Security

Table 9

Matrix of Twenty-One Trade-offs, expanded from Moldoveanu (2021)

	Accuracy D_A	Explainability D_E	Fairness D_F	Privacy D_P	Reliability D_R	Security D_S	Speed D_{SP}
Accuracy D_A		Trade-off 1 T_1	Trade-off 2 T_2	Trade-off 3 T_3	Trade-off 4 T_4	Trade-off 5 T_5	Trade-off 6 T_6
Explainability D_P			Trade-off 7 T_7	Trade-off 8 T_8	Trade-off 9 T_9	Trade-off 10 T_{10}	Trade-off 11 T_{11}
Fairness D_F				Trade-off 12 T_{12}	Trade-off 13 T_{13}	Trade-off 14 T_{14}	Trade-off 15 T_{15}

Privacy D_P					Trade-off 16 T_{16}	Trade-off 17 T_{17}	Trade-off 18 T_{18}
Reliability D_R						Trade-off 19 T_{19}	Trade-off 20 T_{20}
Security D_S							Trade-off 21 T_{21}
Speed D_{SP}							

Study Design

In Chapter 2, I introduced an initial methodology and search criteria used to find relevant literature, then used explicate the trade-off dimensions. This section builds upon and completes that approach, and provides further details on sources of literature and search expressions.

Data Collection, Organization, and Management

Search criteria. An exhaustive search was performed in terms of selection of sources for my research and analysis. Works were selected or de-selected based on my assessment using the following criteria: peer-review, relevance and applicability, validity and reliability, source. No other explicit or implicit sampling technique was employed. Logical operators were also employed: an additive inclusive keyword search using ANDs and ORs was employed, and no specific criteria was created to exclude any particular article or piece of literature (no NOTs or XORs), so long as that it was determined to be: relevant to the topic and research question, peer-reviewed or from a credible source, and in English or translated into English. Literature was deemed to be credible if it was peer-reviewed or from a source with an academic connection.

No other ranking measures such as citation count, author h-index or i10-index were accounted for in the literature search and review. Cases and other forms of literature were subject to sampling based on their level of relevance to this research study and the reliability of the source. For example, the Harvard Business School library system was heavily relied on to collect some of the cases studies and empirical data pieces analyzed.

Databases. Several databases were accessed during the various data gathering phases of this research study thus far, and these and others will be continually accessed as this dissertation progresses from this ~~proposal~~ towards completion and examination. The accessed academic and online databases are:

- Google Scholar using keywords and keywords with Boolean operators
- EBSCO using keywords
- arXiv.org using keywords and logic queries
- Web of Science using keywords
- Athabasca University Library
- University of Toronto Library
- Harvard Business School Library
- MIT Press Direct: Neural Computation
- The academic databases Web of Science, Scopus, Science Direct, ASCE Library, Engineering Village, Wiley Online Library, Sage, and Emerald were used for article search and selection.

Keyword searches. Keywords and keyword searches are an important part of any data gathering, but especially important for this research study since the proposed multi-dimensional trade-off matrix is composed of important discrete concepts vis-à-vis AI and ML technologies and the focus of my research questions. As such, several logic and Boolean operators were used to gather and classify the literature, case studies and other data sources. The keyword index searches included:

- “AI” or “artificial intelligence”
- “ML” or “machine learning”
- “AI trade-off”
- “ML trade-off”

- "AI and trade-off"
- "ML and trade-off"
- "trade-off"
- "accuracy"
- "explainability"
- "fairness"
- "privacy"
- "reliability"
- "security"
- "speed"
- strategy AND artificial intelligence
- strategy formulation AND AI
- AI literature review
- expert systems
- limitations of AI
- human judgment AND AI
- artificial intelligence
- machine learning
- Boolean searches, and combinations thereof:
    - {AI OR ML} AND judgement
    - {AI OR ML} AND "human judgement"
    - {AI OR ML} AND accuracy
    - {AI OR ML} AND reliability
    - {AI OR ML} AND fairness
    - {AI OR ML} AND privacy
    - {AI OR ML} AND security
    - {AI OR ML} AND explainability
    - {AI OR ML} AND speed
- arXiv.org using keywords and logic queries such as include cross list: True; terms: AND abstract=AI; OR abstract=artificial intelligence; OR abstract=ml; OR abstract=machine learning; AND abstract=trade off

In all cases, Boolean operators were utilized along with both acronym and fully spelled out words, for example: {"AI" or "artificial intelligence" or "ML" or "machine learning"} and "accuracy", {"AI" or "artificial intelligence" or "ML" or "machine learning"} and "reliability", and so on.

- {("AI" or "artificial intelligence" or "ML" or "machine learning") and "accuracy"}
- {("AI" or "artificial intelligence" or "ML" or "machine learning") and "reliability"}
- {("AI" or "artificial intelligence" or "ML" or "machine learning") and "fairness"}
- {("AI" or "artificial intelligence" or "ML" or "machine learning") and "privacy"}
- {("AI" or "artificial intelligence" or "ML" or "machine learning") and "security"}
- {("AI" or "artificial intelligence" or "ML" or "machine learning") and "explainability"}
- {("AI" or "artificial intelligence" or "ML" or "machine learning") and "speed"}

Data Analysis

The data for my analysis is sourced from the above mentioned primary, secondary and tertiary sources. In addition, an extensive set of case studies were analyzed. These served to add examples of the processes of organizations undertaking against various strategic activities in various contexts. Examining these sources allows me to explore the problem context, problem class(es), and strategic processes employed for each. This may be beneficial for development of classification or categorization models, and whether such groupings allow for reliable model overlap across things such as industry/context, or problem class.

Approach to Mathematical Modelling

A simple approach to modelling an organization's objective function is to think of a desired outcome such as maximizing a gain *non-exclusive or* minimize a loss—either or both conditions being met may satisfy the objective. While objective functions indeed vary, the literature reviewed demonstrates that the majority of AI technology deployments seek gains in two particular dimensions: 1) accuracy (Gunning & Aha, 2019; Pessach & Shmueli, 2022; others), and 2) speed (Guo, 2019; Lalmuanawma et al., 2020; others). Expanding this simple model, each of the twenty-one pairwise trade-offs contains an objective dimension along with its corresponding trade-off dimension. For example, in the accuracy vs. explainability trade-off, the objective is to increase accuracy while the trade-off is explainability, or put another way, the gain in accuracy is associated with a loss in explainability.

This logic is supported in the literature and from a mathematical perspective, and aligns with the accuracy vs. privacy trade-off in Abowd and Schmutte (2018), in their study of governmental activities to maximize census statistical accuracy while mitigating privacy loss within the surveyed population. Following their approach for my pairwise 1-to-1 trade-offs: for every measure of marginal or differential gain in dimension 1 (e.g. accuracy *I*), how much marginal loss is there of dimension 2 (e.g. explainability ε)? These are scalar dimensions so the trade-off and its corresponding objective dimension can also be to maximize explainability while minimizing accuracy loss. This same pairwise differential gain/loss construct can be utilized for each of the twenty-one trade-offs. There are in fact five (5) forms of the objective, and the goal is to develop models for finding and achieving locations where optima or maxima may exist on the production frontier:

$$
\begin{aligned}
&Max\ \{Accuracy\}\ s.t.\ Min\ \{Explainability\ Loss\}\\
&\qquad = Max\ \{Explainability\}\ s.t.\ Min\ \{Accuracy\ Loss\}\\
&\qquad = Max\ \{Accuracy\}\ s.t.\ Max\ \{Explainability\}\\
&\qquad = Min\ \{Accuracy\ Loss\}\ s.t.\ Min\ \{Explainabiity\ Loss\}\\
&\qquad = Min\ \{Expainability\ Loss\}\ s.t.\ Min\ \{Accuracy\ Loss\}
\end{aligned}
$$

Logically, all of the above forms of the objective function can be shown to provide a consistent goal. But do they yield the same result in terms of the precise function shape and more importantly, the existence and location optima or maxima? There may be divergence here, because a specific objective function of an organization may contain specific requirements, variables or thresholds that alter the shape of a particular curve. A particular objective function, while retaining the primary goal of achieving Max{Accuracy} may set other parameters or variables such that the level of Accuracy must be at least a set target Θ, or the level of acceptable Min{Explainability Loss} may be at most a set target Φ. These thresholds may in turn alter the location of any optima or maxima. Abowd and Schmutte (2018) reach a similar conclusion in that there is no general solution for their transformation function $G(\varepsilon, I)$ where ε = privacy loss and I = level of statistical accuracy. In their analysis, $G(\varepsilon, I)$ represents a proxy for the closed set of production activities available to the organization, such as the publication of private data gathered by statistical agencies:

$$Y = \{(\varepsilon, I) \mid \varepsilon > 0, I < 0 \text{ s.t. } G(\varepsilon, I) \leq 0\}$$

$$PF = \{(\varepsilon, I) \mid \varepsilon > 0, I < 0 \text{ s.t. } G(\varepsilon, I) = 0\}$$

In their analysis, the production frontier provides the "implicit functional relationship between ε and I", "but its direct implementation is problematic. Given the current state of knowledge ... there is no general solution for $G(\varepsilon, I)$" (p. 16). Support for their conclusion can be found at Abowd and Schmutte (2018, pp. 21, App. 2-6). Similarly for my analysis, a PF should exist whereby a non-perfect level of accuracy, $I < 0$ is associated with some level of privacy loss $\varepsilon > 0$, while "obtaining perfect accuracy (I = 0) requires infinite privacy loss ($\varepsilon = \infty$)" (Abowd & Schmutte, 2018, p. 16). Whereas in Abowd and Schmutte (2018), the mitigating transformation function is the publication of this data within a closed set of available production activities, in my analysis the critical variables of accuracy and privacy loss, for example, are dependant variables within a larger production function which likely contains other independent variables that the organization should be able to manipulate in order to land on optimal combinations of accuracy and privacy loss. Continuing the use of ε = privacy loss and I = level of accuracy, consider a simplistic hypothetical scenario where an organization has before it the set of $G(\varepsilon, I)$ production activities, $M(\varepsilon, I)$ machine learning or algorithmic options and $N(\varepsilon, I)$ computational resource options. For the time being at least, I will consider M and N as any set of independent actions available to the organization so long as they are mutually exclusive of the primary trade-off dimensions. Now, the production function PF becomes:

$$Z = \{(\varepsilon, I) \mid \varepsilon > 0, I < 0 \text{ s.t. } G(\varepsilon, I) \leq 0, M(\varepsilon, I) \leq 0, N(\varepsilon, I) \leq 0\}$$

$$PF = \{(\varepsilon, I) \mid \varepsilon > 0, I < 0 \text{ s.t. } G(\varepsilon, I) = 0, M(\varepsilon, I) = 0, N(\varepsilon, I) = 0\}$$

By decoupling other choices from a broad closed set of activities $G(\varepsilon, I)$, this expanded PF contains additional available choices the organization can manipulate to try and reach an optimal or maximum combination of variables on the production frontier.

My suggested approach allows for an expanded set of variables in pursuit an organization's goals, so long as these are discrete choices that an organization can make and they are mutually exclusive to the 1-to-1 trade-off dimensions within the trade-off. This explicit recognition and addition of discrete variables within the realm of contemplation and selection set for organizations, is the first step in the development of new models to broaden the variables captured by organizations in their objective functions—the goal being that a more comprehensive variable set mitigates against explicit or implicit non-accounting for variables. In addition, rather than define the closed activity set as a general function $G(\varepsilon, I)$ without a general solution (Abowd & Schmutte, 2018), I intend to introduce a set of polynomial and power functions of specific order, which I will develop to capture existing empirical findings in the literature.

Does this Vary by Organization, Industry or Functional Area?

I suggest that while organization goals and objectives vary based on industry, competitive environment, or intra-organizational functional areas, the objective functions and trade-offs will not be so different from each other to render them inflexible. This is supported by my finding that the fundamental trade-off dimensions and trade-offs do not change amongst the various industries, functions or cases within the reviewed literature. For example, while some literature focuses on a particular trade-off or industry/functional context (Babcock et al., 2019; London, 2019), other research examines trade-off(s) across domains (Adadi & Berrada, 2018; Bhatt et al., 2020; Rai, 2020). For example, the trade-off of accuracy vs. explainability is observed in the context of healthcare decisions and medical diagnoses (Amann et al., 2020; Arrieta et al., 2019;), legal issues (Adadi & Berrada, 2018; Hacker et al., 2020), and policy-making (Choraś et al., 2020; Gerke et al., 2020). Likewise, the trade-off of accuracy vs. privacy stays largely consistent across public policy (Abowd & Schmutte, 2018), consumer recommender systems (Mazeh & Shmueli, 2019), systems security (Papernot et al., 2016), sales and marketing (Agrawal & Haritsa, 2004)), and social media data privacy (Liu et al., 2020).

My detailed exploration of these trade-off dimensions and the pairwise trade-offs between them, in the context of AI deployments within different organizations, industries and functional areas, expands the breadth and depth of study—it broadens the exploration to what I consider a mutually exclusive and collectively exhaustive set of trade-off dimensions, while also deepening the examination of these specifically to the deployment of AI technologies and techniques. Virtually all studies until now have focused on a limited set of industries, and usually a small subset of the trade-off dimensions. I am not aware of any study that has expanded the research across as wide a set of explicated dimensions and the trade-offs between them, with comprehensive contextual analysis by industry and functional area. My proposed delineation across industries and functional areas should be particularly interesting to strategists and consultants, who often structure their offerings in similar categories[30].

Validity and Reliability

I am confident in the validity and reliability of my research study, first because it is largely based on an examination of peer-reviewed and published works of scholars and practitioners. Second, the proposed dimensions and trade-offs appear to be complete and no literature has been found which speaks to other dimensions or trade-offs in the context of ML and AI technologies applied to organization strategy.

Speaking directly to the validity and reliability question, my proposed models and approaches are based on a synthesis of several established underlying models. The results of simulation testing and associated assessments are presented below.

Study Limitations, Delimitating Boundaries and Implications

This is a theoretical research study and builds upon on existing literature and research. The ever-pervasive development and deployment of many AI techniques and the associated trade-offs and challenges I've explored, is a very active research and development area. The volume of work being undertaken leaves open the possibility that some important contributions

[30] https://www.mckinsey.com/, https://www.bcg.com/, https://www2.deloitte.com/ca/en/pages/about-deloitte/articles/home.html, https://www.accenture.com/ca-en/services, https://www.ibm.com/consulting; all retrieved Feb. 24, 2025.

that may have be under development were not captured. So while my reliance on published peer-reviewed works, rather than primary empirical methods, in some ways adds to the reliability and validity of my results, it is possible that relevant considerations were not captured.

This research study covers an expansive, deep and challenging topic. If one considers the practitioner work and research being undertaken in this topic area, one may consider any single trade-off area as a stand-alone dissertation topic and study—as also empirically seen in my review of the literature wherein most studies have examined a small subset of this research area and the known trade-offs, and within limited contexts. While I did explore all of the explicated trade-off dimensions and trade-offs between them across multiple contexts, in the end I focused on a smaller set of trade-offs that were supportable by my proposed approaches, models and simulations. I believe this study makes several contributions to this topic area and allows for expansion into narrower or adjacent research areas. The results of my research are actionable and should lead to better decision-making and outcomes, and work to mitigate unaccounted-for considerations that have thus far led to suboptimal results for organizations.

From a research outcomes perspective, improvements in the utility and efficacy of AI techniques offer the promise of improved decision-making, decisions, and outcomes. As some relatively recent well-documented examples from the public health context: 1) IBM applied used AI techniques to identify of human antigens and drug compounds in the fight against COVID-19[31], 2) both China and Israel were able to track and analyze the movement and spread of

[31] https://newsroom.ibm.com/US-Dept-of-Energy-Brings-the-Worlds-Most-Powerful-Supercomputer-the-IBM-POWER9-based-Summit-Into-the-Fight-Against-COVID-19/, https://www.zdnet.com/article/ibm-summit-supercomputer-joins-fight-against-covid-19/; both retrieved March 21, 2020.

COVID-19 within their populations[32]. These are examples with profound impacts from a public policy, management and societal impact perspective. More broadly, this research study adds to the discourse and help answer how AI techniques can be applied to improve economic, social, cultural and intellectual well-being.

The benefits of this research study traverse scholarly and societal domains. On the former, I advance the base of knowledge and learning, address gaps in the literature and suggest future research directions. The movement of much scholarly work on this topic to profit-motivated settings has naturally pushed the frontier into, in my submission, towards more technical aspects and perhaps away from an organization strategy poster. My work allows for what I believe to be comprehensive and unbiased advancement in the latter fields of study. The potential benefits are also multi-faceted and include economic via better use of resources, general well-being via better policy, and democratizing via vastly improved analysis of data and inputs—leading to more robust outcomes. Establishment of the limits of known or *contemplated* AI techniques vis-à-vis classes of problems, and improvements in their utility and efficacy, allows organizations, governments and policy makers, and society at large to expend scarce resources for the widest benefit. This is not mutually exclusive from the work being undertaken by commercial actors for their own purposes and benefits, but does allow for potentially greater alignment on outcomes sought by social actors.

[32] https://www.forbes.com/sites/bernardmarr/2020/03/13/coronavirus-how-artificial-intelligence-data-science-and-technology-is-used-to-fight-the-pandemic/#72f8af45f5fc/, https://www.forbes.com/sites/zakdoffman/2020/03/14/coronavirus-spy-apps-israel-joins-iran-and-china-tracking-citizens-smartphones-to-fight-covid-19/#69a514f3781b/; both retrieved March 21, 2020.

The literature has shown that AI technologies have the *potential* to transform medical diagnoses, public policy, legal frameworks, and consumer privacy, to name a few examples. In addition to the trade-off dimensions and trade-offs I've introduced in detail, studies like this one also have the potential to address lingering issues such as socioeconomic impacts, domain or problem specific adoption issues, and the nebulous concept of trust in these systems and the outcomes they produce. I'll leave an analysis of trust to other researchers but in the current context, I suggest that it is at least composed of several other dimensions such as explainability, fairness, privacy, and reliability. Approached and models that allow for improvements in these dimensions should engender greater trust. This should in turn enable greater adoption, which may produce a virtuous cycle through continued improvement and application.

Ethical Considerations

This research study did not involve primary research on human subjects and as such, there are no deemed ethical issues or considerations.

Chapter Summary

My research paradigm is logical positivism or empiricism, ontological position is objectivism or positivism, and epistemological position is positivist. The seven trade-off dimensions explicated from the literature, form twenty scalar pairwise trade-offs, which I placed into a matrix. I described in detail my non-sampled approach to gather sources of primary, secondary and tertiary literature. Preliminary modeling of mathematics constructs indicates that there are five forms or representations of each trade-off. Finally, I highlighted that this theoretical research study builds upon, and is limited by, existing literature and research.

Chapter 4. Maximizing Fitness of a Production Function

Introduction

In this chapter, I will develop approaches that organizations can deploy and utilize to maximize their outcomes while minimizing corresponding trade-offs on the production frontier. To achieve this type of optimization, I will build upon economic and other discipline models and approaches, in the treatment and analysis of the trade-offs described in detail above. I will build upon extensive previous works and application of the NK Model to management science and strategic problems (Kauffman & Levin. 1987; Kauffman, 1988, 1993; Weinberger, 1990; Weinberger, 1996; Levinthal, 1997, 2011; Levinthal & Warglien, 1999; Rivkin, 2000; Rivkin & Siggelkow, 2003; Siggelkow & Levinthal, 2003), and recent advancements in RL-based computational approaches (Bengio et al., 2021; Bengio et al., 2023; He, 2024). To advance more directly into the realm of computational intervention, I will integrate and synthesize concepts from evolutionary biology, organization and management science, economics, mathematics and computer science, as interconnected and propose interoperability. A multi-step approach is developed and proposed to enhance processes to maximize objective and production functions, involving incorporating additional variables, problem modeling based on NK and other models, and inclusion of different AI technique interventions.

My approach to constructing and evaluating mathematical models builds upon the important foundational work of Abowd and Schmutte (2018), and their analysis of the trade-off between Accuracy (I) and Privacy Loss (ε) in the context of protecting and using census and statistical data in the United States. In their analysis, Accuracy (I) is the objective dimension while Privacy Loss (ε) is the trade-off dimension. Their production function (PF) intersects an objective Social Welfare Function (SWF) at a maximum *O(ε, I)* (see Figure 3 below). Their approach to the planner's problem[33] is a maximization—or optimization we we've seen above—of some social utility with some finite constraints (Williamson, 1999). The planning problem is an allocational or assignment problem due to the organization having to face a trade-off in the way it allocates finite resource(s) to achieve an output. The best allocation of resource(s) equates to the allocation that leads to the maximum utility function. In Abowd and Schmutte (2018), it is the maximization of accuracy given some amount of privacy loss within the set of *G*(*ε, I*) statistical agency activities. Economic concepts such as production and objective functions, and maximizations, ғ are also important in my analysis. These concepts help to lay the foundation for trade-off analysis and models organizations may use to alter inputs to try and achieve greater outputs. In the content of this study, this directly speaks to actors such as organizations, making decisions related to input variables, within contexts and with constraints, to produce outputs and achieve outcomes. To that end, I will briefly define and integrate various concepts and terminology from economics, the NK Model, mathematics, computer science and management science. This convergence of theoretical and empirical grounding will become important as I proceed with model development, and simulations and testing.

In the context of computer science, a rugged landscape (Kauffman & Levin, 1987) is the set of allowable configurations in some optimization problem (Weinberger, 1990). The optimization problem at hand, is the maximization of a production or objective function, a MAX{D_O}. This now allows for my proposed M(ε, *I*) and N(ε, *I*), where I've considered M(*ε*, *I*) as the closed set of machine learning options, and N(*ε*, *I*) as the closed set of computational

[33] https://en.wikipedia.org/wiki/Social_planner, retrieved May 2, 2024.

resource options. I want to search the vicinity of a local maxima, e.g. O(ε, I) from Abowd & Schmutte (2018), to find a higher peak on the PF. In their analysis, they posit a $G(\varepsilon, I)$ closed set of production activities, where ε = privacy loss and I = level of accuracy. I've generalized and expanded this to M(i,j) and N(i.j), which respectively form sets of independent actions available to the organization so long as they are mutually exclusive of the primary trade-off dimensions. Now, the production function PF from Abowd and Schmutte (2018) becomes:

$$Z = \{(\varepsilon, I) \mid \varepsilon > 0, I < 0 \text{ s.t. } G(\varepsilon, I) \leq 0, M(\varepsilon, I) \leq 0, N(\varepsilon, I) \leq 0\}$$

$$PF = \{(\varepsilon, I) \mid \varepsilon > 0, I < 0 \text{ s.t. } G(\varepsilon, I) = 0, M(\varepsilon, I) = 0, N(\varepsilon, I) = 0\}$$

By decoupling other choices from a broad closed set of activities G(ε, I), this expanded PF contains additional available choices the organization can manipulate to try and reach an optimal or maximum combination of variables on the production frontier.

Having expanded the articulation of dimensions, I seek to develop optimizations which allow for the improved capture of gains in objective dimensions, while mitigating the trade-off dimension. What does this mean? On a simple 2-axis Cartesian plane, this means that either the PF changes shape or slope such that any point of interest, a local maximum O*(ε, I) is of greater than or equal-to magnitude y-variable (e.g. accuracy), or less than or equal-to magnitude x-variable (e.g. privacy loss). Or alternately, on a same slope PF, the point O*(ε, I) resides at a location of equal-to or higher *positive* magnitude y or equal-to or lower magnitude x, or both. Or, both of these occur. This later depends on the shape of the PF and as I have suggested, only a PF of order three is most practical and interesting for this analysis—higher order functions create untenable situations, and lower order functions as less interesting due to the loss of at least the second derivative.

An exploration of interventions that can be undertaken to drive a higher O*(ε, I) raises several questions. What is the relationship between my proposed D_O and an economic PF, such as in Abowd and Schmutte (2018)? Above, I describe and comment on this, and suggest that the PF in my analysis should capture additional factors and variables, in addition to a desired dimension D_O. Therefore, I propose a prospective PF that captures not only D_O and D_T, but also groups of other independent variables: M(i,j) and N(i.j). *Where* and *when* are interventions via different ML technologies and approaches applicable? In order to dive deeper into this exploration, some analysis of mathematical points of interest is helpful and contained below.

Constructing Mathematical Models

Building on the work of Abowd and Schmutte (2018) and others, I will construct and develop models for organization objective functions and production frontiers for the trade-offs introduced in the Chapter 2. We've seen in the literature that some trade-offs have been more widely studied and are perhaps better understood than others, and a lack of empirical study in an area may limit my ability to develop a model for that trade-off, and/or offer some form of test or support for its reliability and validity. We have also seen that the majority of the trade-off literature highlight gains in accuracy or speed as primary organization objectives. Therefore, it is possible that the number of models may be further reduced if I can mathematically demonstrate that relevant dimensions of a particular trade-off have been captured within another model. It is also possible that specific problem/objective classes exist whereby variations of models germinate, thereby adding to the final set of proposed models.

Before going further, it's helpful to summarize and lay a mathematical foundation from Abowd and Schmutte (2018). They posit that an increase in privacy or a reduction in privacy loss

(ε), the objective in their analysis, *requires* a reduction in accuracy (I). Deductively them, given that these are scalar dimensions, an increase in accuracy would require a reduction in privacy—or increase in privacy loss. In their analysis the trade-off of accuracy vs. privacy is a resource allocation problem and hence, they propose an economic solution: “operate where the marginal cost of increasing privacy equals the marginal benefit” (p. 1). Mathematically, they are proposing to operate where the differential loss in privacy equals the differential gain in accuracy, or vice versa. Their key findings are summarized below:

$$d\varepsilon = dI$$

$$\frac{dI}{d\varepsilon} > 0$$

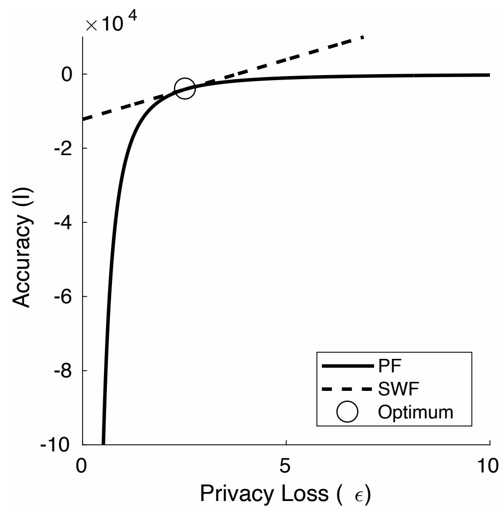
$$\frac{d^2I}{d\varepsilon^2} < 0$$

$$\rightarrow increasing\ (positive)\ quadratic\ concave\ function$$
$$where\ \epsilon > 0\ and\ I < 0$$

The change in sign from positive to negative from the first to second derivative is an important finding because it allows for the identification of a location on the production function where an optimum exists. This O(ε, I) is where a Social Welfare Function (SWF) intersects the production function (PF), and I refer the reader to Abowd and Schmutte (2018) for more detail on this (visualized in Figure 3):

Figure 3

Accuracy (I) vs. Privacy Loss (ε), from Abowd & Schmutte (2018)



$$Y = \{(\varepsilon, I) \mid \varepsilon > 0, I < 0 \text{ s.t. } G(\varepsilon, I) \leq 0\}$$

$$PF = \{(\varepsilon, I) \mid \varepsilon > 0, I < 0 \text{ s.t. } G(\varepsilon, I) = 0\}$$

$$\text{Optimum O}(\varepsilon, I) \text{ where SWF}(\varepsilon, I) = \text{PF}(\varepsilon, I)$$

$$\text{At O}(\varepsilon, I),\ \text{G}'(\varepsilon, I) = \text{SWF}(\varepsilon, I)$$

At O(ε, I), the marginal rate of transformation, which measures the cost of increased loss of privacy, matches the social willingness to accept privacy loss, both measured in units of increased statistical accuracy (p. 2).

Abowd and Schmutte (2018) propose an interesting and helpful concept, which states that databases of information, a proxy for Max{D_A, are an endowment of potential privacy loss, a Min{D_P} (see Table 11). And, some finite amount of this endowed privacy loss must be expended—privacy is given up or lost—in order for statistical agencies to undertaker their production activities, such as “to produce population statistics of any accuracy” (p. 16). They suggest that there is a non-zero loss of privacy for any non-zero gain in accuracy. While they allow for no production activity on behalf of the organization, thereby reducing privacy loss to zero, or approach such a limit, they quickly eliminate this absurd notion by positing that obtaining any non-trivial level of accuracy requires some non-zero privacy loss. Abowd and Schmutte (2018) have thus outlined their *no free lunch*, a common theme that runs through the reviewed literature and other studies of trade-offs.

Many of the concepts and treatment of variables in Abowd & Schmutte (2018) are beneficial to this study as I consider important feasible intersection-/of interest-points and the production frontiers these reside on. The idea of an endowment aligns with one or more variables which account for computational resources, such as CPU, memory, and network. However, these variables stand apart from the trade-off dimensions and are not themselves one of the seven (7) explicated here. I've above introduced the concept of variables such as M(ε, I) and N(ε, I), which account for machine learning algorithm options and computational resource options, respectively. These are but two sets of independent variables, and stand mutually exclusive from the trade-off dimensions. The inclusion of an expanded set of independent and dependent variables is critical for organizations to be able to more fully incorporate relevant variables to capture their choices (input or independent variables), and those that may be impacted (output or dependent variables). As we've seen in virtually all of the trade-off literature, it's insufficient to simply apply an AI technology or ML algorithm against a dataset in the hopes of gaining accuracy or speed—the most common objective dimensions. While there may be observed gains in either dimension, there are real and measurable trade-offs associated with, or 'resulting from' (Abowd & Schmutte, 2018), the explicit or implicit non-accounting of the *traded-off* dimension.

Several approaches have been put forward for modelling one or more of the trade-offs that I'm examining in this study (Agrawal & Haritsa, 2004; Arrieta et al., 2019; Foulds et al., 2019; Guo, 2019; Holzinger, 2018; Holzinger et al., 2019; Liu & Vicente, 2021; Lu & Shen, 2015; Mehrabi et al., 2019; Oneto & Chiappa, 2020; Papernot et al., 2016; Rudin, 2019; Valdivia, 2020; Valera et al., 2018; Xu et al., 2021; Zliobaite, 2015). Most studies focus on a singular trade-off or at most a small number of trade-offs dimensions. While the research and practitioner communities have known about *some* of the trade-offs for *some time*, the literature also indicates that there is still a relative newness to some of these concepts, with most of the literature in these latter areas has been published within the last couple of years. This is especially true for literature focused on how organizations can specifically address and mitigate some of the trade-offs. While I will take into account the work of all of these studies, Abowd and Schmutte (2018) is particularly interesting because it not only explicitly lays out a direct *max:min* trade-off between two dimensions via a mathematical model, it moves to identify a local maximum where the gain in the objective function (accuracy) beyond a certain point results in a *greater-than-gain* loss in the *opposing* trade-off (privacy). In other words, all points to the right of O(ε, II) on their PF result in Privacy Loss ε > Accuracy I, thereby rendering O(ε, I) the point of interest after which there are diminishing returns on their PF. In economic terms, the marginal gain in accuracy to the right of O(ε, I) is less than the marginal loss in privacy.

I incorporate a number of approaches to expanding on the existing models in the literature. First, I will incorporate the *opposing* trade-off dimension into the objective functions, based on the supported notion that there is no free lunch and gains in any dimension will lead to a measurable loss in the opposing dimension in that trade-off. We've already seen that this concept of a trade-off is often deliberately or otherwise not accounted-for by organizations. As a first improvement, these must be explicitly incorporated into objective functions of AI deployment models. Second, I will incorporate variables mutually exclusive from the trade-off dimensions, such as: CPU, memory, network, algorithm choice, and so on. These are my proposed variable sets M(ε, I) and N(ε, I). The purpose here is to introduce a more comprehensive action set, that organizations can effectively dial-in to mitigate for unintended trade-offs or outcomes. Third, and once there is a working set of mathematical models to address the set of trade-offs, I will propose important locations on production frontiers that organizations can *program* into their AI deployments as location(s) of possible maxima or optima, such that empirical data around these regions on the production frontier can be flagged and monitored. Fourth and finally, I hope to present a general framework that organizations can utilize to develop and enhance their own models, based on their objective functions and presumptive need to include different dimensions, metrics, and thresholds.

Foundational Constructs

Without turning this into a discourse on mathematical functions and their respective graphical plot curves, I want to set some directions on what my objective functions and production frontiers will likely look like. Given that these do and will apply to real-world non-hypothetical organization problems with expectation of a solution, I'll start by including some possible mathematical functions, while eliminating functions that are unfeasible in my analysis. I hypothesize that the production frontier for the set of twenty-one 1-to-1 trade-offs and the corresponding production function for each, takes the shape of a curve of at most 3rd order polynomial, and within this class of functions and curves, it is either a cubic curve, cubic polynomial, or a logarithmic function. Following the direction of the 1-to-1 trade-off in Abowd and Schmutte (2018), there seems to be a reasonable likelihood that the right-of or post-optimal portion, e.g. O(ε, I), of the curve will be shaped concave down, such that the gain in the objective dimension results in a loss in the trade-off dimension. I posit that at most, an organization's production function will be a 3rd order function; introducing 4th order (quartic) or higher in the set of possible production functions creates the unfeasible possibility that the function may cross back onto itself or yield the same limit when the argument goes positive or negative, have multiple real roots (crossings of the x-axis), or have two or more inflection points [34,35,36]. These types of 4th order curve characteristics are untenable in the present construct of these dimensions and trade-offs. Juxtaposing on the accuracy vs. privacy trade-off, a gain in accuracy with a corresponding required loss in privacy cannot invert back on itself under real-world constraints such that $\varepsilon > 0$ and $I < 0$ (Abowd & Schmutte, 2018). In addition, while not ruling out linear 1st order or 2nd order quadratic functions, 3rd order or logarithmic functions allow for a non-constant second derivative, which is an important element to maintain for at least the time being, as I develop models to compute the existence and locations for potential optima or maxima on the production frontier. A constant second derivative may be problematic in terms of adequately allowing for the discovery of important locations on the curve, such as inflection point(s), vertex(ices), or cuspidal point(s), to name a few (Bronshtein et al., 2007). While not specifically labelled as such, the $O(\varepsilon, I)$ in Abowd and Schmutte (2018) is akin to an inflection point as this is the point at which the second derivative changes sign. This is a necessary albeit insufficient condition for a point on a curve to be an inflection point (Bronshtein et al., 2007). The sufficient condition requires that the second derivative on "either side of x in a neighborhood of x", $f''(x_0 + \alpha)$ and $f''(x_0 - \alpha)$, have opposite signs (Bronshtein et al., 2007, p. 231). This condition is met at $O(\varepsilon, I)$. Alternately, if at a point the second derivative no longer exists or does not change sign,

[34] https://en.wikipedia.org/wiki/Gallery_of_curves, retrieved Jan 25, 2022.
[35] https://en.wikipedia.org/wiki/Cubic_function, retrieved Jan 25, 2022.
[36] https://en.wikipedia.org/wiki/Quartic_function, retrieved Jan 25, 2022.

then it is called an undulation point and may also be of interest in the analysis (Bronshtein et al., 2007). Of importance to this study is the behaviour of the curve to the left-of and right-of this critical point on the x-axis, where we observe a gain in one dimension with a loss in the other. Such a point is an important consideration in a curve defined by a linear, quadratic, or cubic function.

Based on the above, I can construct a set of relatively simple possible objective functions. General forms of the feasible set of 1st order, 2nd order, 3rd order, power and logarithmic functions, along with their first and second derivatives, are as follows:

1st order linear function: $f(x) = ax + b$

$f'(x) = \frac{df}{dx} = a$ where $a<>0$ and $a<0$

$f''(x) = \frac{d^2f}{dx^2} = 0$

2nd order quadratic function: $f(x) = ax^2 + bx + c$ where $a<>0$, $a>0$ on increasing D_1 and $a<0$ on decreasing D_2

$f'(x) = \frac{df}{dx} = 2ax + b$

$f''(x) = \frac{d^2f}{dx^2} = 2a$

3rd order cubic function: $f(x) = ax^3 + bx^2 + cx + d$

$f'(x) = \frac{df}{dx} = f' = 3ax^2 + 2bx + c$

$$f''(x) = \frac{d^2f}{dx^2} = 6\text{ax} + 2\text{b}$$

General power function: $\text{f(x)} = \text{ax}^\text{b}$ where $0<\text{b}\leq 3$

$$f'(x) = \frac{df}{dx} = \text{bax}$$

$$f''(x) = \frac{d^2f}{dx^2} = \text{ba}$$

General logarithmic function: $\text{f(x)} = \text{a ln (x)} + \text{b}$

$$f'(x) = \frac{a}{x}\frac{df}{dx} = \frac{a}{f(x)}f'(x)$$

$$f''(x) = \frac{d^2f}{dx^2} = -\frac{a}{x^2} = \frac{df}{dx}\frac{a}{x}$$

The first derivative ($f'(x)$ or df/dx) serves to provide the rate of change or slope of a function at any point on the curve *f(x)*, while the second derivative ($f''(x)$ or d^2x/dx^2) serves to provide the rate of change or slope of the first derivative at any point on the curve *f'(x)*. As we saw in Abowd and Schmutte (2018), the location at which their production frontier is optimal O(ε, *I*), *f'(x)* is decreasing in magnitude but remains positive, while *f''(x)* becomes negative. This is exemplary of an increasing (positive) concave function (Bronshtein et al., 2007). Expanding to a more general form, a potential objective function in simple linear form is:

$$Z(x) = c_1x_1 + c_2x_2 + c_nx_n$$

or in a more interesting non-linear form:

$$Z(x) = c_1^{e_1}x_1 + c_2^{e_2}x_2 + c_n^{e_n}x_n$$

$$Z'(x) = \frac{dZ}{dx} = e_1 c_1 x_1^{(e_1-1)} + e_2 c_2 x_2^{(e_2-1)} + \cdots + e_n c_n x_n^{(e_n-1)}$$

$$Z''(x) = \frac{d^2Z}{dx^2} = (e_1^2 - e_1) c_1 x_1^{(e_1-2)} + (e_2^2 - e_2) c_2 x_2^{(e_2-2)} + \cdots + (e_n^2 - e_n) c_n x_n^{(e_n-2)}$$

and more generally:

$$Z(x) = f(c_n^m x_n^m)$$

$$Z'(x) = m c_n^{m-1}\, x_n^{m-1}$$

$$Z''(x) = m^2 c_n^{m-2} 2 x_n^{m-2}$$

where {c} is a set of real coefficients
and {x} is a set of organization activities

Returning to the trade-off dimensions and trade-offs between them, some examples of incremental objective functions are:

$\text{Accuracy}_{\text{GAIN}}$ s.t. MAX{$\text{Explainability}_{\text{GAIN}}$} or MIN{$\text{Explainability}_{\text{LOSS}}$}

$\text{Accuracy}_{\text{GAIN}}$ s.t. MAX{$\text{Privacy}_{\text{GAIN}}$} or MIN{$\text{Privacy}_{\text{LOSS}}$}

$\text{Speed}_{\text{GAIN}}$ s.t. MAX{$\text{Reliability}_{\text{GAIN}}$} or MIN{$\text{Reliability}_{\text{LOSS}}$}

$\text{Accuracy}_{\text{GAIN}}$ s.t. MAX{$\text{Fairness}_{\text{GAIN}}$} or MIN{$\text{Fairness}_{\text{LOSS}}$}

I propose that there are five (5) logically equivalent ways to express an objective maximization function between two scalar dimensions (${}_4C_2$-1 = 5), to maximize the gain of one dimension and minimize the loss of another dimension as a 1-to-1 trade-off, for example the trade-off between accuracy and explainability:

MAX{$\text{Accuracy}_{\text{GAIN}}$} ∩ MIN{$\text{Explainability}_{\text{LOSS}}$}

≡ MAX{$\text{Explainability}_{\text{GAIN}}$} ∩ MIN{$\text{Accuracy}_{\text{LOSS}}$}

≡ MAX{$\text{Accuracy}_{\text{GAIN}}$} ∩ MAX{$\text{Explainability}_{\text{GAIN}}$}

$\equiv$ MIN{$\text{Accuracy}_{\text{LOSS}}$} ∩ MIN{$\text{Explainability}_{\text{LOSS}}$}

$\equiv$ MIN{$\text{Explainability}_{\text{LOSS}}$} ∩ MIN{$\text{Accuracy}_{\text{LOSS}}$}

where ≡ is the logically equal/equivalent operator
and ∩ is the logical AND operator

Expressed as general logical equivalents:

Form 1: MAX {D_1} ∩ MIN {D_2 Loss}

≡ Form 2: MAX {D_2} ∩ MIN {D_1 Loss}

≡ Form 3: MAX {D_1} ∩ MAX {D_2}

≡ Form 4: MIN {D_1 Loss} ∩ MIN {D_2 Loss}

≡ Form 5: MIN {D_2 Loss} ∩ MIN {D_1 Loss}

This can be reduced to three (3) general forms for each 1-to-1 pairwise trade-off:

Form 1: MAX{D_O} ∩ MIN{D_T}

≡ Form 2: MAX{D_O} ∩ MAX{$D^{-1}{}_T$}

≡ Form 3: MIN{$D^{-1}{}_O$} ∩ MIN{$D^{-1}{}_T$}

Using Form 1, a potential objective function becomes:

$$Maximize\ D_1 = \sum_{i=1}^{M} \sum_{j=1}^{N} C_{ij} X_{ij}\ s.t. Minimize\ \{D_2 Loss\}$$

$$\sum_{\substack{0 \leq i \leq m \\ 0 < j < n}} P(i,j)$$

$$\sum_{\substack{0 \leq i \leq m \\ 0 < j < n}} PF(i,j)$$

where P is a production function $P(i,j) = C_{ij} X_{ij}$

The reader will note the various functions above are all forms of maximizations with constraints problems and representative equations.

Given the hypothesis and empirical data thus far supporting the notion that the seven identified trade-off dimensions are indeed scalars, we can deduce that there is some point on the frontier plotting the dimensions as Cartesian coordinates (D_1, D_2), which is a change in slope or direction between the x and y axis. In Abowd and Schmutte (2018), this point is $O(\varepsilon, I)$ where there is a change in sign between $dI / d\varepsilon$ and $d^2I / d\varepsilon^2$. While the shape and slope may vary as we move from left to right on the x-axis, the notion of a 1-to-1 trade-off indicates that at some stage a marginal increase in one trade-off dimension requires a marginal reduction in the other trade-off dimension.

Now including a set of additional organization actions, e.g. $M(\varepsilon, I)$, $N(\varepsilon, I)$, as part of the objective function, the analysis is made possible when the set of available actions is considered a closed set as in Abowd and Schmutte (2018). However, the set of actions is decoupled from the trade-off dimensions themselves, such that the organization can select or alter a particular $M(\varepsilon, I)$ or $N(\varepsilon, I)$ as independent variables, with some measurable impact on their I and ε, referring to their particular trade-off. The explicit inclusion of additional variables is important to make progress on the research question, which are predicated on the empirical finding that organizations currently deliberately or otherwise, ignore important factors in their analysis when deploying AI technologies to their strategic problem solving and decision making. Explicitly identifying and allowing for the incorporation of additional variables moves these from the realm of missed or ignored, to considered or potentially necessary. Secondly, the ability for an organization to specify and dial-in particular variables is critical if they are to improve the outcomes of these projects.

As I make progress in the creation and defining of these models through the examination of existing literature, it's possible that patterns will emerge such that the same or similar variables recur across industries, types or classes of problems. If this is indeed observed, it would allow for organizations to also better account for unknown or inadvertent omissions seen in other industries or problem types, which they themselves may not be familiar with. A test of the quality of my models will be the demonstration of an improvement in the overall objective function, through the achievement of a higher magnitude $\text{Max}\{D_X\}$ or lower magnitude $\text{Min}\{D_Y\}$, or any other combination of the five forms of an objective function. Notwithstanding the overall goals of this study, even an articulation of specific locations of $O(\varepsilon, I)$ for any of the models, and possible regions on the production frontier where important point(s) may be reached, would I believe be valuable contributions.

The above approaches and models assist in laying a foundation for the desired outcomes of this study and subsequent analysis. It would be very valuable if an organization could compute and model in advance, the extent to which a dD_O differential gain in their objective function will be associated with a dD_T differential loss in some other trade-off dimension, and then effectively dial in sets of action variables to maximize their objective function.

Synthesizing Concepts

I realize that while many of the concepts in this analysis come from different fields of study, including but not limited to evolutionary biology, organization and management science, economics, mathematics and computer science, the literature and my forgoing analysis supports their interconnectedness and proposed interoperability (Weinberger, 1990; Kauffman, 1993; Weinberger, 1996; Levinthal, 1997; Rivkin, 2000, others). But to that end, some further exploration, integration and synthesis is likely helpful.

Fitness, Fitness Landscapes, and Fitness Functions

In the NK Model, fitness represents the system's performance or adaptation to its environment—in other words, fitness is determined *against* a fitness landscape (Kauffman & Levin, 1987). In the context of a production function, fitness can be the overall output or profit generated from a set of inputs. We can visualize such a fitness landscape by understanding how different combinations of component states, e.g., labor, capital, technologies, network, algorithms, affect a production function's performance. In the context of evolutionary biology, fitness represents reproductive success or replication rate, and a fitness landscape can be utilized to visualize the distributed relationship between different DNA variations, or encoded genotypes, and their respective associated reproductive success[37]. In effect, this fitness also represents the height of the landscape, where similar genotypes and fitness levels are presumed to be in close proximity, while different genotypes and fitness levels should be non-proximal. The fitness landscape is now composed of the set of all genotypes with their respective similarities, differences, and fitness levels or replication rates. This aligns with Weinberger's description of the Kauffman and Levin rugged landscape as a fitness landscape (1990).

[37] https://en.wikipedia.org/wiki/Fitness_landscape, retrieved May 25, 2024.

If one is a strategist, the landscape is the set of strategic choices which collectively make up the problem being addressed, i.e. the objective function. And if one is a computer scientist, the landscape is composed of the set of allowable configurations, e.g. software, hardware, etc. I've proposed that such configurations can *inter alia* include different machine learning models, and other choices, and where and how these interventions are introduced into the landscape. As explored in detail above, adaptive walks or *hill climbing* by genetic mutation or selection, or some other mechanism, is a type of optimization problem. The space of all the results of these adaptive walks is a fitness space, against which a fitness function can be modelled. Going further, the fitness space is also akin to a state space, which as I will describe below, is the space of all NK nodes. In this context, the state space is the bitwise representations of these NK nodes.

Examples of a fitness space include a peptide space or the traveling salesman problem (Kauffman & Levin, 1987). In the former, a set of 20^{50} peptides represent a high-dimensional space against which a fitness function is induced. The ability of each peptide sequence or node to perform its purpose is an optimization process, which gives us a landscape of peaks of different heights, collectively making up the fitness landscape. While each sequence in the former is a biological or evolutionary process, in the traveling salesman problem, the set of solutions to fulfil the basic TSP problem and visit each city once, represent the nodes or peaks on the fitness landscape. The set of shortest total route distances, and therefore time taken, are the local maxima. In both examples, the optimization process is to find the best performing multivariate node sequence. Stated more generally, a fitness function is the property of a solution or fitness space, that induces the optimization process of finding the best performing actions or combinations of actions, resulting in the highest point vertices.

N and its Importance

In the NK Model, N represents the number of peptides in a sequence (Kauffman & Levin, 1987; Kauffman, 1993). In my analysis, I propose that N represents the number of variables the organization can control or influence, and this is supported by Rivkin (2000) wherein he models N as the number of decisions a firm faces, each simplified to either 0 or 1 within a discrete Boolean option set. These N decisions become variables in support of a particular strategy S, which is a configuration of decisions S_i as a vector $[S_1, S_2,...S_N]$. Each S_i is either 0 or 1, giving us 2^N possible configurations of decisions. Kauffman's adaptive walk is conceived as an optimization process that moves from a current location on the fitness landscape, to higher peaks via 1-mutant variants. Similar to the protein spaces that Kauffman describes, we can alter single-variable changes in pairwise trade-offs (1993). A gain is represented by 1, while a loss is represented by 0. If N=2 and there are two (2) possible choices for each location of N, then we get $2^2 = 4$ possible combinations of length 2: 00, 10, 01, 11. The number of possible combinations or node strings is 2^N where 2 represents the number of possible choices for each location N. However, earlier discussion on this has already provided support for the notion that 00 and 11 are not feasible in a true trade-off construction of these concepts. Biggiero (2016, p. 66) supports this notion for examples of a 2^4 combination example: "0000 and 1111, respectively – can be eliminated because they have no practical meanings in the real world." Both the objective dimension and trade-off dimension cannot *move together* as mutual gains or losses. Doing so would violate the underlying premise of pairwise trade-offs. In Table 10 below, I provide a detailed table of bitwise state space representations with increasing values of N.

K from NK Model

K represents the number of interactions for each change in a node in a peptide sequence, and is a measure of the richness of epistatic interactions among sites (Kauffman & Weinberger, 1989). The specification of K provides the degree to which the fitness of the organization depends on interaction effects among attributes or variables, with K influencing the level of contribution of a given attribute to the organization's overall fitness (Levinthal, 1997). It is expressed as an integer ranging from minimum 0, meaning no interactions, to maximum N-1 where every change interacts with every other peptide (Kauffman & Levin, 1987). In other words, K=0 implies that the contribution of any elements, such as structure or personnel, is independent of all other attributes, while K=N-1 means that the fitness contribution of any single attribute depends on all other attributes. Rivkin (2000) posits a model where each decision N makes a contribution to C_i firm value, where Ci depends on S_i (see above) and K. Putting it together, where K=0, we get Ci = Ci(S_i), and where K=N-1, we get $C_i = C_i(S_i; S_{i1}...S_{iK})$. K is also a measure of landscape ruggedness (Kauffman & Levin, 1987; Kauffman, 1988; Kauffman & Weinberger, 1989; Weinberger, 1990; Kauffman, 1993; Weinberger, 1996; Levinthal, 1997; Rivkin, 2000). In my analysis, I propose that K represents the interaction of each variable the organization can control or influence, with other variables. This can also be considered and modeled as the intensity of interaction between the variables via the parameter K, but does not restrict the functional form of the interaction (Levinthal, 1997).

NK Nodes

In the NK Model, NK nodes represent a genome, or allele of the sequence of N peptides (Kauffman & Levin, 1987; Kauffman & Weinberger, 1989; Kauffman, 1993). In my analysis I propose that a NK node would represent a set of control/decision/independent variables taken together, that collectively form a PF as a function or curve of node combinations (var1, var2, varM, varN, ..., varZ). The precise definition of these variables will depend on the nature of the problem and the variables the organization seeks to control or choose. Connecting these concepts, the N choices each form a single bit within the overall sequence and taken together, form the node sequence or string. So, if N = 4 and each N is a single choice of 2 possible options, then we return to an example with $2^4 = 16$ possible nodes.

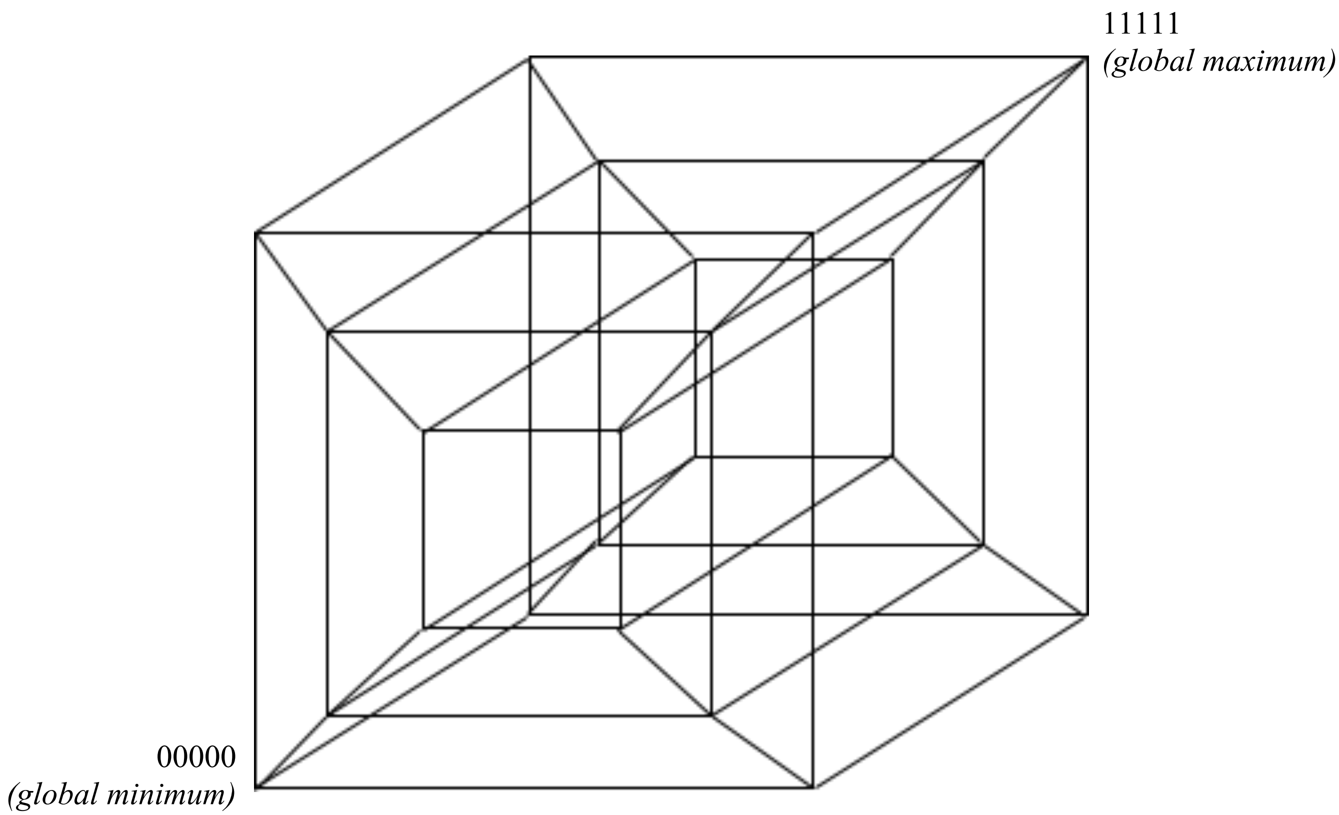
Now that we have a clearer picture of N, K and NK nodes, one can visualize the relationship between these, leading to the Boolean hypercube representation Kauffman and Levin (1987) provide where the cubes of N-dimensions can be constructed from layers of nested 3-dimensional cubes. In other words, each cube within a hypercube construct is a simple 3-dimensional cube with 8 vertices (think of a dice). Each vertex on each cube represents a NK node with a bitwise designation, e.g. 0000, 0001, 0010, ...1111. By placing cubes within cubes and connecting the vertices, we get the hypercube construct whereby a move can take place along the same cube or to another cube, thereby proxying a greater than 3-dimensional construct. For greater N, the number of cubes can continue to be layered and the number of vertices grows at 2^N if there are, as proposed, 2 possible values for each N. Figure 4 visualizes a five-dimension hypercube where N=5 and is Boolean.

Objective Function

Taken from the study of economics, an objective function may be a mathematical expression whereby an actor—economic or otherwise—seeks to maximize or minimize some

Figure 4

A Five-Dimensional Hypercube with Binary Sequences as Vertices, expanded from Kauffman & Levin (1987)



Number of possible values (vertices) for each binary bit (0,1) where N=5 → $2^5 = 32$

Non-repeating binary sequence set: [00000, 00001, 00010, 00011, 00100, 00101, 00110, 00111, 01000, 01001, 01010, 01011, 01100, 01101, 01110, 01111, 10000, 10001, 10010, 10011, 10100, 10101, 10110, 10111, 11000, 11001, 11010, 11011, 11100, 11101, 11110, 11111]

quantity[38] (Nomiks, 2023). The quantity being maximized or minimized may be profits, costs, or in the context of this research study, any of the seven objective/trade-off dimensions. In the

[38] https://www.stockmaster.com/objective-function, retrieved Dec 16, 2023.

trade-offs I've proposed and their pairwise constructs, organizations seek to maximize a D_O and/or minimize a D_T.

In order to solve for an objective function optimization such as maximization or minimization, the organization must determine independent decision variables, which the organization can control to achieve an objective. These are privacy loss (ε) and accuracy (I) in Abowd & Schmutte (2018), and expanded to included M(i,j) and N(i,j) in my proposed model. Now, an optimal solution is one where the combination of independent decision variables *collectively* optimizes the objective function.

Production Function

A production function (PF) in economics, is a function that represents the *total* output of an entity given a combination and quantity of inputs[39,40]. Typical empirical examples contain a progression of input quantity positively correlated with output quantity, with the latter exhibiting at some point, diminishing returns in the short run. This means that the marginal output starts to decrease for each unit of input, with a limit of 0, where additional input results in no additional output (marginal output = 0). This is the O(ε, I) in Abowd and Schmutte (2018) and in the context of my study, where the shape of the curve starts to exhibit an inflection (Bronshtein et al., 2007) and is a point of interest, i.e. where the $dD_T > dD_O$.

---

[39] https://en.wikipedia.org/wiki/Production_function, retrieved Dec 16, 2023.
[40] https://en.wikipedia.org/wiki/Production–possibility_frontier, retrieved Dec 16, 2023.

As elaborated on extensively above, Abowd and Schmutte (2018) posit that acts such as data publication by statistical agencies entail some level of accuracy (I) requiring some level of privacy loss (ε). Each node or pair (ε, I) is a *production activity*:

> Production activities usually represent vectors of inputs and outputs such that the inputs can be transformed into the outputs. Our insight is to think of the information in the database as akin to an endowment of potential privacy loss. Some of the privacy loss endowment must be expended by the data custodian to produce population statistics of any accuracy. (Abowd and Schmutte, 2018, p. 16)

Taken together, the set of production activities gives us a production function or production frontier. Where there is a gap between the MAX{PF}of an organization, the delta between the maximum possible outputs given constraints, and the Actual{PF}, an objective function optimizes inputs to increase performance or location relative to MAX{PF}—the goal being to maximize D_O, e.g. profits, or Minimize D_T, e.g. losses.

Optimization via Computation

Mathematical optimization against *optimization-class* problems can be divided into two categories: discrete and continuous[41]. The former involves discrete objects such as integers, permutations or graphs and come from a countable set, the latter continuous category involves variables from a continuous non-bounded set and can include multimodal problems. In the context of ML, this translates into continuous evaluation of data quality by algorithmically minimizing the cost by optimizing parameters with lowest error—or maximum accuracy. Modelled as an optimization of an objective function:

[41] https://en.wikipedia.org/wiki/Mathematical_optimization, retrieved Nov 16, 2024.

$$A \subset \mathbb{R}\ and\ A(f)\ represents\ a\ search\ space$$

$$x_o \in A\ and\ all\ x \in A$$

$$f(x_o) \leq f(x)\ minimization$$

$$f(x_o) \geq f(x)\ maximization$$

The elements of f(x) in A represent a search space and $f(x_o)$ represents the objective function, which in this case may be a minimization or maximization., and optimization is the solution that achieves MAX{$f(x_o)$} or MIN{$f(x_o)$}.

The above is a simple example of a single-objective optimization. The addition of more than one objective to an optimization leads to not only added complexity, but also trade-offs when the objectives conflict or operate in opposing directions. We've of course seen this throughout this analysis. To mitigate against a *pure* improvement or maximization of one objective at the expense of another, the Pareto sets that collectively form the Pareto frontier, I have proposed an approach that goes towards a global optimization for convex and non-convex objective functions. Iterative optimization techniques and algorithmic approaches fail to achieve such multi-model or global optimizations[42]. A multi-model or global optimizer can be achieved via several approaches, including evolutionary algorithms, Bayesian optimization and simulated annealing.

Binary Representation and Bitwise Operations

Kauffman and Levin (1987) have provided a model for representing NK nodes as N-digit binary sequences. While they model random walks on a fitness landscape via changes in the

[42] https://en.wikipedia.org/wiki/Mathematical_optimization, retrieved Nov 16, 2024.

value of a single bit within the binary sequence, there are also logical and mathematical operations that allow for computations and other calculations between and against such sequences. This will become an important consideration as I proceed into optimizations via algorithmic enhancements of production functions. Delving deeper into this, there are different bitwise or Boolean operators that allow for computation of 2 or more bits. Here are some simple examples of how bitwise computations with AND and OR logical operations function[43]:

1 AND 1 = 1 ≡ TRUE

1 AND 0 = 0 ≡ FALSE

0 AND 0 = 1 ≡ TRUE

1 OR 1 = 1 ≡ TRUE

1 or 0 = 1 ≡ TRUE

0 or 0 = 0 ≡ FALSE

In addition to AND and OR, the set of common bitwise or Boolean operators includes XOR, NOT, NAND, NOR, XNOR, and so on, which I'll leave for reference elsewhere (Sparks, 2018). And, this starts to get much more interesting when we increase N beyond the 1-bit examples above, and certainly with higher N. In a computational context, these logical operators also represent logic gates, which are the building blocks of all computational systems—utilized to encode simple mathematical operators into computer hardware and microchips. A logic gate is a small electronic circuit made of one or two transistors and used to process binary data. These form the foundational microscopic calculators that, composed and assembled as literally billions of gates and transistors, allow microchips, CPUs and GPUs to perform advanced computational

[43] Adapted and expanded from https://en.wikipedia.org/wiki/Bitwise_operation, retrieved Nov 26, 2023.

operations. As a logic gate, 0 represents false or off or closed, while 1 represents true, or on, or open. A bit is either on/off, true/false, yes/no, or 1/0.

Fitness Space or State Space

Returning back to a binary representation of the possible combinations of the nodes in the fitness space, and while not violating the premise of trade-offs moving in opposing directions, for this analysis the value of N can certainly be any integer from 1 to the maximum a state machine can compute. If we increase the value of N, or number of variables the organization can control or influence, we start to get an exponential increase in the number of possible combinations or single-variable states. As a binary to 2-possible-bit representation of zeros (0s) or ones (1s), the number of bits is the exponent on 2, giving us 2^N number of possible combinations of each location within the fitness or state space. In Table 10 below, I provide numerical examples of how these related dimensions interact to provide the numerical state space against which any computational approach can be deployed. As readers will recognize, the absolute and numerical ranges for the fitness landscape grow exponentially into very large values as the value of N increases.

Table 10

How the Value of N Translates into Bitwise States and Spaces

No. of Bits[44]	**Numerical Value by Bit**	**Binary Range**	**Possible Combinations (2^N)**	**Absolute Range (0 to 2^N-1)**	**Numerical Range**[45] **($–2^{N-1}$ to 2^{N-1}-1)**
1	1	0 to 1	$2^1 = 2$	0 to 1	$–(2^0)$ to 2^0-1 = -1 to 0
2	2 1	00 to 11 [00, 01, 10, 11]	$2^2 = 4$	0 to 3	$–(2^1)$ to 2^1-1 = -2 to 1
3	4 2 1	000 to 111 [000, 001, 010, 011, 100, 101, 110, 111]	$2^3 = 8$	0 to 7	$–(2^2)$ to 2^2-1 = -4 to 3
4	8 4 2 1	0000 to 1111 [0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111]	$2^4 = 16$	0 to 15	$–(2^3)$ to 2^3-1 = -8 to 7
5	16 8 4 2 1	00000 to 11111	$2^5 = 32$	0 to 31	$–(2^4)$ to 2^4-1 = -16 to 15
6	32 16 8 4 2 1	000000 to 111111	$2^6 = 64$	0 to 63	$–(2^5)$ to 2^5-1 = -32 to 31
7	64 32 16 8 4 2 1	0000000 to 1111111	$2^7 = 128$	0 to 127	$–(2^6)$ to 2^6-1 = -64 to 63

[44] https://study.com/academy/lesson/16-32-64-128-bit-integers.html, retrieved Nov 26, 2023.

[45] Negative numerical values are represented in binary via a sign bit, either within the existing bit register space (same number of bits), or an extra bit (https://en.wikipedia.org/wiki/Signed_number_representations/, retrieved Jan 20, 2024. For the purposes of this analysis, I will only consider positive numerical values.

No. of Bits[44]	Numerical Value by Bit	Binary Range	Possible Combinations (2^N)	Absolute Range (0 to 2^N-1)	Numerical Range[45] (-2^{N-1} to 2^{N-1}-1)
8	128 64 32 16 8 4 2 1	00000000 to 11111111	$2^8 = 256$	0 to 255	$-(2^7)$ to 2^7-1 = -128 to 127
...					
16	32768 16384 8192 4096 2048 1024 512 256 128 64 32 16 8 4 2 1	0000000000000000 to 1111111111111111	$2^{16} = 65{,}536$	0 to 65,535	$-(2^{15})$ to $2^{15}-1$ = -32,768 to 32,767
32	2147483648 1073741824 536870912 268435456 134217728 67108864 33554432 33554432 8388608 4194304 2097152 1048576 524288 262144 131072 65536 32768 16384 8192 4096 2048 1024 512 256 128 64 32 16 8 4 2 1	00000000000000000000000000000000 to 11111111111111111111111111111111	2^{32} = 4,294,967,296	0 to 4,294,967,295	$-(2^{31})$ to $2^{31}-1$ = -2,147,483,648 to 2,147,483,647
64	9223372036854775808 ... 128 64 32 16 8 4 2 1	0000000000000000000000000000000000000000000000000000 to 11111111111	2^{64} = 18,446,744,073,709,551,616	0 to 18,446,744,073,709,551,615	$-(2^{63})$ to $2^{63}-1$

No. of Bits[44]	**Numerical Value by Bit**	**Binary Range**	**Possible Combinations (2^N)**	**Absolute Range (0 to 2^N-1)**	**Numerical Range**[45] **(-2^{N-1} to 2^{N-1}-1)**
		11111111111 11111111111 11111111111 11111111111 111111111			
128	1701411834604 6923173168730 3715884105728 ... 128 64 32 16 8 4 2 1	00000000000 00000000000 00000000000 00000000000 00000000000 00000000000 00000000000 00000000000 00000000000 00000000000 00000000000 0000000 to 11111111111 11111111111 11111111111 11111111111 11111111111 11111111111 11111111111 11111111111 11111111111 11111111111 11111111111 1111111	2^{128} = 340,282,366,9 20,938,463,46 3,374,607,431, 768,211,456	0 to 340,282,366 ,920,938,46 3,463,374,6 07,431,768, 211,455	$-(2^{127})$ to $2^{127}-1$

Some readers may be interested to learn that there are currently no commercially available 128-bit CPUs, thus, 64-bit is the largest memory address currently operable by a CPU. As a fun fact, a CPU with 128-bit addressing could address 2^{128} or 3.4×10^{38} bytes, which "would exceed the total data captured, created, or replicated on Earth as of 2018", estimated as of that date to be approx. 2^{74} bytes[46]. It's also interesting to note that Reinsel et al. (2018) expected this to grow to 2^{175} by 2025. It is unclear whether the world is on track to surpass this target between now and 2025. In any event, the move to 128-bit CPUs has not yet occurred.

How to Solve this Optimization Problem?

A Boolean hypercube is a representation of a *natural* adaptive walk from some/any point on the fitness landscape to higher local peaks via single-location changes or mutations, with a *cessation* of walk activities once a local maximum, e.g. O(ε, I), is reached where all neighboring vertices away from this local maximum are of lower height or less optimal (Kauffman & Levin, 1987). As an organism moves from one location to another via single-N-location variation, one can envision single-node-location moves along vertices of a Boolean hypercube. Once a local maximum is reached, the movement stops since any further move is less optimal in terms of the *value of* the node sequence or string. Similarly, there is no gain in moving beyond O(ε, I) because there the differential loss in privacy is greater than the gain in accuracy, *or* there is less accuracy gain than privacy lost (Abowd & Schmutte, 2018). It can be modeled and constructed as a line on a 2-axis Cartesian plane. And as per Abowd and Schmutte (2018) for a PF containing a trade-off, this is a positive concave function. The *zero* starting point for each can be represented in binary as the point 0, 00, 0000, 00000000, or larger depending on the value of N. However, an optimization can begin at any point on the curve or Hypercube. An improved PF via *higher* node locations renders each O*(ε, I) to be higher than or superior fitness compared to each O(ε, I). In effect, the Boolean hypercube in Kauffman and Levin (1987) can be represented

[46] https://en.wikipedia.org/wiki/128-bit_computing, retrieved Mar 21, 2024.

as, and leads to the same result as, the increasing (positive) concave down curve in Abowd and Schmutte (2018). Both represent a series of choices against certain variables leading to places on a production function. The NK Model and explanation are as valid with policy decisions trying to balance the trade-off of accuracy vs privacy loss, as it is with biological organisms undertaking walks towards higher peaks on a fitness landscape.

Now that our production function can be represented as a line of *at most* 3rd order, with some characteristic shape. this function or line is in effect an alternate representation of a sequentially ordered set of Hypercube vertices, where there is an *upward walk* from some starting point, some minimum such as a sequence of zeros, e.g. 0000, to some maximum such as a sequence of ones, e.g. 1111. Both represent a random walk on a landscape with certain ruggedness characteristics, or an output of a production function with certain input variables. In other words, organisms and organizations will produce outputs based on inputs, or traverse adjacent vertices (Kauffman & Levin, 1987) until they reach a local maximum. Each local maximum can be a node on the production function. Boolean hypercubes of varying dimension N have multiple local optima, and of these, there is one that is the highest or global maximum. This global maximum is the $O(\varepsilon, I)$ in Abowd and Schmutte (2018) where the PF intersects the SWF. Collectively, the set of production nodes make up the *natural* production function, with the possibility of one or more local maxima. I will propose an approach to altering/improving this via interventions in process and algorithms.

The question that now arises is, how the Boolean representation of a 1 in an objective dimension mathematically compares to a 0 in the trade-off dimension. This creates the solution space whereby an organization can undertake analysis to reduce the gap via mechanisms such as such as M(ε, *I*) and N(ε, *I*), and others. This does not contradict Kauffman (1993) wherein he states that *only* N and K matter, because I'm pre-supposing a model whereby extra-dimensional interventions can alter the distribution and therefore, attained local maximum. Expanding our view and analysis to include additional variables which individually and collectively contribute to the D_O and D_T, we get an altered PF.

$$PF = \{(\varepsilon, I) \mid \varepsilon > 0, I < 0 \text{ s.t. } G(\varepsilon, I) = 0, M(\varepsilon, I) = 0, N(\varepsilon, I) = 0\}$$

In the NK Model, "as the main parameters are altered, the model generates a *family* of increasingly rugged multipeaked landscapes" (Kauffman, 1993, p. 40). These peaks are local optima, and points of interest in my analysis of an organization's production function. In context of current proposed pairwise trade-off matrix, if N=2 and K=1, then we get $2^{K+1} = 2^2 = 4$ possible combinations of each interaction of trade-off dimensions [00, 01, 10, 11]. Now capturing my suggestion that 00 and 11 are not possible, as supported by Biggiero (2016), this low dimension fitness landscape further reduces to 01 and 10.

NK Model as an Optimization Utility and Unanswered Complexity Considerations

Can the adaptive walk conceived by Kauffman and Levin (1987), an optimization process that moves from a current location (or fitness) to higher peaks via 1-mutant variants, be used as a utility applied to other forms of optimization including organization strategic decision making? How can this be modeled and tested, and if so, how? In their own words:

> Lacking a good theory, we have resorted to numerical simulations to discover the features of landscapes for two apparently dissimilar problems: (1) adaptation of desired cell types via 1 to 5 connection or 1 to 5 Boolean function mutants in K = 2 input genetic networks; (2) heuristic approaches to the traveling salesman problem by 1-mutant to 4-mutant

> moves, where a one-mutant move involves exchanging the positions of two cities in the circuit through the N cities. Both complex combinatorial optimization problems nevertheless exhibit similar features. (Kauffman & Levin, 1987, p. 37)

Does Porter's alignment and fit, up-down-across, exert pressure towards an increased K? Then how are higher peaks reached? Or is higher K the result in misaligned organizations which exhibit low fit? I posit that higher-fit organizations can reach higher peaks without increasing K and without suffering from Kauffman's *complexity catastrophe* (1993). Complexity catastrophe occurs when the expected fitness of any local optima f* drives to a mean of 0.5. This means that as N increases, the probability of finding a higher peak is just as likely as finding a lower peak. This same f* is the point of interest on the production function, as it is a local maximum of the walk on the landscape, but in complex systems with high N, the distribution of peaks reached is lower than those observed with lower N. How do we increase this location while not violating Kauffman's findings with increasing N? He proposes a limited solution whereby the cost per part is ignored, thereby permitting both the complexity maximum possible fitness, to increase without bound. But even this approach is bounded because at some N, the total cost of the system is greater than the fitness, and increasing N beyond this point is no longer profitable—there must always be a marginal gain in fitness for each input. The reader may be interested to explore Kauffman's further study of, amongst other things, the addition of variables and recombination as an adaptive strategy on rugged landscapes (1993). His key finding of value to this study, is that recombination is most useful when K=0, and diminishes to a point of no value on completely random landscapes where K=N-1. When K=0, akin to genes in the genome making independent contributions to fitness, in other complex systems, functional parts within adapting systems make independent contributions to overall system fitness. So, in high-fit organizations, K is low, and the implicitly aligned parts of the organization operate so as to make independent contributions to the overall organization fitness.

Proposed Approach to Increase a Production Function

I will now develop an approach to optimize a production function using a combination of the underpinnings developed above, namely: the NK Model from Kauffman & Levin (1987) and its applications into management sciences (Weinberger, 1996; Levinthal, 1997; Rivkin, 2000; others), $O(\varepsilon, I)$ and the SWF from Abowd and Schmutte (2018), and concepts from economics and management science. As we've seen in Kauffman and Levin (1987) and other literature applying the NK Model to management and organization science (Kauffman & Weinberger, 1989; Weinberger, 1990; Kauffman, 1993; Weinberger, 1996; Levinthal, 1997; Rivkin, 2000; others), the NK Model is a theoretical framework used to study the dynamics of complex adaptive systems. We know that it's been used within a wide variety of contexts from biological evolution to economics and even management theory (Weinberger, 1996; Levinthal, 1997; Rivkin, 2000; others), but not as a tool to help increase a production function. However, insights gained from the NK Model, other complex systems theory and problem complexity theories, help inform how production functions in complex systems behave. A comprehensive approach may be developed to optimize a production function. I suggest that the NK Model can also serve to provide a framework for optimizing the performance of complex systems. Principles of the NK Model, such as component interaction, adaptation, and robustness can be applied to enhance and optimize the performance of a production function within a dynamic and complex environment.

A production function can be conceptualized as a complex adaptive system with multiple, dependent or independent, variables or components. Abowd and Schmutte (2018) rely on two: privacy loss and accuracy. In common economic contexts, independent input variables can include capital, labour, technology, raw materials, and other resources that influence a production function. In my thus-far conceptualized model, I've proposed two new groups of variables, M(ε, I) algorithmic options and N(ε, I) computational resource options, each composed of other variables. These multi-variate nodes in my proposed model, as also seen in the NK Model, now represent different variables or factors that *collectively* affect the PF. By applying these to specific types of production functions and conducting experiments and simulations, I will endeavor to demonstrate improvements in a production function. Drawing on the above, I propose the following multi-step approach to increase and optimize a production function.

Identify Critical Components of the Production Function

As a first step, identify the critical components or variables that contribute to the output of the production function. Each of these components and their combinations thereof, can be represented as and are akin to, the nodes in the NK Model. These components represent the fundamental factors that contribute to the production process of whatever output is under analysis. In other words, these components form the foundation of the system and contribute directly to its output. In contexts such as manufacturing for example, these may include raw materials, machinery, labour, technologies and so on. In other contexts, the production function and its critical components may be very different, e.g. privacy loss as the input in Abowd and Schmutte (2018), used by their broader set of activities G(ε, I), which represents their production function. To effectively identify these critical components, it is essential to conduct a comprehensive analysis of the production system. This analysis should involve not only examining the physical elements that may be involved in production, but also considering the broader operational and logistical aspects. Using a manufacturing process again, in addition to identifying the specific machinery used, it's important to understand the role of factors such as supply chain management, inventory control, quality assurance, and others that may be relevant.

Once these critical components have been identified, a sensitivity or regression analysis allows for a prioritization based on their relative importance and impact on the overall PF. Some components may have a more significant influence on output quantity or quality, while others may be more critical for ensuring operational efficiency or cost-effectiveness. By prioritizing these components, efforts can be focused on optimizing the most impactful variables of the production process. This analysis feeds into and helps with the next stage, where the interactions K are assessed. The number of components will then drive the number of dimensions that need to be modelled and may be visualized via the Boolean hypercubes discussed above. A multi-variate analysis is inherently difficult to model via this proposed framework due to the rapid increase in relationship complexity if one goes past pairwise trade-offs. While I will not cover these scenarios in this analysis, an approach to handling >2 dimensions is to model these via multi-dimensional state spaces—greater than two axes—where the order of the production function would likely also be greater than cubic.

These components, while they may seem disparate, contribute to and should be factored into the analysis. In some quantifiable way, all components whether they are in the initial consideration set or reduced critical set, surface within one or more of accuracy, explainability, fairness, privacy, reliability, safety, or speed. In effect, the identification of critical components of the production function forces organizations to examine multiple D_O:D_T pairwise trade-offs, thereby expanding the variable set to include previous implicitly or explicitly ignored variables. Simply stated, I'm suggesting that no critical component identification exercise can be limited to, e.g. speed vs accuracy.

Assess Interactions

We've seen via the NK Model that components within systems interact with each other—within and between biological organisms (Kauffman & Levin, 1987; Kauffman, 1988; Kauffman, 1993), inter-/intra-organizationally (Kauffman, 1990; Levinthal, 1997; Levinthal & Warglien, 1999; Siggelkow & Levinthal, 2003; Ghemawat & Levinthal, 2008), or more broadly (McKelvey, 1999; Rivkin, 2000; Rivkin, 2003; Moldoveanu & Bauer, 2004; Rivkin & Siggelkow, 2007). The NK Model provides one vehicle to analyze and understand these interactions, and there is strong support for the notion that components in complex systems rarely operate in isolation—instead, they interact with and influence one another in various ways. In an economic PF analysis, this can be the interplay between labor, capital, technology, and other factors. Understanding these interactions between components is crucial for modeling the dynamic behavior of a PF.

To that end, the next step I propose is to assign an interaction value between nodes—the K from the NK Model—for the components identified in the first step. To reiterate, this value represents the degree of landscape ruggedness, impact by or interdependence between, the components. Kauffman and Weinberger (1989) also give us the concept of *the transition* from a single-peaked "Fujiyama" landscape to a multi-peaked "badlands" landscape as K increases, where walks to local optima on such latter landscapes become shorter as K increases. And as we've seen, a higher K means higher impact of each change on another [neighbouring] component. K can increase up to a maximum of N-1, wherein every component impacts every other component. A lower K, down to a minimum of 0, means there are no interaction effects between components.

> The main conclusions to bring away are that increasing K relative to N increases the number of local optima, shortens the lengths of walks to optima, increases the rate at which fitter neighbors dwindle to 0 along adaptive walks, increases the ruggedness of the landscape, reduces the fraction of optima accessible from a given point, reduces the number of points which can climb to a given optimum and leads to a complexity catastrophe in which accessible optima fall toward the mean of the space. All these features presumably reflect the fact that, as K increases, more conflicting constraints, or what in spin glass models is called frustration, sets in. (Anderson, 1985, as cited in Kauffman & Weinberger, 1989, p. 222)

It is also possible that different K-values exist within the overall system, indicating that some components have higher magnitude interactions (higher K), while others are more independent (lower K). In any PF, different factors and variables interact to influence the outcome, and gaining insights into the dynamics of the system helps identify opportunities for optimization. Understanding K interactions further helps identify which components of the PF have the most significant impact and warrant closer attention. In a pairwise 1-to-1 model, only the D_O and D_T are considered to the exclusion of all other dimensions, thereby limiting K to define the degree of interaction between bivariate nodes contributing to the PF. Any mathematical or computational optimization models are now focused on factors that contribute to higher nodes on the production frontier.

To assess these interactions, organizations can employ a variety of techniques, including data analysis, modeling, and simulations. Data analysis techniques such as correlation or regression analysis can help identify statistically significant relationships between different components of the production system. Modeling techniques such as system dynamics modeling can provide a more holistic understanding of how changes in one component can propagate throughout the system and affect other components. Simulation tools can also be leveraged to experiment with different scenarios and observe how changes in one component impact the overall performance of a PF.

Interested researchers may want to examine inter-/intra-nodal interactions using network analysis models. The NK Model is essentially a network of interconnected nodes. Applying network analysis techniques to such production functions may help to understand the relationships between different components and how they contribute to overall system performance. This can help identify other areas of improvement, and seems to open up additional avenues of research, including complex or yet to be defined/understood interactions between dimensions or nodes, how the inverse relationship between trade-off dimensions ebbs and flows, or even how node composition may change in complex systems.

Explore State Spaces

Next, I propose an exploration of the state spaces of these components, where state spaces represent the possible configurations or states that each component can assume within the system. I've covered this in detail above where I discuss N and its importance, and as demonstrated, it can have dramatic impact on the size of the overall problem and state space. By systematically exploring these state spaces, organizations can gain insights into the complexity and variability of the PF and identify opportunities for optimization.

While it is possible, and indeed likely, that these state spaces vary by component, the NK Model provides a useful simplification: treat the universe of state spaces as Boolean values of 1 or 0. These can now be subject to bitwise operations at the component level. By any approach, the idea here is to explore how different combinations of the states of each component affect the overall performance and local maxima of the production function.

By exploring these state spaces, organizations can identify optimal configurations that maximize output or other D_O, minimize costs or other D_T, or some other combination of components in their OF. In my analysis, the exploration of state spaces is via quantitative techniques, including statistical analysis, optimization algorithms, and simulation modeling can then be used to validate and refine hypotheses and identify optimal solutions. Different ML algorithms and techniques can also be deployed to explore and identify optimal state spaces or optimal node combinations, and simulations run on real or simulated data combinations.

Optimize Fitness

Having identified the critical components, assessed their interactions, and explored their state spaces, the proposed next step is to optimize the PF, or fitness of the PF. As I've discussed above and, in this context, fitness refers to the overall performance or efficiency of the production system. I've defined an improvement or optimization in fitness as a value where the production frontier has shifted to yield an $O^*(\varepsilon_2, I_2) > O(\varepsilon_1, I_1)$, or a point $O^*(\varepsilon_2, I_2) > O(\varepsilon_1, I_1)$ on the same production frontier, or both. More generally, the overall performance can be measured by specific metrics such as output quantity, quality, cost-effectiveness, or resource utilization. These are all, in effect, forms of optimizations of a production function, and in all cases, the model pre-supposes a D_O and D_T as 1-to-1 pairwise trade-off dimensions.

Define a fitness function that quantifies the performance or output of your production function based on the states of the variables. The fitness function could represent production output, cost efficiency, or any other relevant metric. To optimize fitness, I propose conducting simulations or empirical experiments with different configurations of critical components to maximize fitness, and then evaluate their impact on overall performance. This presents an iterative process of experimentation and analysis allows an organization to fine-tune a PF, with the ultimate goal of enhancing its performance and effectiveness.

One computational approach to achieve this is to use a class of algorithms known as evolutionary algorithms, and specifically within this class, the use of genetic algorithms. Like ML is a type of AI technology, genetic algorithms are types of evolutionary algorithms that algorithmically emulate genetic events that occur in nature and biological contexts, such as birth, selection, reproduction/mating, cross-over, and mutation (Holland, 1992; Whitley, 1993). These algorithms are a class of optimization algorithms inspired by the principles of natural selection—indeed concepts from the NK Model like evolution and adaptation—to iteratively search for optimal solutions within the complex state space of the PF. By using them in conjunction with the NK Model, such algorithms allow for the search and identification of configurations of components states to maximize fitness, thereby leading to an optimized PF. Genetic algorithms and how they operate is described in more detail below, and will be modelled and simulated in subsequent sections of this study.

Test Robustness

Taking the concept of complex system behaviour a bit further, the NK Model may be used to assess robustness, by analyzing how a production function performs under various scenarios and unplanned external impacts—thereby helping to optimize a PF to be more resilient. While this is also a form of optimization, it is different that the $O^*(\varepsilon, I)$ vs $O(\varepsilon, I)$ discussed previously. Here, the goal is to reinforce the PF against external factors or shocks. The assessment and subsequent potential adjustments, are compared to a baseline PF which captures less relevant variables, and presumably reaches lower peaks than a more robust one. The implication is that a more robust system performs better, i.e. reached higher peaks, against perturbations or unexpected changes or disruptions. In effect, a more robust model captures and accounts for more variables while maximizing its fitness, and this is *visible* via higher $O^*(\varepsilon, I)$ on the PF.

Similar to as occurs naturally in nature and explained through the NK Model, an organization can experiment with different combinations of input factors (states) and assess their performance (fitness). A systematic approach to testing different configurations allows for the identification of the most favourable combination of factors that optimize fitness of the PF.

Ongoing Improvement

None of the above stages is one-time and each may change and evolve over time, making this an ongoing process. As the production environment evolves and changes, new factors may emerge as critical, while others may decrease in importance. Regular monitoring and assessment of the production system are necessary to ensure that the identified critical components remain relevant and aligned with the organization's objectives. Along the same vein, interactions within a production system are dynamic and subject to change over time. External factors such as market conditions, technological advancements, and regulatory requirements can influence the nature and intensity of interactions between components. Therefore, it's crucial to regularly monitor and adapt to these changes to ensure the continued effectiveness and efficiency of the production function.

As such, the entire concept of optimizing fitness becomes an ongoing and iterative process, requiring monitoring, evaluation, and refinement. As the production environment evolves and changes, new challenges and opportunities may arise that necessitate adjustments to the production function and its optimization strategies. Such periodic reviews allow for adjustments to the factors that contribute to the production function.

Interventions via AI Techniques

One of the main differences between evolutionary algorithms and systems based on deep learning—such as LMs—is that the former are typically an optimization through maximization, while the latter are usually through minimization. For example, a LLM reduces distance (time and computation) to a next word, n-gram or output parameter. Therefore, an evolutionary fitness function maximization can be considered and deployed as a negated version of a deep learning loss function minimization.

Evolutionary Algorithms

Deployed in the context of this study, evolutionary and genetic algorithms[47] work by iteratively and sometimes in parallel, *generating* and *evaluating* a population of candidate solutions, each represented as a set of component configurations—akin to nodes in the NK Model. Potential solutions to the optimization problem play the role of individual organisms in a population. In general, a mix of potential solutions to the problem is first modelled as a random population. This population is then tested for fitness—how well/quickly it solves the problem—assessed via tests of speed or some other dimension at the conclusion of each offspring generation. Next, the fittest individuals *or solutions* are selected for reproduction. This cycle is then repeated and with each population fitness evaluation, the least fit solutions are eliminated from subsequent iterations. At each iteration, candidate solutions are subjected to evolutionary/genetic *events* such as selection, crossover, and mutation operations, to simulate the process of genetic natural selection. As in biological or natural contexts, this iterative reproductive process drives the population towards better solutions over time. By iteratively refining and improving candidate solutions, genetic algorithms can converge towards optimal configurations that maximize fitness and meet the desired objectives and constraints of a PF. Genetic algorithms can also be used to evolve and optimize interaction values between components, based on the desired objectives and constraints of a PF. And, they can also be employed to explore and optimize the state spaces discussed above, allowing for the identification of optimal configurations that maximize performance or fitness. In this way, deploying a genetic algorithm gives us a stochastic global search optimization algorithm. The

[47] https://en.wikipedia.org/wiki/Evolutionary_algorithm, retrieved Nov 16, 2024.

analysis of non-stochastic or deterministic, non-random, or mixed types of variables will be left to others or subsequent studies.

Self-Learning and Self-Refining

For genetic algorithms and other techniques discussed above, organizations can leverage simulation and modeling techniques to assess the performance of candidate solutions under different scenarios and conditions. This allows organizations to identify robust and adaptive solutions that perform well across a range of potential operating environments and uncertainties.

In the NK Model, randomness is introduced to simulate real-world variability. This can be incorporated via experimentation and random elements in the PF design to explore different scenarios, and then its performance optimized under various conditions. Simulation of the NK Model on different combinations of variable states allows for the assessment and observation of their impact on the fitness of the PF. Naturally then, optimal combinations or configurations lead to optimal production outcomes. The search and identification of such configurations that maximize fitness, can also be automated and turned over to computational techniques or algorithms. As I've alluded to above, such points of interest may lie around the O(ε, I) point from Abowd and Schmutte (2018), where the second derivative changes sign. In effect, I'm proposing deploying ML algorithms to find configurations(s) that optimize the fitness of the PF, and also use MLAs to identify such configurations once they *occur* within the state space. This form of an iterative process, can create models that are self-learning and self-refining.

Chapter Summary

In this chapter, I developed mathematical models to analyze trade-offs, particularly between accuracy and privacy, drawing on Abowd and Schmutte (2018). I explored how improvements in one dimension often come at the cost of another and provides a framework for optimizing AI deployments. By hypothesizing that trade-offs follow at most a 3rd-order polynomial or logarithmic function, I focus on production frontiers and objective functions, incorporating additional organizational actions to enhance AI-driven decision-making.

I apply the NK Model (Kauffman & Levin, 1987) to organization strategy, viewing performance through fitness landscapes where decision variables (N) and interaction complexity (K) shape ruggedness. Organizations navigate these landscapes via adaptive walks, optimizing production functions while balancing competing objectives. Boolean hypercubes illustrate how small changes impact local optima, and evolutionary strategies help organizations maximize efficiency while avoiding excessive complexity.

A structured approach is proposed for optimizing production functions by integrating the NK Model with economic and management science concepts. This involves identifying critical components, assessing interactions (K-values), exploring state spaces, and using evolutionary algorithms to refine strategies iteratively. The framework emphasizes robustness, ensuring adaptability to external changes through simulations, machine learning, and data-driven modeling, ultimately enhancing organizational performance while managing trade-offs.

Chapter 5. Simulations and Results

Introduction

Having developed the above multi-step approach and conversion models and, I will now turn to simulating and testing these computational interventions. These indicative simulations are designed to enable organizations to more optimally select and deploy AI technique(s) against their strategic problem, and as a result, their outcomes. I developed and executed several Monte Carlo simulations, designed to simulate the application of individual and combinations of AI techniques to defined problems, and introduce randomness and variability. These simulations allow for a performance comparison across different problem types, and recommendations on the optimal conditions to apply different AI techniques.

I will in some cases, develop my own code and algorithms as foundational elements, and in other cases, adopt and utilize existing algorithms. These will then be used for simulations and graphical visualizations. That said, this is not a dissertation in computer sciences, and I humbly submit that there are limitations to my coding abilities. So where necessary, I will alternate between pseudo-code and logical descriptions. In other cases, concepts will be introduced and proposed with references to more detailed algorithms or code in Appendices, or elsewhere.

While many of the below modelling is focused on the accuracy vs. privacy loss trade-off, the concepts are certainly applicable to others via revisions in code and frameworks utilized. I selected this trade-off as a starting point due to its robust mathematical analysis in Abowd and Schmutte (2018), and logical ease of expansion with my proposed maximization approach.

Coding Languages and Software Used

For the purposes of developing and testing my approaches, I utilized the computer language Python version 3.12 (Van Rossum & Drake, 2009). Software code was written, compiled, tested and plotted using The Scientific Python Development Environment version 5.5, also known as Spyder (Spyder-IDE.org) (Raybaut, 2009), and Microsoft Visual Studio Code version 1.87.2 (Microsoft, 2024).

Maximization Model

Zooming in on Concavity Region Area of Interest and Intervals

I will begin by modelling and codifying a basic positive concave down function akin to the Planning Problem in Abowd and Schmutte (2018). In their analysis, they seek to maximize accuracy given some amount of privacy loss within the set of $G(\varepsilon, I)$ statistical agency activities. One simple form of a function that yields such a concave down curve is:

$$y(x) = \frac{1}{2}x^3 - x^2 + x + 1$$

While the three numerical coefficients ½, -1 and +1 are not particularly important in my analysis, their sign is. As we saw previously, a positive concave down function undergoes a change in sign in its second derivative at *or near* a point of interest $O(\varepsilon, I)$. To determine the interval over which the curve is concave up or concave down, a second derivative test is necessary (Bronshtein et al., 2007), and the concavity of the function changes at *all* points where its second derivative changes sign. For a third order function, this may occur at more than one location.

$$y'(x) = \frac{d}{dx}\left(\frac{1}{2}x^3 - x^2 + x + 1\right) = \frac{3}{2}x^2 - 2x + x$$

$$y''(x) = \frac{d}{dx}\left(\frac{3}{2}x^2 - 2x + x\right) = 3x - 2$$

Now, set $y''(x) = 0$ to find the points of interest (Bronshtein et al., 2007):

$$3x - 2 = 0$$

$$3x = 2$$

$$x = \frac{2}{3}$$

To determine the concavity intervals, test points in the interval determined by the critical point x=2/3, thereby dividing the curve into two intervals. Using the second derivative, the result is: on the interval $x = \left(-\infty, \frac{2}{3}\right)$ the curve is concave down, then becomes concave up from $x = \left(\frac{2}{3}, \infty\right)$. There are many other similar curves and one can easily alter the numerical coefficients. An example of Python code to plot such a function is available in Appendix D. The reader will note the definition of a cubic production function with the requisite parameters:

```
def cubic_production_function(input_variable, a=1/2, b=-1, c=1, d=1)

output_variable = a * input_variable**3 + b * input_variable**2 + c * input_variable + d
```

This allows for the plotted visualization of a segment of the curve that exhibits concave down shape between the above-noted value range of x. For simplicity, I have used a limited global range of x as -5 ≤ x ≤ 5.

Now taking a more general form of a cubic polynomial function:

$$y(x) = ax^3 - bx^2 + cx + d$$

$$y'(x) = 3ax^2 - 2bx + c$$

$$y''(x) = 6ax - 2b$$

Critical points occur where $y''(x) = 0$, giving us:

$$6ax = 2b$$

$$x = \frac{b}{3a}$$

This function is of order three, as stipulated above, and has changes in slope and sign in different intervals. Again using the second derivative, and depending in the signs of a and b, the concavity regions or intervals are as follows:

- If a>0 and b>0, the curve is concave down on (−∞,b/3a) and concave up on (b/3a,∞).
- If a>0 and b<0, the curve is concave up on (−∞,b/3a) and concave down on (b/3a,∞).
- If a<0, the concavity of the curve is reversed from the above.

This pattern creates a region of interest. In addition the general form of a cubic polynomial, other forms of functions that exhibit a positive concave down curve/form are as follows:

Vertex: $y(x) = (x - h)^3 + k$

In this form, *h* and *k* determine the position of the vertex or local maximum/minimum. Adjusting *h* will shift the curve horizontally, while adjusting *k* will shift it vertically.

Intercept: $y(x) = a(x - p)(z - q)(x - r)$

Here, *p*, q, and *r* are the x-intercepts of the function. Adjusting these intercepts will change the position of the roots and thus the shape of the curve.

Factored: $y(x) = ax(x - p)(x - q)$

Here, p and q are the x-intercepts, and a scales the function. Adjusting a changes the steepness of the curve, while adjusting p and q changes the position of the roots of the function. In all the above forms, including the general form, the resulting y(x) is third (3^{rd}) order function. The selection of most appropriate form is in part dependent on the number of intercepts of the curve and the interaction of variables in the function.

For my purposes here, I'll model the increasing (positive) concave down curve in Abowd and Schmutte (2018, p. 2) in vertex form due to there being no other variables considered other than ε and I. The result is a positive concave down curve where the x-axis variable is > 0, i.e. $\varepsilon > 0$. One form of such a function, to model their PF in shape and behaviour, is:

$$O(\varepsilon, I) = (1 - \varepsilon)^I$$

However, I need to take the negative to ensure concave down where $\varepsilon > 0$.

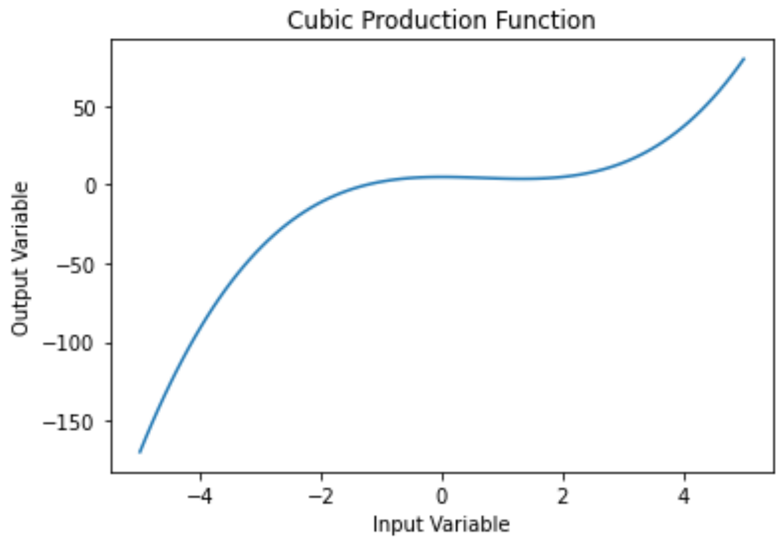
$$O(\varepsilon, I) = -(1 - \varepsilon)^I = -(1 - e)^I$$
$$\equiv O(\varepsilon, I) = 1 - (1 - \varepsilon)^I = 1 - (1 - e)^I$$

Cubic and Concave Down Production Functions

I have previously asserted that we're interested in at most 3^{rd} order cubic functions. The production function modelled in Abowd and Schmutte (2018) is such a function. To begin this analysis, I need to create the same—Appendix D contains a piece of simple python code to model a cubic production function with a single input variable and single output variable, as a trade-off. On this curve, we're interested in the portion of the curve where the curve is concave down. The plot of this function is Figure 5 below:

Figure 5

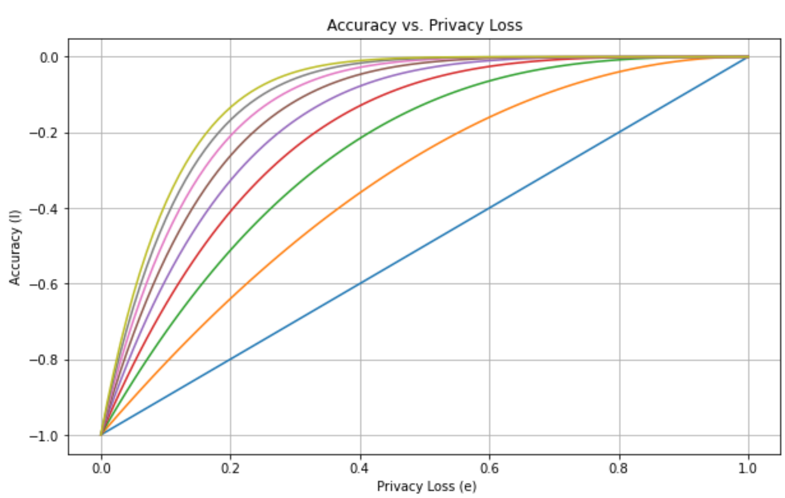
Plot of Generic Cubic Function



Next, we're interested in plotting a positive concave down function, like in Abowd and Schmutte (2018). I vary the I in $G(\varepsilon, I)$ to demonstrate different shapes of the production function. The plot of this function is Figure 6 below:

Figure 6

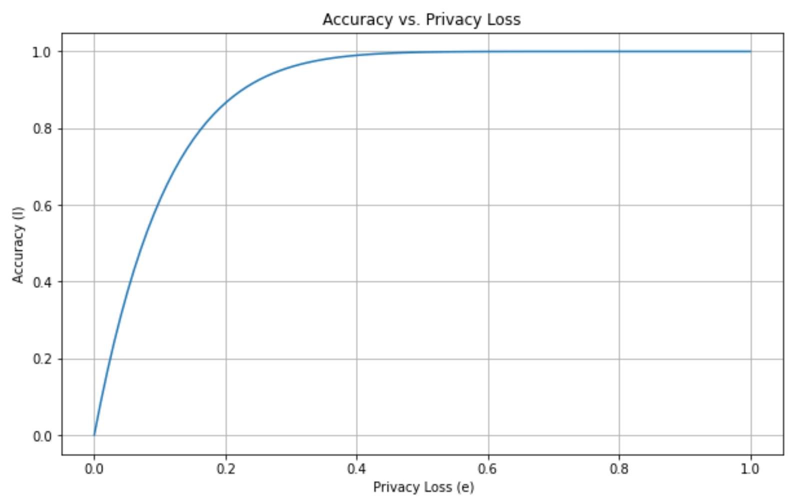
Plot of Positive Concave Down Functions



Next, I enhance the above plot to include M(*e,I*) closed set of machine language options, and N(*e.I*) closed set of computational resources such as CPU, memory, network. A plot of this function is Figure 7 below:

Figure 7

Plot of Positive Concave Down Function



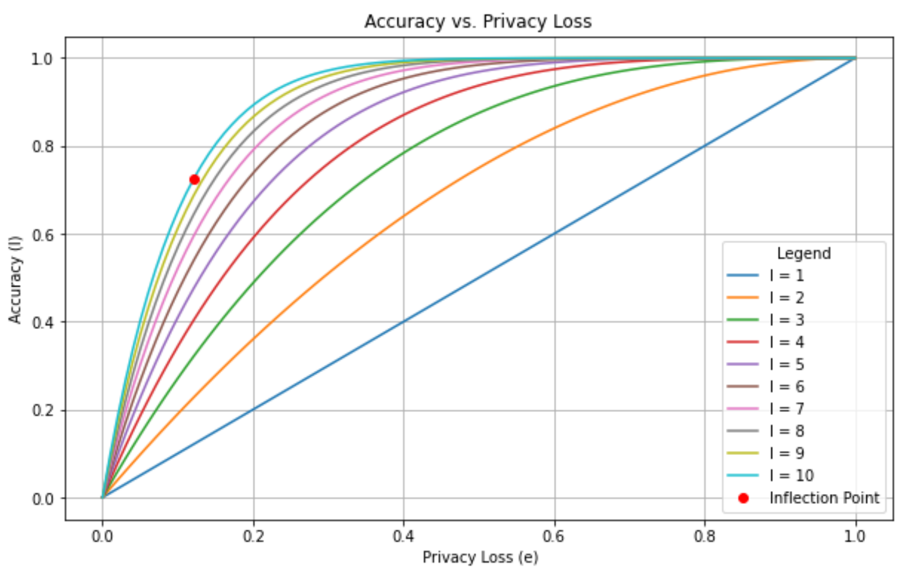
For my purposes here, I have simply provisioned for defined placeholder functions for M(e,I) and N(e,I) in the code. Practitioners should replace these functions with actual functions that represent the impact of machine learning options and computational resources on their production function. For example, choice of M(e,I) may be modelled to steepen the curve before or in the vicinity of the O*(e,I) where the production function intersects the SWF, thereby increasing the accuracy I for a given magnitude of privacy loss e. Likewise, enhanced N(e,I) can be modelling for similar effect, as has been observed with the use of vast computational resources and the impact on the accuracy and speed dimensions.

Inflection Points as Areas/Points of Interest

I've then adapted the code to calculate and display inflection point(s) on the curve using the second derivative test, i.e. the location on the frontier where $f''(x)$ changes sign. The plot of this function is Figure 8 below:

Figure 8

Plot of Positive Concave Down Functions with Identified Inflection Point



Genetic Evolutionary Algorithm Intervention to Increase O*(ε, I)

An AI technique particularly suited to optimization through maximization is evolutionary or genetic algorithms, applied as a maximization utility around the above curves and inflection point(s).

No Free Lunch Theorem. Returning to the concept no free lunch, Wolpert and Macready (1997) posit in their *no free lunch theorem* of optimization, that all optimization techniques and approaches are equally effective against the set of *all* optimization problems. This can be extended to evolutionary algorithms in general, but only if the set of problems is restricted. In other words, when the set of all problems is restricted, no evolutionary algorithm is fundamentally superior than another. They suggest that some form of problem limitation, or other form of upfront problem knowledge—a form of problem limitation—inevitably takes place in practice. An improvement in the performance of an evolutionary algorithm is predicated on exploitation of some problem knowledge, and the latter also implicitly drives comparative differences in algorithm performance. I won't go into details here, but there any many ways to introduce problem or problem-specific knowledge, e.g. non-random generation of the first-generation population, heuristics or other local search procedures in offspring generation, etc. Such approaches allow for gains in speed of search, while increasing robustness, up to a local or global maximum.

In effect, the above no free lunch theorem leads to the conclusion that, 'for certain types of computational problems, the computational cost of finding a solution, averaged over all problems in the class, is the same for any solution method.'[48] The implication of this for this study, is that the choice of evolutionary algorithm is perhaps less critical than the set or class of problems such algorithms are deployed against.

Are There Other Techniques?

Other AI techniques may be feasible in addition to genetic evolutionary algorithms, but the approach must be modified because deep-learning based techniques are more suited for minimizations than maximizations. That said, logic can be inverted or negated for maximization

---

[48] https://en.wikipedia.org/wiki/No_free_lunch_in_search_and_optimization, retrieved Jan 23, 2024.

purposes. Some of these have been introduced above, e.g. GAN and RAG, and other feasible and applicable options include: small language models (SLMs), GFNs, and MoEs. These are defined with utility examples in Appendix A. Other hybrid or quasi-approaches also exist, but will not be explored in this study.

Having explored types and locations of computational interventions, I now turn to a categorization model, whereby organizations can better select between AI techniques based on several factors and considerations.

Three-Level Categorization for Selecting AI Techniques

The very inception and design—computational problem being solved and associated solution approach—of different techniques creates the possibility of categories for deployment. This indeed holds and I propose a three-level categorization of organization problems and goals, against which different AI techniques should be deployed for higher performance against specific dimensions.

Due to the persistence of trade-offs despite the significant gains made, categories help optimize outcomes. It's also clear that while these proposed categories may be valid, there is a strong likelihood that they will change as AI techniques advance and new techniques are developed. This categorization is based on a combination of grounding of different AI techniques against different problem types, and also my assessment of their respective performance against same.

Multi-Level Categorization

I propose a multi-level categorization while offering organizations the ability to try and select discrete or inter-/intra-level categories based on their specific problem/goal. They can find themselves inside a category coordinate (x, y, z), e.g. organization Z_1 using an AI technique to perform a simple classification, while organization Z_2 is using an AI technique to perform an optimization *using* generation and retrieval *within* the topic of healthcare. Likewise, many permutations and combinations are available as optimal starting points for organizations to select and deploy different AI techniques.

Level 1: Trade-off Tolerance by Problem Context. Tolerance for trade-offs and the context of AI technique deployment is perhaps the most important and foundational category for AI technique selection. To take an extreme example, what works in mission critical or life and death contexts, is very different than what may be acceptable for recommendation-type optimizations in what movie to watch. This categorization level in Table 11 involves an organization making a context-based determination of what magnitude of trade-off tolerance they can tolerate for the problem or goal at hand, or generally for their purposes.

Table 11

Level 1: Trade-off Tolerance by Problem Context

	Trade-off Dimension Tolerance			
Topic or Context (from Table 8)	**High**	**Medium**	**Low**	***Zero (life:death, mission-critical)***
Healthcare				X

	Trade-off Dimension Tolerance			
Topic or Context (from Table 8)	**High**	**Medium**	**Low**	***Zero (life:death, mission-critical)***
Public Policy and Law			X	
Finance and Insurance			X	
ICT and Telecom			X	
Consumer Products	X			
Social Media	X			
Natural and Earth Sciences		X		
Security Services			X	

Level 2: Trade-off Dimensions by Problem Context. The next level I suggest is to designate which trade-off dimensions should be considered in the same context as Level 1. If we suppose that magnitude of tolerance is zero or near-zero within the context of health decision-making, then Level 2 asks the organization to determine which of the seven trade-off dimensions are intended to be maximized, and this naturally creates a set of corresponding trade-offs. To take a relatively simpler example from Table 12: within the context of Consumer Products, e.g. product recommendations, the dimensions of accuracy, privacy, reliability and speed would create a matrix of 4-choose-2 → 6 trade-offs (accuracy vs. privacy, accuracy vs. reliability, accuracy vs. speed, privacy vs. reliability, privacy vs. speed, and reliability vs. speed).

Table 12

Level 2: Trade-off Dimensions by Problem Context

	Trade-off Dimension Importance (X = High)						
Topic or Context	**Accuracy**	**Explainability**	**Fairness**	**Privacy**	**Reliability**	**Security**	**Speed**
Healthcare	X	X	X	X	X	X	
Public Policy and Law	X	X	X	X	X	X	X
Finance and Insurance	X			X	X	X	X
ICT and Telecom	X				X	X	X
Consumer Products	X			X	X		X
Social Media							X
Natural and Earth Sciences	X	X			X		X
Security Services	X					X	X

Level 3: AI Technique by Trade-off Dimension and Problem Context. Having analyzed the previous two levels, organizations can select AI technique(s) based on their feasibility and viability against both previous *sets of choices*. Taking the same Consumer Products example from Table 12 above, Table 13 portrays AI technique recommendations by trade-off dimension and context. For example, in some cases LMs present an optimal AI technique choice, while in other cases, other AI techniques are more optimal for different trade-off dimensions.

Table 13

Level 3: AI Technique by Trade-off Dimension and Problem Context

	Important Trade-off Dimensions						
Topic or Context	**Accuracy**	**Explainability**	**Fairness**	**Privacy**	**Reliability**	**Security**	**Speed**
Health	LM + GAN + RAG	Non-DL, RAG	RL	RAG	RAG	Internal LM	
Public Policy and Law	Internal LM + RAG	Non-DL, RAG	RL	RAG	RAG	Internal LM	RAG, MoE
Finance and Insurance	Internal LM + RAG				RAG	Internal LM	RAG, MoE
ICT and Telecom	Internal LM + RAG				RAG	Internal LM	RAG, MoE
Consumer Products	RAG				RL		LM
Social Media							LM
Natural and Earth Sciences	LM + RAG	Non-DL, RAG					
Security Services	LM + Gan + RAG					Internal LM	LM, MoE

Decision Trees

Another way to work through a multi-level categorization matrix is via decision trees, such as the examples I present below in Figures 9-12. Organizations can map their problems/goals or other intents based on a set of answers to questions. They benefit from suggested AI techniques or algorithmic architectures.

Figure 9

Decision Tree 1

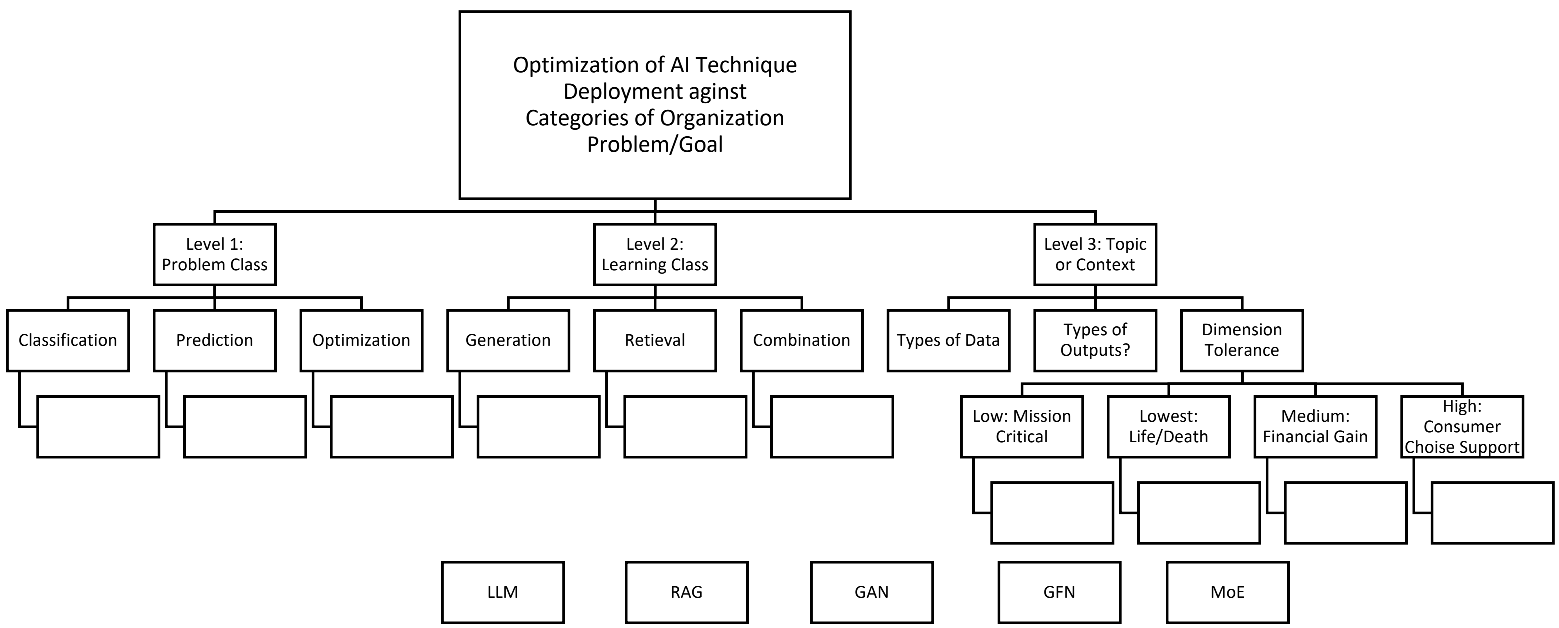


Figure 10

Decision Tree 2

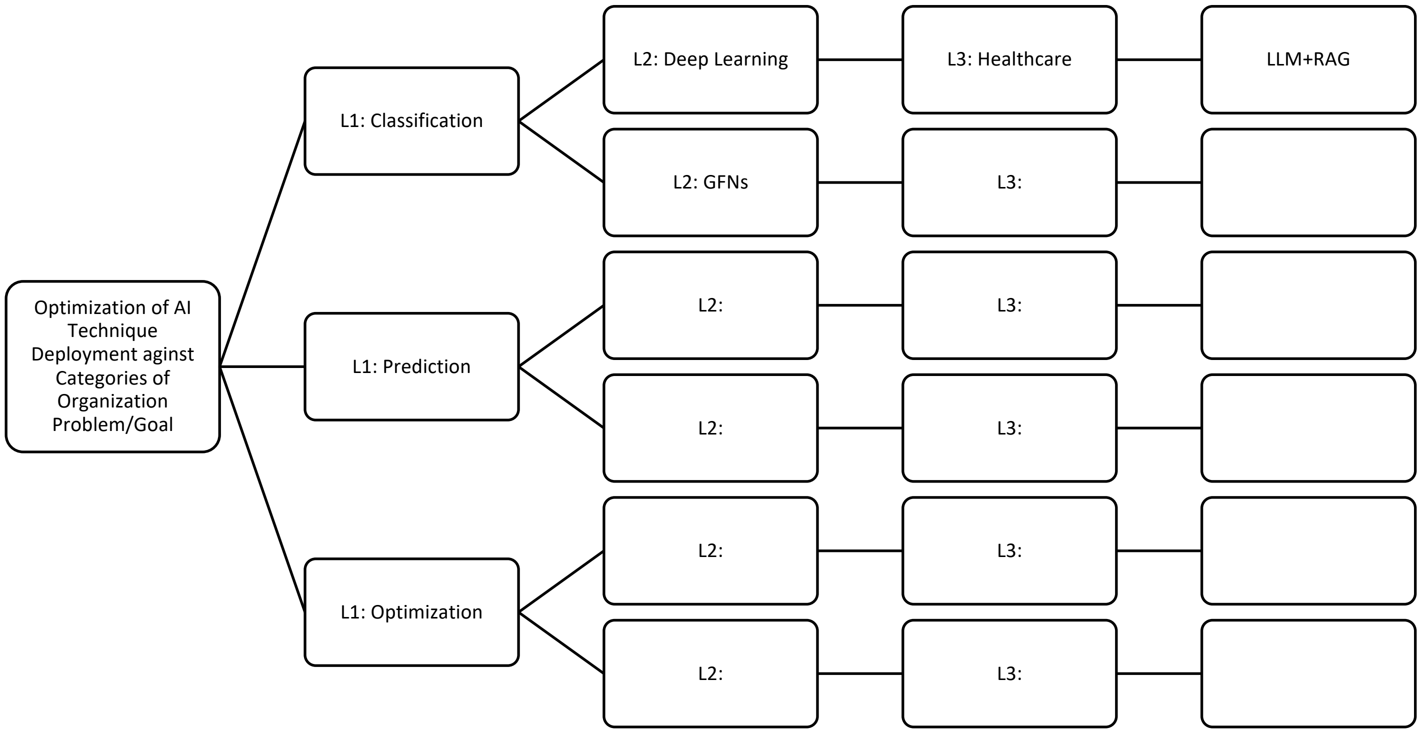


Figure 11

Decision Tree 3

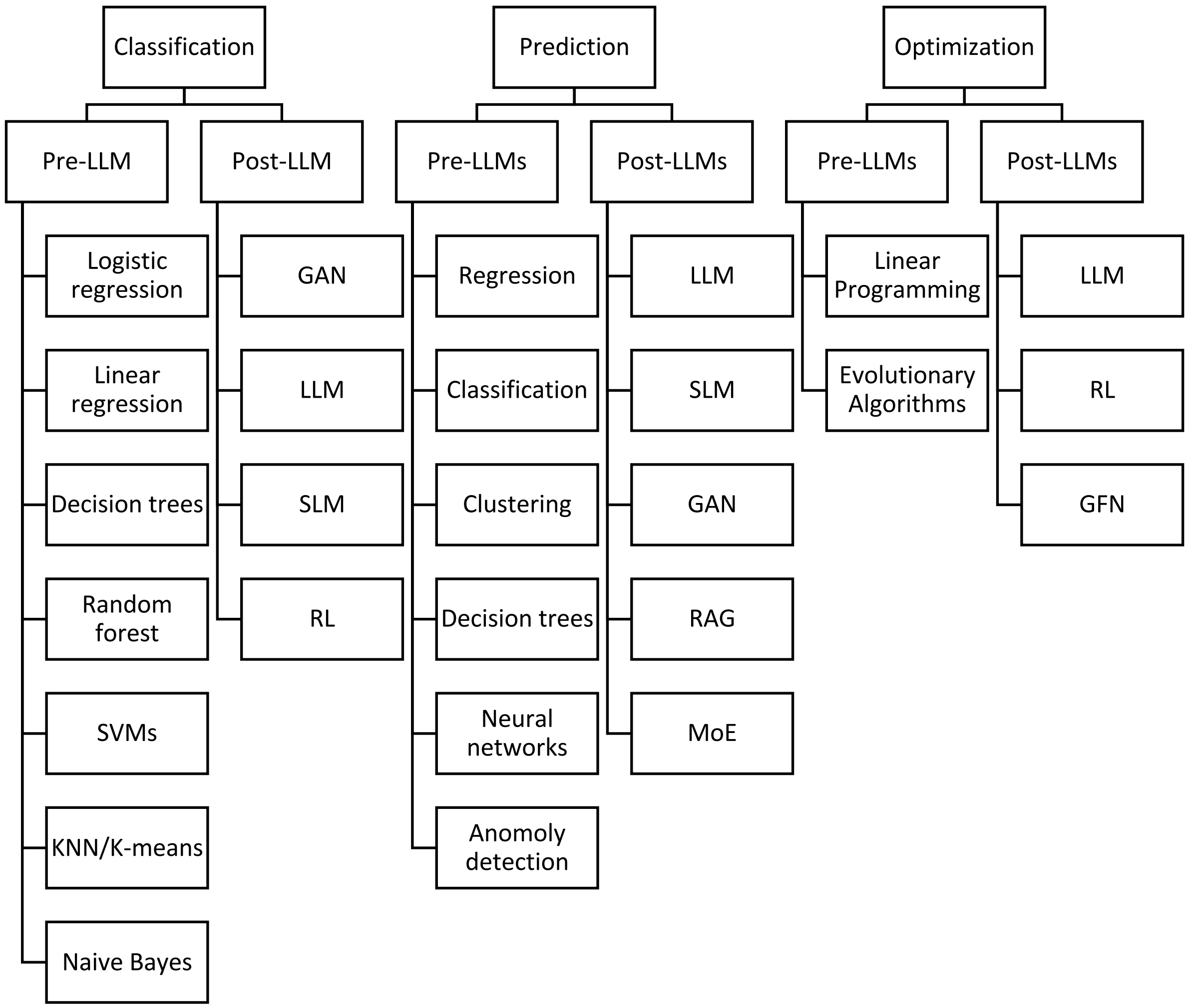
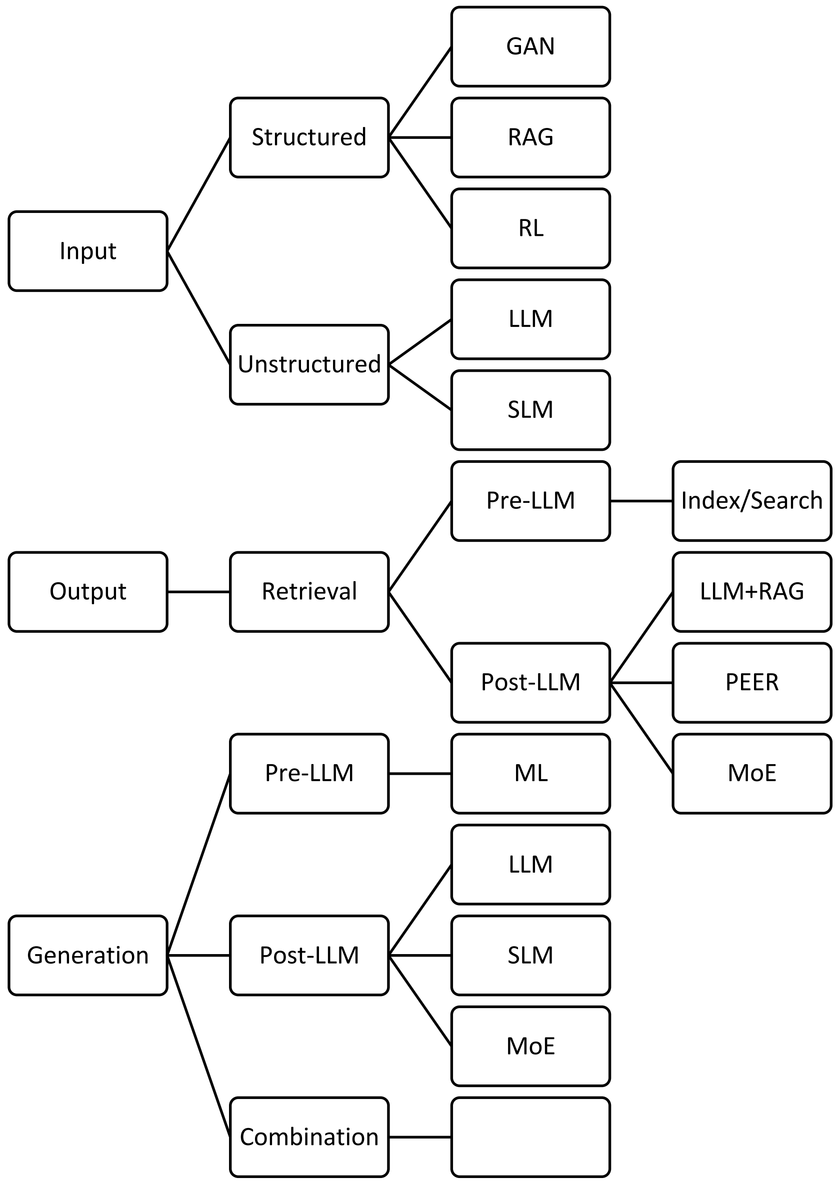


Figure 12

Decision Tree 4



Colloquial Attributes

Tokens and Parameters

I consider colloquial attributes as qualitative reflections of main areas in the literature reviewed. I will examine and make use of the commonly used attributes of tokens and parameters. Since the advent of LLMs and other generative technologies that leverage these language models (LMs), attributes commonly used as proxies for the robustness or capabilities of a model include the number of parameters or tokens[49]. Like other MLs before them, LMs such as ChatGPT are pre-trained on data. For pre-training data types like text, LMs break down this textual information into tokens. A token is considered to be approximately half to three-quarters the size of an average English word[50]. This is based on the notion that English words—or other languages for that matter—employ concatenation and other linguistic mechanisms in word construction. There are limits to how many total tokens any LM can process. For example, it is believed that ChatGPT-4 can process a total context length of 32,000 tokens, which would equate to approximately 24,000 English words[68,51]. This is the total amount of input a user can submit to a ChatGPT-4 prompt.

Parameters are what determines how a LM can process these tokens. While there are differences based on algorithms and architectures, parameters are the internal settings and optimizations inside a LM that determine how it takes inputs and predicts textual outputs, as a

---

[49] https://explodingtopics.com/blog/gpt-parameters, retrieved Nov 16, 2024.

[50] https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them, retrieved Nov 16, 2024.

[51] https://explodingtopics.com/blog/gpt-parameters, https://lingarogroup.com/blog/whats-new-with-gpt-4-features-and-limitations, https://vivekupadhyay1.medium.com/openai-chat-gpt-4-32k-detailed-document-ce3486e4786e; all retrieved Feb 26, 2025.

sequence of predicted individual words (or other forms of output such as pixels, or binary data such as sound), They are a quasi-discrete unit inside the DL model, and sometimes compared to neurons in the human brain. As a comparison, the human brain has some 86 billion neurons[52]. The connections and interactions between these neurons are fundamental for everything the human brain, and indeed other parts of the human body, does. They interpret information and determine how to respond. Similarly, in training a large language model, parameters are features of the LLM that are adjusted in order to optimize the model's ability to predict the next token in a sequence. After each *training sequence*, the model learns and adjusts its parameters to correct any incorrect predictions in the previous textual output. Over millions to billions of iterations, these parameters get adjusted and optimized for that particular LM to maximize its output correctness—a proxy for accuracy as I define it.

Brief Survey of LLMs

The size difference in terms of number of parameters between order of magnitude difference in the parameter size of LMs from as recently as 2018 with the launch of ChatGPT-1, but more interestingly, between more recently developed models that are architected differently than their peers and competitors. For example, based on the below figures, ChatGPT-4 is 1.5 million times larger than ChatGPT-1 in terms of number of parameters: 1.76 trillion / 117 million. Table 14 ranks different LMs by number of parameters.

---

[52] https://hms.harvard.edu/news/new-field-neuroscience-aims-map-connections-brain, retrieved Jan 26, 2025.

Table 14

Language Models ranked by Number of Parameters[53,54] *(ascending)*

LM Name	Number of Parameters
GPT-1	117 million
GPT-2	1.5 billion
Gemini Nano-1	1.8 billion
Gemini Nano-2	3.25 billion
Llama 3 8b	8 billion
GPT-4o mini	8 billion*
Llama 3 70b	70 billion
Claude 2	130 billion
GPT-3	175 billion
Gemini Pro	500 billion
DeepSeek-R1	670 billion*
Gemini Ultra	1.1 trillion
GPT-4	8 x 220 billion → 1.76 trillion*
GPT-4o	>= 8 x 220 billion → 1.76 trillion*

Note: * = estimated based on various sources, but unverified due to lack of public disclosure.

Are More Parameters Always Better?

A LM or other deep-learning architecture is usually considered to be better at processing information—think of a larger computational engine or an engine with a greater number of performance-enhancing capabilities. Parameters work to enhance the performance of models. However, there are downsides and other types of trade-offs as the number of parameters grows.

[53] Extracted from *IGS 1395 Standards Landscape for GenAI and LLMs in Telecommunications v1.1.0 DRAFT (ANP-1161),* (B. Pasagic, personal communication, December 16, 2024).

[54] https://en.wikipedia.org/wiki/Large_language_model, retrieved Nov 16, 2024.

The most obvious of these is costs associated with training the models and optimizing the parameters. This has been a key driver in the advancement of development and commercialization of techniques such as Small Language Models (SLMs) and Mixture of Experts (MoEs) (Jacobs et al., 1991). These are techniques whereby forms of separation and specialization is employed to limit the [pre-]training effort while maximizing correctness and/or performance against limited sets of problem classes or topics. OpenAI's ChatGPT-4o mini, despite being a relatively cost-efficient small language model with far fewer parameters than ChatGPT-4, has demonstrated outperforming ChatGPT-4 on several benchmarks based on internal and external testing[55]. ChatGPT-4 itself is a different algorithmic architecture that its OpenAI predecessors, employing what is known as MoE architecture. While this is not a new or novel architecture, it is believed to be new to OpenAI. Instead of piling all the *system-wide* parameters into a single model, ChatGPT-4 is divided into eight smaller models, each composed of two experts[56]. In total, ChatGPT-4 has sixteen (16) *specialized* experts, each with a small number of parameters[57]. The interconnectedness of these smaller models is achieved through routing algorithms that employ NLP and NLU to route the prompt-entry to the most appropriate expert model(s), whichever is deemed to be most appropriate to compute a response. This routing is also an area of optimization since routing a request to an inappropriate expert is obviously suboptimal. In effect, MoE and other similar architectures are able to response to a

[55] https://explodingtopics.com/blog/gpt-parameters, retrieved Nov 16, 2024.

[56] https://platform.openai.com/docs/models, retrieved Jan 26, 2025.

[57] ChatGPT-4 is estimated to have roughly 1.8 trillion parameters, consisting of 8 models, with each internal model having approx. 220 billion parameters, thereby architected as a MoE with 16 experts of approx. 118 billion parameters each. This makes ChatGPT-4 several times smaller than ChatGPT-3, yet able to outperform it on several benchmark tests.

range of requests in a more cost-efficient and shorter timeframe—far fewer than the total number of parameters in the *system* are being used at any given time to respond to any given request.

Other Characteristics or Attributes?

While not widely covered in the literature, there are of course additional characteristics or attributes associated with other techniques. For example, in the architecture of GANS, the number of adversarial networks, or networks of networks can be modified. Different architectures or specializations may also between neural networks. In addition, the training data sets can be different between different neural networks. A proposal, tests and attempt to optimize such attributes is beyond the scope of this study, but may be an area of exploration for future research. There are also characteristics or attributes with respect to MoEs. We've seen in the exploration of some LMs that this is already in active development and optimization, e.g. architectural differences between earlier generation LLMs and MoE-based language models composed of specialized "expert" smaller models. These architectures also open up possibilities around optimizing the number of experts, how these are delineated, routing between *or within* experts, and so on.

Different AI techniques offer variation in how their architectures and mechanisms act to address different trade-off dimensions. MoEs were at least partially developed to mitigate the pre-training computational time and associated costs, which factor into my trade-off dimension of speed.

Experiment 1. Simulating Performance of AI Techniques using Colloquial Attributes

Several Monte Carlo simulations were designed and run on various AI techniques, ranging from LLMs only, to other individual techniques such as GANs and RAG, and combinations of the forgoing. I've run these simulations using colloquial dimensions, and again after converting these attributes to my seven trade-off dimensions.

Simulation Design

By extracting and synthesizing various baseline performance metrics in the literature (Eibich et al., 2024; Fabbri et al., 2021; Fu et al., 2023; He et al., 2023; Ip, 2025; Jiang et al., 2024; Li et al., 2024; Liu et al., 2023; Lo et al., 2024; Phan et al., 2025; Roychowdhury et al., 2024; Vats et al., 2024; Yan, 2024; Yun et al., 2024; Zhou et al., 2022), these first set of simulations allowed a comparison of several individual AI techniques: LLMs, GANs, RAG, GFNs, and MoE. The performance of each was measured by evaluating the effectiveness (a colloquial proxy for my dimension of accuracy) of each technique against the colloquial attributes of computational cost and applicability. It's assumed that the objective dimension (effectiveness) and trade-off dimensions (cost and applicability) are influenced by randomness. As such, these simulations were designed to capture this uncertainty, randomness and variability.

I simulated the application of each AI technique to strategic decision-making, such as identifying market opportunities, resource allocation, or risk management. Different iterations of minimum 1000 runs were simulated. The results were analyzed to compare performance across the different AI techniques, and synthesized with context or conditions under which each technique excels or performs poorly. For both non-converted and converted dimensions, AI techniques were evaluated as individual techniques and in combinations of two, three, four, and five techniques. Table 15 contains all possible combinations of the five AI techniques considers.

Effectiveness. A proxy for accuracy and quality of generated responses (scale: 0-1).

Computational Cost. Resources required (normalized between 0-1).

Applicability. How broadly the technique applies to various sub-problems (scale: 0-1).

Assumptions. All trade-off dimensions are influenced by randomness, e.g. variability in problem complexity, noisy training data, and other random disruptions. Each AI technique has a baseline performance and variability range. Scores are modeled as normal distributions.

Parameters. Different iterations of minimum 1000 runs. Randomly sample scores for each dimension across the different AI techniques.

Table 15

All Possible Combinations of Five (5) AI Techniques

Number of AI Techniques				
1	2	3	4	5
LLM	LLM+GAN	LLM+GAN+RAG	LLM+GAN+RAG+GFN	LLM+GAN+RAG +GFN+MOE
GAN	LLM+RAG	LLM+GAN+GFN	LLM+GAN+RAG+MOE	
RAG	LLM+GFN	LLM+GAN+MOE	LLM+GAN+GFN+MOE	
GFN	LLM+MOE	LLM+RAG+GFN	LLM+RAG+GFN+MOE	
MOE	GAN+RAG	LLM+RAG+MOE	GAN+RAG+GFN+MOE	
	GAN+GFN	LLM+GFN+MOE		
	GAN+MOE	GAN+RAG+GFN		
	RAG+GFN	GAN+RAG+MOE		
	RAG+MOE	GAN+GFN+MOE		
	GFN+MOE	RAG+GFN+MOE		

Results of Simulation of Individual AI Techniques

Based on the results of the simulation (summarized in Table 16 below), RAG scored the highest on average (0.90 ± 0.05) on the dimension of Effectiveness, making it the most effective for generating solutions, while GAN has the lowest effectiveness (0.75 ± 0.05). On the trade-off dimension of Computational Cost, LLMs and MoE had lower computational costs, while RAG and GAN were more resource intensive. On Applicability, LLMs were the most broadly applicable (0.90 ± 0.07), while GANs were the least (0.60 ± 0.08).

Analysis of boxplots of the above also reveal interesting insights (Figure 13 below). On Effectiveness, RAG consistently outperformed other techniques, while GANs showed the largest variability, suggesting sensitivity to problem complexity. LLMs and GFNs maintained lower Computational Costs, but RAG, required the most resources. As seen above, LLMs demonstrated the highest Applicability to various scenarios, making them suitable for diverse problems.

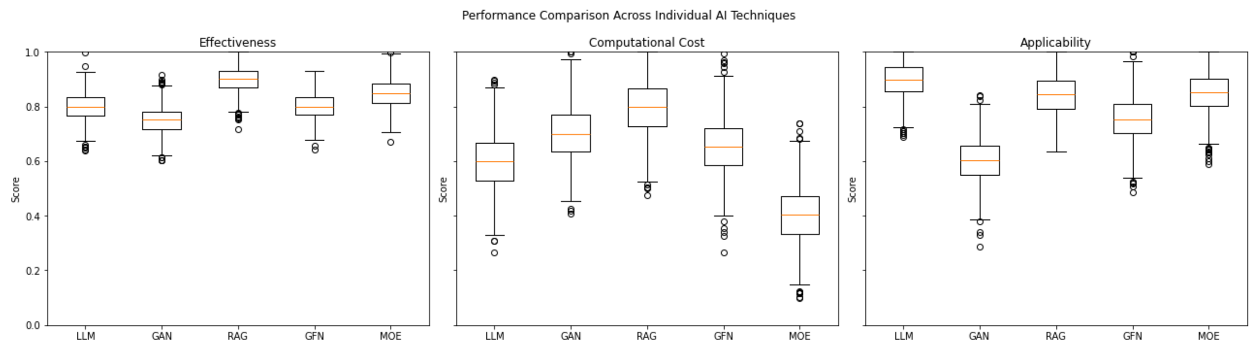
Table 16

Summary Table of Results of Monte Carlo Simulation of Individual Techniques

	Dimension		
AI Technique	Effectiveness	Computational Cost	Applicability
LLM	0.80 ± 0.05	0.60 ± 0.10	0.90 ± 0.06
GAN	0.75 ± 0.05	0.70 ± 0.10	0.60 ± 0.08
RAG	0.90 ± 0.05	0.80 ± 0.10	0.84 ± 0.08
GFN	0.80 ± 0.05	0.65 ± 0.10	0.75 ± 0.08
MoE	0.85 ± 0.05	0.40 ± 0.10	0.85 ± 0.08

Figure 13

Boxplots of Results of Monte Carlo Simulation of Individual AI Techniques (Colloquial Attributes)



Combinations of AI Techniques

To simulate the performance of all combinations of these AI techniques, I extended the Monte Carlo simulations for all possible combinations of two, three, four and all five of the techniques, totaling 26 possible combinations. The same colloquial trade-off dimensions are considered as above, and analyze their trade-offs.

Parameters. Different iterations of minimum 1000 runs. Randomly sample scores for each dimension across the different combinations of AI techniques. For each combination, compute the *combined* mean and standard deviation of the selected techniques' dimensions, then sample performance scores.

Results of Simulation of Combinations of AI Techniques

Based on the results of the simulation (Table 17 below), the combinations LLM+RAG (0.88 ± 0.05) and RAG+GFN (0.85 ± 0.05) have the highest scores on the dimension of Effectiveness. Adding more techniques appears to generally balance the effectiveness but does not exceed the baseline set by RAG. The lowest Computational Cost is achieved by a combination of LLM+GFN (0.62 ± 0.10). Combinations involving RAG and GAN, e.g., RAG (0.80 ± 0.10) and GAN+RAG (0.75 ± 0.10)) have the highest Computational Costs. The combination LLM+GFN appears to offer an optimum trade-off between Effectiveness and Computational Cost. In terms of Applicability, combinations LLM+RAG (0.87 ± 0.07) and LLM+RAG+GFN (0.84 ± 0.07) maintain high applicability, while combinations involving GAN generally score lower in this dimension.

Analyzing multi-dimension balance, which means to find an optimal balance of more than two dimensions (all three in this case), a combination of LLM+RAG has high Effectiveness (0.88 ± 0.05). moderate Computational Cost (0.70 ± 0.10). and high Applicability (0.87 ± 0.07). A combination of RAG+MOE has high Effectiveness (0.88 ± 0.05), low Computational Cost (0.60 ± 0.10). and high Applicability (0.86 ± 0.08). Table 17 below summarizes these results.

Analysis of boxplots of the above also reveal interesting insights (Figure 14 below). On Effectiveness, RAG consistently outperformed other techniques, while GANs showed the largest variability, suggesting sensitivity to problem complexity. LLMs and GFNs maintained lower Computational Costs, but RAG, required the most resources. As seen above, LLMs demonstrated the highest Applicability to various scenarios, making them suitable for diverse problems.

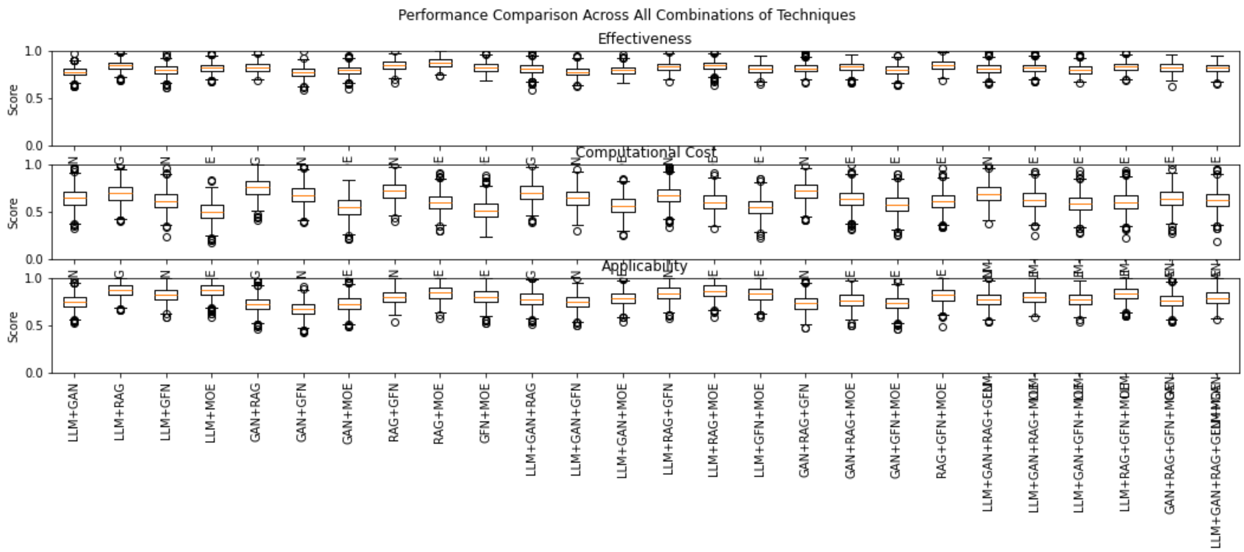
Table 17

Summary Table of Results of the Monte Carlo Simulation of AI Technique Combinations

	Dimension		
AI Technique Combination	Effectiveness	Computational Cost	Applicability
LLM+GAN	0.78 ± 0.05	0.65 ± 0.10	0.75 ± 0.07
LLM+RAG	0.88 ± 0.05	0.70 ± 0.10	0.87 ± 0.07
LLM+ GFN	0.80 ± 0.05	0.62 ± 0.10	0.82 ± 0.07
LLM+MOE	0.83 ± 0.05	0.50 ± 0.10	0.87 ± 0.07
GAN+RAG	0.82 ± 0.05	0.80 ± 0.10	0.72 ± 0.08
GAN+ GFN	0.77 ± 0.05	0.68 ± 0.10	0.67 ± 0.08
GAN+MOE	0.80 ± 0.05	0.55 ± 0.10	0.72 ± 0.08
RAG+GFN	0.85 ± 0.05	0.72 ± 0.10	0.80 ± 0.08
RAG+MOE	0.88 ± 0.05	0.60 ± 0.10	0.86 ± 0.08
GFN +MOE	0.82 ± 0.05	0.53 ± 0.10	0.80 ± 0.08
LLM+GAN+RAG	0.81 ± 0.05	0.70 ± 0.10	0.78 ± 0.08
LLM+GAN+ GFN	0.78 ± 0.05	0.65 ± 0.10	0.75 ± 0.08
LLM+GAN+MOE	0.80 ± 0.05	0.57 ± 0.10	0.78 ± 0.08
LLM+RAG+GFN	0.83 ± 0.05	0.68 ± 0.10	0.84 ± 0.08
LLM+RAG+MOE	0.85 ± 0.05	0.60 ± 0.10	0.87 ± 0.07
LLM+ GFN+MOE	0.82 ± 0.05	0.55 ± 0.10	0.83 ± 0.08
GAN+RAG+GFN	0.82 ± 0.05	0.72 ± 0.10	0.73 ± 0.08
GAN+RAG+MOE	0.83 ± 0.05	0.63 ± 0.10	0.77 ± 0.08
GAN+GFN+MOE	0.80 ± 0.05	0.58 ± 0.10	0.73 ± 0.08
RAG+ GFN+MOE	0.85 ± 0.05	0.62 ± 0.10	0.82 ± 0.08
LLM+GAN+RAG+GFN	0.81 ± 0.05	0.68 ± 0.10	0.78 ± 0.08
LLM+GAN+RAG+MOE	0.82 ± 0.05	0.62 ± 0.10	0.80 ± 0.08
LLM+GAN+GFN+MOE	0.80 ± 0.05	0.59 ± 0.10	0.78 ± 0.08
LLM+RAG+GFN+MOE	0.84 ± 0.05	0.61 ± 0.10	0.84 ± 0.08
GAN+RAG+GFN+MOE	0.82 ± 0.05	0.63 ± 0.10	0.77 ± 0.08
LLM+GAN+RAG+GFN+MOE	0.82 ± 0.05	0.63 ± 0.10	0.79 ± 0.08

Figure 14

Boxplots of Results of Monte Carlo Simulation of Combinations of AI Techniques (Colloquial Attributes)



Conversion from Colloquial Technique Attributes to the Seven Trade-off Dimensions

Colloquial Attributes are insufficient to capture the underlying reasons for suboptimal results in AI technique deployments. The set of seven dimensions are not captured in attributes such as parameters or tokens, or other dimensions such as computational cost or effectiveness. Therefore, a mapping is necessary from these abstractions and computational dimensions to the 7, which have a meeting in both computer science and management science. The driving *independent* variables, e.g. technique attributes, are subject to change. Even if my conversion is reliable and valid, it is subject to change as the driving independent variables (technique attributes) are frequently changing in definition and magnitude. Of importance is the need for a conversion, not necessarily my proposed conversion. Even as AI technique/model attributes and characteristics change over time, I've not found any published work that substantively presents alternatives to my seven trade-off dimensions—as I state in detail, most relevant studies simply cover a subset of the seven. Unless my seven trade-off dimensions and the twenty-one pairwise trade-offs between are widely understood and analyzed as a comprehensive set or as some sort of more complicated manifold, publicly-disclosed/used attributes must be translated into the seven trade-off dimensions I focus on. I posit that the need for a conversion is more important than the conversion itself.

Why Care if not Building or Training a Model?

We've seen that amongst other dimensions, parameters are being used a significant proxy for accuracy, but analysis has also shown that using models with more parameters may *or may not* lead to more accurate results. While an increase in parameters intuitively aligns with notion of increased accuracy, other trade-offs such as speed (time to train), are pushing developers to seek other approaches, e.g. MoE and sub-techniques therein. Unlike *democratized* search, which is fundamentally driven by the ability to maximize both the input field and speed of indexing, LLMs are closed systems limited to their training set and access controls. Cost to train and re-train is immense, now entering realm of tens of billions of dollars. Cost to access is also increasing, as are costs to operate, etc. Many organizations are trying to build or at least their own models as a way to mitigate perceived impacts of the trade-offs of privacy and security.

Conversion Mappings

Underscoring a finding from above, the ever-increasing parameter count of LLMs is based on the underlying premise that this correlates with dimensions such as accuracy, performance and confidence (using colloquial dimensions). How do these convert and map to my seven trade-off dimensions? Let's take a look at some of these in detail and connect empirically observed behavior with my trade-off dimensions.

Explainability. Hallucinations in AI technique deployments are phenomena where an explainable AI technique like a LLM, generates factually incorrect or entirely fabricated outputs. This can either be for many reasons, including that the AI model wasn't trained on enough data or because it was trained on only one type of data, which makes it create solutions or results that match *its* dataset.

Security. LLMs present important security risks when not managed or surveilled properly. They can leak people's private information, participate in phishing scams, and produce spam. Users with malicious intent can reprogram algorithms with malicious intent, and contribute to the spread of misinformation.

Fairness. The data used to train LMs will affect the outputs a given model produces. As such, if the data represents a single demographic, or lacks diversity, the outputs produced by the LM will also lack diversity. This perpetuation of biases present in the training data can lead to unfair or discriminatory outcomes.

Privacy. LLMs are trained on trillions of datasets, some of which might not have been obtained consensually. When scraping data from the internet, for example, LLMs can ignore copyright licenses, plagiarize written content, and repurpose proprietary content without getting permission from the original creator or owner. Then when it produces responses, there is no way to track data lineage, which can expose users to copyright infringement issues. Lack of robust protections can lead to training on data within public domains that should not have been, e.g. customer data, medical information or sensitive corporate details.

Scaling. It is difficult and time-/resource-consuming to scale and maintain LLMs.

Proposed Conversions and Weightings

The above examples point to the need for a conversion, both because these trade-offs persist, but also because it's suboptimal without conversion, to try and provide recommendations on AI technique utilization against different problems, or how to use AI techniques within complicated strategic decision-making processes.

In Table 18, I decompose the colloquial attribute of Number of Parameters against each of my seven trade-off dimensions. I suggest that there is varying interaction in each case, in both direction and magnitude, i.e. direct or inverse correlation of varying magnitude. The ratios I propose below are quantitative reflections of the interaction between the attribute of Parameters and my seven trade-off dimensions, and based on my analysis of the literature reviewed. As linear equations, this can be considered as follows, for example:

LLM = x1 for accuracy, y1 for speed, z1 for Dx

RAG = x2 for accuracy, y2 for speed, z2 for Dx

Table 18

Vector Mapping of Colloquial Attributes to Seven Trade-off Dimensions

Colloquial Attribute	Trade-off Dimension	Relationship	Ratio	Reasoning
Number of Parameters in Language Model	Accuracy	Correlated (↑ Parameters → ↑ Accuracy)	0.7	Perhaps single most correlated attribute, supports efforts and resources being spent here \$100M → \$1B → \$10B models → \$100B models?[58,59]
Number of Parameters in Language Model	Explainability	Inversely correlated	-0.3	Explainability? [likely ↓ since accuracy and explainability are inversely related, and an increase in parameters in-effect creates a *deeper* deep neural network] Ratio? [assume -0.3]
Number of Parameters in Language Model	Fairness	Correlated	0.2	↑ Parameters → Fairness ? [likely ↑ since an increase in parameters allows for a larger corpus of training data, which one can assume should be more comprehensive and less prone to bias] Ratio? [assume 0.2]
Number of Parameters in Language Model	Privacy	Inversely Correlated	-0.4	↑ Parameters → Privacy ? [likely ↓ since corpus training data is immense and may not have been obtained with consent or with safeguards to protect private information ... see above bullet on this] Ratio? [assume -0.4]
Number of Parameters in Language Model	Reliability	Correlated	0.6	↑ Parameters → Reliability? [likely ↑ since greater corpus of training data should lead to more consistent results against same training data

[58] Extracted from *IGS 1395 Standards Landscape for GenAI and LLMs in Telecommunications v1.1.0 DRAFT (ANP-1161),* (B. Pasagic, personal communication, December 16, 2024).
[59] https://en.wikipedia.org/wiki/Large_language_model, retrieved Nov 16, 2024.

Colloquial Attribute	**Trade-off Dimension**	**Relationship**	**Ratio**	**Reasoning**
				Ratio? [assume 0.6]
Number of Parameters in Language Model	Security	Inversely Correlated	-0.2	↑ Parameters → Security ? [likely ↓ since an increase in corpus training data may not have been obtained with consent or may inadvertently or otherwise cross into secure sources through vulnerabilities] Ratio? [assume -0.2]

Number of Parameters in Language Model	Speed	Pre-training: Inversely Correlated
		
		Post-training: Correlated
		
		Net Inversely Correlated
			-0.9
			
			0.6
			
			-0.5
				↑ Parameters → ↓ Speed (pre-training, post-training?, RL?)
				• Pre-training [known ↓ since increase in corpus data and parameters requires longer to compute and organize into language models or other structures such as vector DBs]
				  o Ratio? [assume -0.9 or worse?]
				  o Not linear but close, maybe worse?
				  o Other interventions being developed and introduced to non-linearize or decouple this
				• Post-training [likely ↑ due to greater volume of text encoded and embedded into vector DBs → lower time to access.
				  o Ratio? [assume 0.6]

The above reveals some interesting patterns. As seen in the literature, the quest for accuracy seems to the only dimension that is being *attacked* by most AI technique developers. The recent developments in MoE architectures are in in response to unsavoury trade-offs such as computational cost and pre-training time. For the thrust of number of parameters, the trade-off always appears to be accuracy vs D_x. And, as the accuracy dimensions is being focused on, other dimensions continue to be detrimentally impacted by accuracy gains via increased parameters. Given this, is the *overall* picture improving, or getting worse, or same? Below in Table 19, I present a comprehensive listing of different underlying AI technologies and AI techniques and my assessment of the impact of each on my seven trade-off dimensions.

Table 19

Impact of AI Technique on Seven Trade-off Dimensions

AI Technology or Technique	Dimension	Direction (↑ = improvement, ↓ = detriment)
Machine Learning	Accuracy	↑↓
	Explainability	↓
	Fairness	↓
	Privacy	↓
	Reliability	↑↓
	Security	↑↓
	Speed	↑
Deep Learning	Accuracy	↑
	Explainability	↓
	Fairness	↓
	Privacy	↑↓
	Reliability	↑
	Security	↑↓
	Speed	Pre ↓, Post ↑
LLM	Accuracy	↑↓
	Explainability	↓
	Fairness	↓
	Privacy	↓
	Reliability	↓
	Security	↑↓
	Speed	Pre ↓, Post ↑
SLM	Accuracy	↑↓

AI Technology or Technique	Dimension	Direction (↑ = improvement, ↓ = detriment)
	Explainability	↓
	Fairness	↓
	Privacy	↓
	Reliability	↓
	Security	↑↓
	Speed	Pre ↑, Post ↓
RL	Accuracy	↑
	Explainability	↑
	Fairness	↓
	Privacy	↑↓
	Reliability	↑
	Security	↑↓
	Speed	↑↓
GAN	Accuracy	↑
	Explainability	↓
	Fairness	↑↓
	Privacy	↑↓
	Reliability	↑↓
	Security	↑↓
	Speed	↑↓
RAG	Accuracy	↑
	Explainability	↑
	Fairness	↓
	Privacy	↑↓
	Reliability	↑
	Security	↑↓
	Speed	↑
GFN	Accuracy	↑
	Explainability	↓↓
	Fairness	↑↓
	Privacy	↑↓
	Reliability	↑↓
	Security	↑↓
	Speed	↑
MoE	Accuracy	↑
	Explainability	↓

AI Technology or Technique	**Dimension**	**Direction (↑ = improvement, ↓ = detriment)**
	Fairness	↑
	Privacy	↑
	Reliability	↑
	Security	↑
	Speed	↑

Experiment 2. Simulating Performance of AI Techniques using Converted Attributes

The following Monte Carlo simulations were designed and run on the AI techniques: LLM, GAN, RAG and MoE, as individual techniques and in combinations, after converting these attributes to my seven trade-off dimensions.

Results of Simulation of Individual AI Techniques with Converted Attributes

On the dimension of Accuracy, RAG scored the highest on average (0.80 ± 0.05), making it the most effective for generating solutions, while GANs had the lowest effectiveness (0.60 ± 0.05). On the trade-off dimension of Speed, MoE performed the best, while RAG and GAN performed the worst. Table 20 below summarizes these results.

Analysis of boxplots of the above also reveal interesting insights (Figure 15 below). On Accuracy, RAG outperformed other techniques on almost all dimensions, while MoE showed the smallest variability, suggesting less sensitivity to problem complexity. MoE also showed the highest speed, and combination of accuracy and speed.

Table 20

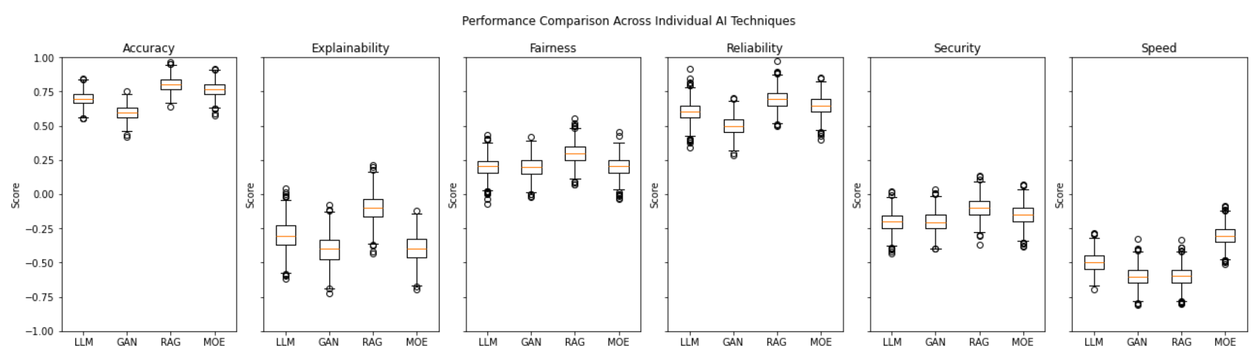
Summary Table of Results of the Monte Carlo Simulation of Individual Techniques (Converted Dimensions)

	Dimension					
AI Technique	Accuracy	Explainability	Fairness	Reliability	Security	Speed
LLM	0.70 ± 0.05	-0.30 ± 0.10	0.20 ± 0.07	0.60 ± 0.07	-0.20 ± 0.07	-0.50 ± 0.07
GAN	0.60 ± 0.05	-0.40 ± 0.10	0.20 ± 0.07	0.50 ± 0.07	-0.20 ± 0.07	-0.60 ± 0.07
RAG	0.80 ± 0.05	-0.10 ± 0.10	0.30 ± 0.07	0.70 ± 0.07	-0.10 ± 0.07	-0.60 ± 0.07
MOE	0.77 ± 0.05	-0.40 ± 0.10	0.20 ± 0.07	0.65 ± 0.07	-0.15 ± 0.07	-0.30 ± 0.07

Note: GFN technique was removed from simulation due to lack of basis to convert dimension for this technique. Privacy dimension was removed from simulation due to lack of basis to change score-mapping for this dimension.

Figure 15

Boxplots Results of Monte Carlo Simulation of Individual AI Techniques (Converted Dimensions)



Results of Simulation of Combinations of AI Techniques with Converted Attributes

To simulate the performance of all combinations of these AI techniques, I extended the Monte Carlo simulations for all possible combinations of two, three, and four of the techniques, totaling 15 possible combinations. The above colloquial attributes were converted to my set of seven trade-off dimensions.

Based on the results of the simulation (Table 21 below), the combinations RAG+MOE (0.78 ± 0.05) and LLM+RAG+MOE (0.76 ± 0.05) have the highest scores on the dimension of Accuracy. Adding more techniques appears to generally balance the effectiveness but does not exceed the baseline set by RAG.

Based on my construction of the simulation and conversion coefficient, all techniques suffer a negative Explainability. Given this, the best and worst performing combinations were LLM+RAG (-0.20 ± 0.10), and GAN+MOE (-0.40 ± 0.10), respectively. On the dimension of Speed, the combination of GAN+RAG performed the best (-0.39 ± 0.07), while GAN+MOE has the lowest score (-0.60 ± 0.07). On Speed, GAN+RAG performed the best (-0.39 ± 0.07).

Analyzing multi-dimension balance, a combination of RAG+MOE has high Accuracy (0.78 ± 0.05), relatively moderate Explainability (-0.25 ± 0.10), moderate Fairness (0.26 ± 0.07), moderate Reliability (0.67 ± 0.07), relatively higher Security (-0.13 ± 0.07) and Speed (-0.45 ± 0.07). Table 21 below summarizes these results.

Table 21

Summary Table of Results of the Monte Carlo Simulation of AI Technique Combinations (Converted Dimensions)

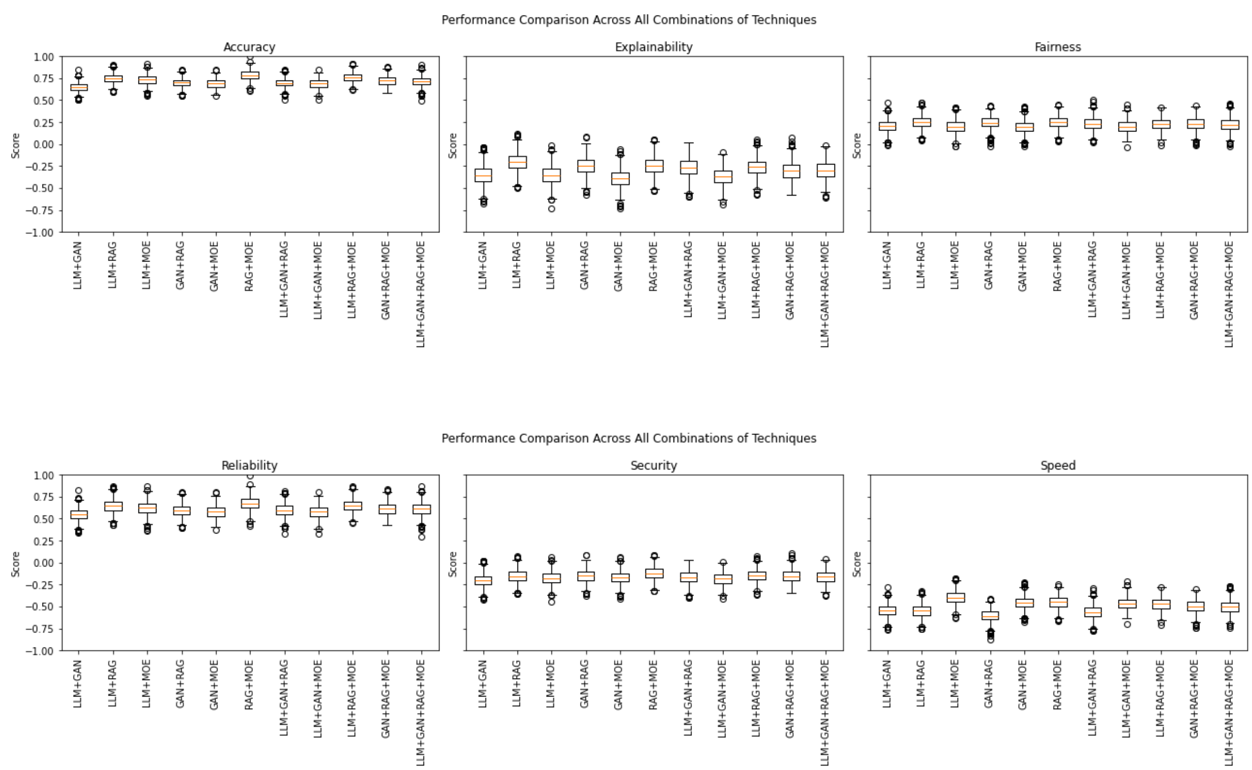
			Dimension			
AI Technique	Accuracy	Explainability	Fairness	Reliability	Security	Speed
LLM+GAN	0.65 ± 0.05	-0.35 ± 0.10	0.20 ± 0.07	0.55 ± 0.07	-0.20 ± 0.07	-0.55 ± 0.07
LLM+RAG	0.75 ± 0.05	-0.20 ± 0.10	0.25 ± 0.07	0.65 ± 0.07	-0.20 ± 0.07	-0.55 ± 0.07
LLM+MOE	0.74 ± 0.05	-0.35 ± 0.10	0.20 ± 0.07	0.62 ± 0.07	-0.15 ± 0.07	-0.55 ± 0.07
GAN+RAG	0.70 ± 0.05	-0.25 ± 0.10	0.25 ± 0.07	0.60 ± 0.07	-0.17 ± 0.07	-0.39 ± 0.07
GAN+MOE	0.69 ± 0.05	-0.40 ± 0.10	0.20 ± 0.07	0.57 ± 0.07	-0.15 ± 0.07	-0.60 ± 0.07
RAG+MOE	0.78 ± 0.05	-0.25 ± 0.10	0.26 ± 0.07	0.67 ± 0.07	-0.13 ± 0.07	-0.45 ± 0.07
LLM+GAN +RAG	0.70 ± 0.05	-0.26 ± 0.10	0.23 ± 0.07	0.60 ± 0.07	-0.17 ± 0.07	-0.57 ± 0.07

LLM+GAN +MOE	0.69 ± 0.05	-0.37 ± 0.10	0.20 ± 0.07	0.58 ± 0.07	-0.18 ± 0.07	-0.47 ± 0.07
LLM+RAG +MOE	0.76 ± 0.05	-0.27 ± 0.10	0.23 ± 0.07	0.65 ± 0.07	-0.15 ± 0.07	-0.47 ± 0.07
GAN+RAG +MOE	0.72 ± 0.05	-0.30 ± 0.09	0.23 ± 0.07	0.62 ± 0.07	-0.15 ± 0.07	-0.50 ± 0.07
LLM+GAN +RAG+MOE	0.72 ± 0.05	-0.30 ± 0.10	0.23 ± 0.07	0.61 ± 0.07	-0.16 ± 0.07	-0.50 ± 0.07

Note: The trade-off dimension of Privacy was removed from simulation due to lack of basis to change score-mapping for this dimension.

Figure 16

Boxplots of Results of Monte Carlo Simulation of Combinations of AI Techniques (Converted Dimensions), split up for enhanced readability



Chapter Summary

This chapter began with simple concave down functions, where zones of interest are identified on the curves—as possible locations to apply AI techniques. I then developed a three-level categorization scheme: trade-off tolerance, problem context, and purpose/efficacy of AI technique. Organizations can work through the categorization or decision trees to better-select AI technique(s) for their intended outcomes. I next explored AI technique colloquial attributes such as parameters and tokens, and suggest a conversion to map from these to my seven dimensions.

I next ran several simulations, as individual and combinations of AI techniques against both colloquial attributes and my seven dimensions. The simulations indicate that the LLM+RAG (0.88 ± 0.05) and RAG+GFN (0.85 ± 0.05) combinations achieve the highest Effectiveness, while LLM+GFN (0.62 ± 0.10) offers the lowest Computational Cost. RAG-based combinations, particularly with GAN, tend to have higher computational demands. LLM+GFN presents an optimal trade-off between Effectiveness and Computational Cost, while LLM+RAG and RAG+MOE demonstrate high overall balance across all three dimensions, including Applicability. Simulations to evaluate AI technique combinations across seven trade-off dimensions yielded: RAG+MOE (0.78 ± 0.05) and LLM+RAG+MOE (0.76 ± 0.05) achieved the highest Accuracy, while adding more techniques balanced effectiveness but did not surpass RAG's baseline. All techniques exhibited negative Explainability, with LLM+RAG performing best (-0.20 ± 0.10) and GAN+MOE worst (-0.40 ± 0.10). In terms of Speed, GAN+RAG performed best (-0.39 ± 0.07), while GAN+MOE scored the lowest (-0.60 ± 0.07). RAG+MOE demonstrated strong multi-dimensional balance, maintaining high Accuracy, moderate Explainability, Fairness, and Reliability, with relatively higher Security and Speed.

Chapter 6. Optimizing an Algorithmic Architecture

Approaches to Optimize Algorithmic Architecture for the Accuracy vs. Speed Trade-off

As we've seen in the literature and my analysis herein, the accuracy vs. speed trade-off is a fundamental challenge when using AI techniques and perhaps where most resources are being expended, and where much value may be obtained if substantive progress is made. Typically, improvements in accuracy require more computation—due to increased model size, more layers, longer processing time—while faster performance may lead to sacrifices in predictive precision. There are, however, several algorithmic/computational and architectural approaches that can be employed to optimize this trade-off: The following is intended to introduce concepts from computer sciences and quickly moves into highly technical areas beyond the scope of this dissertation and therefore, I will not undertake a thorough analysis of any. These approaches are applicable against the accuracy vs. speed trade-off and others, and point to areas of further study in the application of AI techniques to intractable problems classes such as organization strategy.

Model Optimization Techniques

Model Compression. Compressing large models can reduce their size and improve inference speed without significantly sacrificing accuracy. Techniques such as pruning (removing less important neurons), quantization (reducing the precision of weights), and knowledge distillation (training a smaller model to mimic the performance of a larger one) can help reduce the complexity of models without sacrificing performance too much (Bucila et al., 2006; Cheng et al., 2020; Choudhary et al., 2020).

Knowledge Distillation. This technique involves transferring knowledge from a large, highly accurate model (teacher) to a smaller, faster model (student). The smaller model can then achieve similar performance while requiring less computation (Gou et al., 2021; Cho & Hariharan, 2019; Park et al. (2019); Kim & Rush, 2016).

Efficient Hardware Utilization

Parallel Computing. Using parallel computing[60] (on multiple GPUs, TPUs, or other distributed systems architectures can speed up inference time without compromising accuracy (Skillicorn & Talia, 1998). For example, model parallelism can split a large model into smaller parts that run concurrently (Zeng et al, 2021).

Asynchronous or Batch Processing. Grouping multiple requests or data points into a single batch for processing can lead to more efficient use of resources, improving speed without a direct loss in accuracy. This is commonly done in recommendation systems and search engines, where multiple queries can be processed together. Instead of running all operations in sequence, asynchronous processing (such as running different sub-models concurrently) can achieve faster performance, allowing high accuracy models to be invoked only when necessary.

Hardware-Specific Optimizations. Tailoring the model for specific hardware accelerators, e.g. NVIDIA GPUs, Google's TPUs, or even custom silicon, can help reduce

---

[60] https://www.heavy.ai/technical-glossary/parallel-computing, retrieved Jan 26, 2025; https://en.wikipedia.org/wiki/Parallel_computing, retrieved Jan 26, 2025.

latency while maintaining accuracy. Optimized techniques such as TensorRT[61] for NVIDIA hardware or XLA[62] for TensorFlow can optimize model execution.

Approximation Techniques

Approximate Inference. For some use cases, approximate algorithms like Monte Carlo and Markov Chain Monte Carlo methods, or early pruning techniques, can be used to generate fast but reasonably accurate results[63]. These methods may skip detailed computations in cases where full or high precision isn't critical, yielding faster performance—a gain in the speed dimension—with smaller magnitude loss in the accuracy dimension (Rezende et al, 2014).

Sampling. For use in LMs and tasks involving NLP, use sampling and temperature parameters[64] to generate partial responses like partial sentence generation. Instead of generating a complete response, use refinement techniques to incrementally improve the output if more time is available. This technique allows for both speedy responses and high accuracy when needed, while also gaining *creativity* via randomness[65].

Reinforcement Learning for Adaptive Systems

Reward-Based Optimization. In real-time systems including user-prompted systems, RL can be used to find an optimal trade-off between accuracy and speed, where the system adjusts its behavior dynamically based on user feedback, which can trigger different parameters

---

[61] https://developer.nvidia.com/tensorrt, retrieved Jan 26, 2025.
[62] https://openxla.org/xla, retrieved Jan 26, 2025.
[63] https://www.geeksforgeeks.org/approximate-inference-in-bayesian-networks/, retrieved Jan 26, 2025.
[64] https://huyenchip.com/2024/01/16/sampling.html, retrieved Jan 26, 2025.
[65] https://medium.com/@shashankag14/understanding-sampling-techniques-in-large-language-models-llms-dfc28b93f518, retrieved Jan 26, 2025.

on/off as necessary to optimize against the output being sought, Another approach could be to optimize against a predefined cost-benefit analysis or such a parameters, e.g. minimizing response time while maximizing user satisfaction or output accuracy (Aydın et al., 2001; Song et al., 2017).

Table 22

Summary of Approaches to Optimize Algorithmic Architecture for the Accuracy vs. Speed Trade-off

Approach	Description
Model Compression	Reduces model size and speeds up inference.
Knowledge Distillation	Transfers knowledge from large to small models.
Hardware Optimization	Leverage specialized hardware to improve speed. Batch or parallelize tasks to speed up processing.
Approximations and Sampling	Use approximate methods to speed up computations.
Reinforcement Learning	Use reinforcement learning to adapt in real-time.

By combining these approaches (summarized in Table 22 above), organizations can design and implement systems of AI techniques that find an optimal balance for their purposes and use-cases, for the accuracy vs. speed trade-off. Any super-optimal approach depends on the specific constraints, such as real-time performance requirements, the nature of the task, and available computational or other resources.

Architecture of AI Techniques within an Organization Strategy Development Process

An optimal approach for applying advanced AI techniques such as LMs, GANs, RAG, GFNs, and others, within an organizational strategy development process is best accomplished via an algorithmic architecture. Such an architecture should consist of different layers, each optimized to perform a set of functions within a strategy development process. The design of the system is proposed as a multi-tier system that takes data in, generates insights, and iteratively refines its outputs. This proposed architecture and its components are summarized in Table 23 and visually presented Figure 17 (both below).

While this is a proposed algorithmic architecture, and tries to address the shortcomings and trade-offs highlighted throughout this paper, the end result builds upon the simulated performance measures above of individual and combinations of AI techniques, and my proposed conversions and weightings from colloquial attributes to my seven trade-off dimensions.

Data Collection Layer

At the foundation, data is gathered from multiple internal and external sources. Internal data sources include financials, market performance, inter and intra-department data, customer feedback, sales and analytics, and so on. External data includes market trends, competitor analysis, news and events, regulatory and legal requirements, and social media.

Various technologies and tools can be employed to gather this data, such as web scraping for external sources and robust data integration systems with single sources of record/truth for internal data sources.

Preprocessing and Embedding Layer

The next payer in this multi-tier architecture acts to clean, process, and transform the data from its respective sources to format(s) useful for various AI techniques, such as tokenization n-gram data conversions for LMs. Pre-processing of text and numerical data must take place. For the former, this is accomplished via tokenization, stemming and vectorization (insertion and storage into vector databases). Numerical data is transformed via scaling, normalization and feature extraction, and various other methods.

Technologies for textual processing are largely based on LMs such as GPTs or other similar techniques to embed the textual data into useable formats and constructs. Data transformation algorithms are more varied and perhaps less black-box and include autoencoders for dimensionality reduction or feature engineering, again designed to embed the numerical data into useable numerical constructs.

Insight Generation or Solutioning Layer

This is where the proverbial magic happens, and several AI techniques are architected into a *system*, to operate in series and parallel, to generate insights to assist in strategy formulation and strategic decision-making processes, based on the processed and embedded data from the previous tier.

Sentiment Analysis and Stakeholder Insights. LMs analyze unstructured data like social media, surveys, and internal reports to assess stakeholder sentiment. This is augmented using GANs, to simulate realistic market scenarios by generating synthetic data for rare events or strategic disruptions.

External Market and Trend Information Analysis. RAG is used to retrieve relevant information, documents or reports from internal and external sources, then generate concise, retrieval-supported incrementally accurate insights on market conditions and emerging trends. This allows for incrementally accurate, real-time insights for market forecasting and other externally-sourced information. GFNs can be used for simulating complex decision processes and evaluating potential strategic option sets over time, accounting for multiple possible outcomes and helping to identify the most promising paths forward.

Scenario Planning and Simulation. RL models allow for simulating different strategic choices, from within the option sets, and help predict how they would evolve over time, e.g. strategic decision(s) in furtherance of market expansion. Next, Monte Carlo Simulations and other *agentic modelling*, perhaps powered by GFNs, can simulate numerous strategic scenarios in parallel to help the organization choose the optimal decision(s).

Decision Support & Strategy Refinement Layer

In this layer, AI techniques assists in the generation of specific strategic recommendations, evaluate trade-offs, and *collaborative* decision-making (more on this later).

GFNs help to simulate and rank potential strategic decisions based on various performance metrics, such as ROI, long-term sustainability, and risk. LMs help to generate initial draft of strategic choices, providing text-based reports, executive summaries, or even more comprehensive outputs resembling a strategy. As we've seen with other approaches before the advent of and widespread access to generative techniques, an human-machine collaboration and iterative looping now plays a role. I propose a process whereby the AI techniques present a range of options, insights, and feedback that are presented to human decision-maker(s), who make the final strategic call. This may be thought of as the ultimate RAG and allows for further mitigation of unintended consequences that may arise due to technical deficiencies of the underlying or upstream techniques employed to get to this stage.

Monitoring and Adjustment Layer

Once a set of strategic decisions are implemented, AI techniques can continuously track performance and make recommendations for adjustments.

Continuous Feedback. Real-time monitoring systems powered by LMs should be deployed to analyze incoming data streams. Depending on the scope of the decisions being implemented, these data streams may be, e.g. sales data, performance metrics, market share shifts. Once analyzed, LMs can generate actionable insights for possible areas of adjustment. Any suggestions can iterate through the human-decider mechanism suggested above. Other AI techniques such as RL and GFNs and continually adapt the strategic plan or execution sequence based on real-time feedback, in an attempt to adjust actions to optimize long-term outcomes. RL models are adept at dynamic adaptation, with deep-learning models to predict trends and suggest refinements using time-series forecasting.

Flow of Architectural Components Working Together

Raw data is injected from internal and external data sources. This data is cleaned, tokenized and transformed into embeddings for consumption by ML models and AI techniques. Next come the critical and value-add techniques to generate useable insights. LMs generate initial insights, analyses, and summaries. GANs generate synthetic, realistic data to augment insights. RAG helps refine answers to strategic questions by retrieving relevant data and documents. And, GFNs simulate multiple strategic decisions over time and provide probabilistic outcomes.

Decision support can be achieved in a number of ways—a simple approach may be to have AI techniques generate the solution set, rank them based on potential outcomes, and assist in collaborative human-machine decision-making. Outcomes are continuously monitored and actions optimized against desired targets.

Table 23

Use of AI Techniques Across the System Architecture

AI Technique	**Function Served**
LMs	Generating insights, processing unstructured data, and summarizing vast amounts of information.
GANs	Simulate market events, generate synthetic competitors, or create diverse future scenarios to inform decision-making.
RAG	Optimize retrieval of knowledge from internal and external databases, incrementing and improving system-generation with the current relevant information.
GFNs	Model probabilistic strategic decisions, helping the system explore a variety of strategic options and providing insight into their potential future outcomes.
RL	Dynamically adjusts the decisions as real-world feedback comes in, helping optimize the decision-path *and outcomes* over time.

Figure 17

Representation of Proposed Algorithmic Architecture of AI Techniques

```
                        +------------------------------+
                        |     Data Collection Layer    |
                        +------------------------------+
                                       |
                                       v
                        +------------------------------+
                        | Preprocessing and Embedding  |
                        +------------------------------+
                                       |
                                       v
     +--------------------+--------------------+---------------------+
     |                    |                    |                     |
     v                    v                    v                     v
+-----------------+   +----------------+   +------------------+   +-------------------+
| Sentiment &     |   | Market Trends  |   | Scenario         |   | Decision Support  |
| Stakeholder     |   | & Forecasting  |   | Simulation &     |   | & Strategy        |
| Insights (LLM)  |   | (RAG, GANs)    |   | Analysis (RL,    |   | Refinement (LLM,  |
|                 |   |                |   | GFlowNets)       |   | GFlowNets)        |
+-----------------+   +----------------+   +------------------+   +-------------------+
     |                    |                    |                     |
     v                    v                    v                     v
    +-----------------------------------------------------------------------+
    |         Strategy Monitoring & Continuous Adjustment (RL, LLM)         |
    +-----------------------------------------------------------------------+
```

Chapter Summary

I revisit the definition of AI, and argue that AI should be referred to as *AI Techniques*, rather than solely as systems (OECD, 2024) or some other term used as a noun. This also allows for their fluid and combinable nature. AI techniques can consist of various computational architectures, including LLMs and RAGs.

A key challenge in the application of AI techniques is the accuracy vs. speed trade-off, as improvements in accuracy often require more computation. Several optimization techniques, such as model compression, knowledge distillation, parallel computing, and reinforcement learning, can help achieve this trade-off. Organizations can strategically combine these methods to optimize AI systems based on specific performance constraints.

A multi-tier algorithmic architecture is proposed, for integrating AI techniques into an organization's strategy development process. The architecture consists of five layers: 1) a data collection layer gathers internal and external data, 2) a preprocessing and embedding layer cleans and transforms data into structured formats for AI technique consumption, 3) an insight generation layer employs AI techniques such as LMs, GANs, RAG, and GFNs for various segmented tasks, 4) a decision support and refinement layer uses AI techniques to rank strategic options, provide decision-making support, and generate strategic reports, ensuring human-AI collaboration for final decisions, and finally, 5) a monitoring and adjustment layer that implements real-time monitoring using AI techniques like RL and LMs to track outcomes, refine strategies dynamically, and optimize long-term results. These layers work together by processing raw data, generating insights, simulating decisions, and continuously refining strategies through human-AI collaboration.

Chapter 7. Conclusion

This study began as an exploration of the abstract topic of organization strategy and how AI technologies and techniques have evolved and been applied against such intractable problem classes to try and improve outcomes. It was grounded by technologies and approaches published in an extensive literary record, or used by practitioners. I intended to advance understanding of the nature of the problem space, why outcomes until now have been less than expected, and develop new management models and approaches to address process or technical shortcomings.

In this concluding chapter, I will summarize the study and highlight the contributions of this dissertation, including to the literature and theory, and practice of strategic decision-making. I will then discuss limitations and implications for future research. I close with some final words and commentary on artificial intelligence vis-à-vis human intelligence, and my journey.

Summary of Findings

This dissertation explored improving the intractable problem of strategic decision-making, via the application of AI techniques. While the latest generative AI techniques hold great promise in many areas, they are thus far unable to solely-computationally address the complexity of strategic decision-making. I have developed a set of seven (7) trade-off dimensions: accuracy, explainability, fairness, privacy, reliability, security, and speed; and there are twenty-one (21) pairwise trade-offs between these dimensions. Despite gains made with respect to trade-off dimensions such as accuracy or *post-training* speed, these seven trade-off dimensions still persist. The trade-offs have not been addressed, but rather, abstracted away.

As ways to make improvements, several management models and approaches were introduced and developed, including using the NK Model (Kauffman & Levin, 1987) as a unifying model to more comprehensively address strategic problems and achieve enhanced organization fit (Porter, 1996). This presents the opportunity to apply maximization AI techniques, such as evolutionary algorithms, at important locations on an organization's production function.

Indicative simulations of individual and combinations of AI techniques against colloquial and converted attributes, yielded interesting results. The techniques of RAG and MoE outperformed other techniques across the broadest subset of the seven trade-off dimensions. LLMs offer broad applicability and reasonably high accuracy. The loss of speed due to pre-training (and other resources such as CPU and memory also captured in the speed dimension), appears to suggest that more specialized architectures such as MoE as more optimal.

Overall, based on the AI techniques I examined herein, an optimal management model and approach to apply AI techniques to organization strategic decision-making is achievable via a hybrid multi-layer human-machine architecture with feedback loops. This allows for the maximization:minimization of dimensions by leveraging AI techniques selected and placed at specific points in the strategic process, and using human judgment and other *computation-exceeding* capabilities to mitigate areas where AI techniques fall short due to detrimental trade-off dimensions.

Discussion of Findings

The findings of this research study systematically identify trade-off dimensions and the trade-offs between them, and a proposed structured hybrid human-AI approach for organizations to use AI techniques as tools in their strategic decision-making processes.

On the trade-offs, the identification of seven persistent trade-off dimensions is crucial, as it highlights the inherent limitations of AI in strategic decision-making. My research suggests that AI technique advancements have not eliminated these trade-offs but have only reframed them, which aligns with broader discussions on issues like AI ethics, fairness, and interpretability.

The findings also indicate that specialized architectures like MoE outperform general LLMs in many scenarios, particularly where resource efficiency and speed are critical. This suggests a shift toward more *modular* AI techniques and/or model-types, rather than monolithic LLMs for strategic decision-making.

The proposed multi-layer hybrid Human-AI approach with feedback loops, acknowledges the limitations of AI techniques and leverages human expertise to mitigate specific trade-off dimensions and other technical shortcomings. This aligns with real-world AI deployments in high-stakes decision-making, such as in healthcare and finance, where human oversight currently remains essential. This study underscores the necessity of combining artificial intelligence with human intelligence rather than seeking a purely computational solution for strategic decision-making. It provides a structured framework that could guide future AI implementations in various contexts.

Contributions to the Literature and to Theory

As I embarked on this study, it quickly became obvious how expansive and significant this topic is and how the volume of literature was simultaneously growing and specializing. While exploration and understanding were growing in certain areas, there was lack of convergence or unifying research to address challenges such as under-achievement of objectives, deliberate or inadvertent non-inclusions, or trade-offs, all of which persist until today. AI technique deployment into strategic decision-making is accelerating yet under-theorized.

I have developed and proposed several management models, from a thorough multi-step approach to more comprehensively examine the strategic problems and solution space, amongst other constructs. Without simplification—in fact the opposite—this transforms often abstract and intractable problems into computationally improvable form. I posit that this directly challenges the common tendency and practice of simplifying problems and solution spaces—from NP-class to P-class as I discuss at length above. Beyond language flexons which no doubt assist (Moldoveanu & Leclerc, 2015), I've put forward mathematics and computation science based approaches. I note that foundational works of Porter (1996) on critical elements of superior strategy, and the NK Model of evolutionary out-performance by Kaufman and Levin (1987), and others, were not connected with more recent study of the application of AI techniques against these same classes of problems—superior strategic decision-making or other intra-organization competitive performance. Finally, my quantitative analysis of AI techniques as individual techniques and in combinations, and against colloquial attributes and my set of seven trade-off dimensions, counters the far too frequent reliance by organizations and practitioners/developers on LMs. There are many reasons for this, including lack of awareness and understanding of the specific targeting of AI technique against problem type or step in strategic decision-making process. I believe my analysis supports exploration and expansion of AI technique selection.

This study makes several contributions to the literature and theory: 1) explication and extensive exploration of a set of seven (7) trade-off dimensions and twenty-one (21) pairwise trade-offs between them that have relevance within both organization strategy and computational optimizations, 2) finding that these trade-off dimensions persist despite advancements in and performance of AI techniques, 3) use of Kauffman's NK Model (1987) as a basis to propose a unifying model the connects Porter's strategic fit (1998) *with* computational resources used to optimize strategic decision-making, and 4) proposition that *AI Technique* is the most accurate way to refer to anything AI, e.g. as opposed to AI as a *nebulous* noun.

Contributions to Practice

The nature and commercial implications of the topics I studied, point to practitioners as key drivers of advancement in both theory and application (Abreu & Grinevich, 2013; Domingos, 2015; Kinnunen et al., 2018). The vast commercial resources applied to advance the technologies and techniques to their current state, and for activities such as model training have further blurred the lines between theoretical research and practitioners.

Therefore, contributions I make to the practice to some degree also contribute to theory being studied and advanced by academics. In this context, contributions I made are: 1) a categorization of trade-offs by industry and functional area, common ways that strategists and consultants delineate their expertise and problem domains, 2) an approach to achieve optimized outcomes by more comprehensively defining the problem, via a multi-step approach to fully capture problem variables, with identification of points of interest on a production function, 3) finding that different AI techniques are better suited for different problems/goals, 4) a multi-level categorization that organizations can use to select AI technique(s) for their specific problem/activity, 5) a conversion from widely used colloquial attributes to my set of seven trade-off dimensions, and 6) management models and approaches to improving outcomes from/with AI techniques generally, and in the context of commonly used multi-step organization strategy processes.

Limitations of the Study and Implications for Future Research

This research study relied on and was informed by existing published literature and other publicly accessible source information. Also, I did not undertake direct empirical data gathering on individual or representative subjects. Another limitation is that much research and development is shrouded in corporate secrecy or otherwise inaccessible in the public domain.

Future Direction 1. Expand Exploration of Trade-off Dimensions and Interactions

My review of the literature suggests that the majority of the trade-offs are not well explored or understood. This lack of extensive study and publication presents a challenge but also an opportunity for subsequent research. I comment several times about the inequity of literary volume into certain trade-offs, including within an industry or functional context. Deeper and wider coverage into less-explored areas should be addressed, especially as the application domains of AI techniques expands (see Tables 7-9 for obvious visible gaps in the literature I found and reviewed).

Future Direction 2. Explore Multi-Dimensional Manifold

I have extensively explored each of the trade-off dimensions, and pairwise 1-to-1 trade-offs between them. However, I did not explore any multi-dimensional analysis where three or more dimensions and the interactions between them are examined. I have so far found no academic or other literature examining this, and this presents an opportunity to broaden understanding of more complex interaction between trade-off dimensions.

Future Direction 3. Explore Formulation of AI Techniques as Combinations

Based on my models of combinatorial inclusion of AI techniques, technical researchers and practitioners may want to explore new combinations of underlying AI technologies that can be composed and deployed as novel *AI technique sets*, that may yield interesting results in general and against specific problem classes or domains. This is also interesting in view of research directions into multi-dimension trade-off analysis.

Future Direction 4. Other Lines of Inquiry based on Findings

How can organizations practically implement his hybrid model like the one proposed herein?

What tools and frameworks best support the adaptive deployment of AI techniques at different strategic decision-making layers?

How do regulatory and ethical considerations shape the adoption of AI-based or driven decision-making?

Future Direction 5. Lingering Questions

As this research study draws to a conclusion, I wanted to address some questions that may still linger. In addition to the inherent value of such a succinct Q&A matrix in Table 24, it also suggests areas of attack for researchers, technologists and developers, as these are the thresholds to try and breach as these technologies and techniques continue to evolve, and are deployed in ever-expanding areas.

Table 24

Some Questions that Remain, with My Answers

Question	**Answer**
Are AI Techniques able to add value?	Yes—as automation tools for simple tasks and as human-aids for limited sets of more complex tasks.
Are AI Techniques able to solve intractable problems?	Not yet in a way that can be relied upon.
Are AI Techniques able to formulate strategy?	Not yet in way that can be relied upon.
Can AI techniques achieve judgment?	Not yet in way that can be relied upon.
Can AI techniques exhibit reasoning?	Yes, to limited degrees.
Will reasoning improve?	Yes, but it’s unclear exactly how or how far.

The Future of Work, Ethical and Other Strategic Considerations

There are many issues and considerations beyond technical limitations, which can be contemplated as AI techniques continue to evolve, and the lines blur between human and computation-based decision making. While important on their own, I mention them because how these are treated, perceived and prioritized will influence investments and societal push and acceptance of greater AI technology adoption, which will impact how any such systems will be designed and implemented.

Meek et al. (2016) present a review of the ethical and risk implications of AI, and how these can be better managed. Of interest to this paper is the consideration and acknowledgement that AI-enabled systems may “become capable of self-determination and super intelligence in only a few decades” (p. 682). This further highlights some of the cautionary commentary in Meek at al. (2016). They call out the important distinction of AI which is domain-specific from artificial general intelligence. This latter category are those systems and technologies that can analyze and respond to unknown or unspecified data and situations. These forms of AI can learn, create knowledge, make decisions—truly simulate human brain function (p. 683). While the implications and gaps Meek et al. put forward are important today and will become more important as the technology advances, I’m more interested in (or concerned about) the specific and far-reaching commentary on where AI is in terms of traversing into the realm of human brain function, and where it is already envisioned to be going—an increasingly blurry line if one still exists (Brynjolfsson & McAfee, 2014; Domingos, 2015). Indeed, it is contemplated that AI systems may be able to completely mimic or perform like human brains within a few decades. These will be important considerations as we look to using AI for general strategy formulation and the potential for organizations and employees to follow directions and plans created by machines. As noted above, these ideas of the capabilities and limitations of AI in context of human cognition or thinking, and true general intelligence, are not new and were seen as early as Meinhart (1966) and McCarthy and Hayes (1969) and by many others since then.

The analysis contained in Acemoglu and Restrepo (2018) on the impact of automation on labour, uses models including economics and efficient use of labour capital and real costs, eventually leading to an equilibrium. While there is rampant fear and ad hoc perception that ever-increasing automation and AI will make labour ever more redundant, drive down relative or real wages, the authors move the discussion into the realm of constructing a comprehensive framework to study these effects along with potentially countering forces. Of note is their premise that while automation replaces certain labour tasks, other technologies will complement labour—new tasks will emerge where humans have a comparative advantage. Their use of the term comparative is interesting in that it implies the economic definition of same: humans can and will perform certain tasks more efficiently than machines and thus will be able to make available their excess production. Similar to Ng (2016), Acemoglu and Restrepo (2018) rely on an underlying assumption that humans have a comparative advantage in new and complex tasks. Humans can learn or apply on a novel subject almost instantly whereas AI must first learn from enormous amounts of data. Based on this supposition, Acemoglu and Restrepo (2018) claim that humans can maintain their long run labour share due their advantage being significant and the continuous creation of new tasks (p. 1489). Their analysis is highly technical, like Agrawal et al. (2018b), ultimately algebraically and logically demonstrating their thesis. It is beyond the scope of this paper to perform a detailed analysis of this position, but it’s worth considering and thinking about as we think about the ability of machines to perform the complex task of strategy formulation and models under which this would produce economic advantage. A question which time may help answer, is how far automation can go and how far must humans *come up* with new complex tasks where they seem to have a comparative advantage?

Continuing with the theme of the impact of the future of work and the impact of technology and computerization, Ojanpera et al. (2018) suggest that the future is far from homogeneous in any one direction and that multiple future scenarios exist and several factors will play a role, including “context, choices and adaptability” (p. 1). Their general findings are similar to what we’ve previously seen (Acemoglu & Restrepo, 2018; Agrawal et al., 2018b) in that there is data to support a premise and assumption of new opportunities arising out of automation and technological advancements. Like others, Ojanpera et al. (2018) delineate routine or less cognitive tasks from complex tasks as a possible dividing line on how far AI can go—while the former is more prone to replacement, the latter are less so and there are also new roles which are being created as a result of automation.

Final Words

The title of this dissertation asserts that AI may soon emulate Picasso and Tolkien. Is it “soon” or is it already here? I posit that, despite advancements made via generative AI techniques, true emulation remains some distance away. More accurate words to use with respect to generative technologies and approaches, would be: copy, mimic, or parrot. In all cases, there remains a necessity to start with something that can be used to train the models—be that words or images or any other binary-encodable object. Once trained, advanced models can perform all sorts of amazing and wonderful things. Entire videos and movies can be created. Entire musical works can be created. Entire textual compositions can be created. The argument that people like Picasso and Tolkien also started with knowledge of words, and string words together in celebrated ways, is flawed. There is fundamental difference between the intelligence and the creations exhibited by such artists, and the mechanisms utilized by generative models like GPTs to string together words. The most important of these is their current inability *to know*, and whether their outputs/responses are in fact accurate. Improvements are no doubt being made all the time to existing AI techniques or development of new ones, but my research suggests that gaps remain.

What is Human Intelligence?

While a deep exploration on this subject is far outside the scope of this study, I believe some thoughts and reflections are warranted on the subject of human intelligence. As we've seen, LLMs, or LMs more generally, are a critical foundation and driving force behind much of the recent advancements. These techniques allow for sense-making sequences of words, pixels, and so on. In effect, LMs have simulated/emulated some communication and generative capabilities of brain-based approaches. Other techniques have also been inserted to augment the output of various generative techniques with some "grounding", "information", "knowledge", etc., via RAG or some other AI technique, or other human-machine technique, and voila, we have something akin to intelligence? Do we really?

When a human child learns, they learn different forms of language, e.g. physical, verbal, etc., and through additional of forms of reinforcement learning, their inherent retrieval augmentation, and brain-based adversarial generations (thinking of and mentally simulation alternatives), intelligence is born and develops.

We've seen that with the leap gained via LMs, other building blocks of human brain-based intelligence systems, can start to be developed in microchip-based intelligence systems. For example, GFNs are one approach to model reinforcement learning. MoE are an approach to create specialization within microchip-based intelligence, thereby managing and reducing the amount of data or information an *expert* is trained on and is responsible for. Does this sound familiar?

This prompts several questions, and not to minimize the achieved technical feats, but is that all *we* are, and human brains are? Can human brain-based intelligence really be emulated by creating and connecting the similar building blocks within an algorithmic architecture? Despite the advancements, this study suggests that human intelligence seems to be more than a computational sum of the parts, as are complex processes such as strategic decision-making. This is supported by the nature of model training and persistence of speed and other trade-offs, and other limitations. Also, the ability of the human brain to attack intractable problems is unmatched in design and performance in microchip-based approaches. There are glimpses of potential, and as I've proposed, hybrid models where capabilities can be collectively harnessed to try and improve performance against certain dimensions while better-mitigating corresponding trade-offs. There may come a time when the lines are more blurred, or at least more blurred in the context of certain problems, functions and applications. In other words, some tasks may be turned over to AI techniques for demonstrably reliable and valid outcomes.

What's interesting is that current research and practitioner approaches, and associated algorithmic architectures, seem to be moving towards emulating human brain and indeed, societal structures. The MoE technique built upon LMs is a form of specialization that we as humans do when gaining expertise in area(s) of study or career. MoE have similar specialization in their algorithmic architecture being specialized against domains, forms of input data and expected outputs, e.g. different languages (human or computer programming) vs. computation or mathematics vs. sounds or visualizations vs. other specializations. The resulting architecture is a form of load-balanced routing of information to a shared pool of quasi-isolated experts. Outputs are then tested and improved via RL. The challenges posed by a single *omni* artificial intelligence appear unsurmountable for myriad of reasons, the most obvious being time and cost—captured in my Speed trade-off dimension. And, approaches and techniques seem to be converging with what works with humans and human brains. It remains to be seen how far these approaches will advance, and what results they will yield.

Problem Frontiers

I suggest that the problem frontier is not training (or pre-training). Human intelligence seems to develop at a faster rate on less data than the corpi being used to train current models with current AI techniques. More pre-training requires more compute. Therefore, there are trade-offs that exist and persist.

Is this a compute problem? More pre-training, or more advanced post-training algorithmic “inference” requires more compute. Again, there are trade-offs that exist and persist. To advance the science and math, without detrimental effects of the trade-offs, focus should be less about compute and more about combinations of techniques in layers and sequences, thereby allowing for computational gains within each algorithmic architectural element.

Greater progress with less detrimental trade-offs may be achieved with more appropriate selection of technique(s), more appropriate selection of management models and approaches, and may involve iterative and hybrid machine-human approaches. And there are and may forever be, certain classes of organization problems and goals, that are not best-suited to be addressed via AI techniques. There are also certain classes of problem/goals that are and will continue to benefit from these techniques.

Architecture

Architecture is a construct of hardware and systems, but also algorithmically. In the context of AI techniques, architecture is an algorithmic construct—this means that differences in how different MLAs or other elements are pieced/chained together leads to different outcomes and quality of outcomes.

In the context of language models, this can be dense, sparse (MoE), or hybrid. Without getting into technical details, the density of transformers in the architecture is a critical driver of computational and algorithmic complexity and costs[66]. Of these three established architectural patterns, MoE is the least dense and most advantageous in terms of gains in accuracy efficacy and computational accuracy for the last detrimental trade-off, e.g. speed (i.e. compute costs). Some examples of MoE architecture-based LMs are ChatGPT-o1, ChatGPT-4o, DeepSeek[67] (Dai et al., 2024).

The current answer lies at the intersection of AI techniques, and combinations thereof. I suggest, and invite others to explore such an *algorithmic architecture*—a purpose-built optimizations built for: 1) data collection, 2) preprocessing and embedding, or 3) continuous improvement layers.

[66] https://wandb.ai/zaiinn440/hybridMoe/reports/MoE-vs-Dense-vs-Hybrid-LLM-architectures--Vmlldzo3NzYwNzAw, retrieved Jan 26, 2025.
[67] https://github.com/deepseek-ai/DeepSeek-MoE, retrieved Jan 26, 2025.

References

Abreu, M., & Grinevich, V. (2013). The nature of academic entrepreneurship in the UK: Widening the focus on entrepreneurial activities. *Research Policy*, *42*, 408–422. https://doi.org/10.1016/j.respol.2012.10.005.

Abowd, J. M., & Schmutte, I. M. (2018). An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices. Forthcoming in *American Economic Review.*

Abowd, J. M., & Schmutte, I. M. (2019). An economic analysis of privacy protection and statistical accuracy as social choices. *American Economic Review*, *109*(1), 171-202.

Acemoglu, D., & Restrepo, P. (2018). The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment. *American Economic Review,* 108(6), 1488-1542. https://doi.org/10.2139/ssrn.2781320.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access*, *6*, 52138-52160.

Agrawal, A., Gans, J., & Goldfarb, A. (2017). What to expect from artificial intelligence. *MIT Sloan Management Review*, 58(3), 23–26.

Agrawal, A., Gans, J., & Goldfarb, A. (2018a). A Simple Tool to Start Making Decisions with the Help of AI. *Harvard Business Review Digital Articles*, 2–6.

Agrawal, A. K., Gans, J. S., & Goldfarb, A. (2018b). Exploring the Impact of Artificial Intelligence: Prediction Versus Judgment. Working Paper Series. https://doi.org/10.3386/w24626.

REFERENCES

Agrawal, S., & Haritsa, J. R. (2005, April). A framework for high-accuracy privacy-preserving mining. In *21st International Conference on Data Engineering (ICDE'05)* (pp. 193-204). IEEE.

Agrawal, A. (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence*. Lecture presented at Deep Learning and Reinforcement Learning Summer School in Rotman School of Management, University of Toronto, Toronto.

Aguiar, R. (2019). An overview of model explainability in modern machine learning. Retrieved Aug 9, 2020, from https://towardsdatascience.com/an-overview-of-model-explainability-in-modern-machine-learning-fc0f22c8c29a.

Ai.gov (2023). 2023 Consolidated AI Use Case Inventory (PUBLIC).csv. https://ai.gov/ai-use-cases/, retrieved Nov 12, 2024.

Albescu, F., Paraschiv, D., & Pugna, I. (2008). Business Intelligence & Knowledge Management - Technological Support for Strategic Management in the Knowledge Based Economy. *Informatică Economică*, 4(5).

Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, *20*(1), 1-9.

Aristodemou, L., & Tietze, F. (2018). The state-of-the-art on Intellectual Property Analytics (IPA): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data. *World Patent Information*, 55, 37–51. https://doi.org/10.1016/j.wpi.2018.07.002.

REFERENCES

Arrieta, A.B., Rodríguez, N.D., Ser, J.D., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2019). Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion, 58*, 82-115.

Aydin, H., Melhem, R., Mosse, D., & Mejia-Alvarez, P. (2001). Optimal reward-based scheduling for periodic real-time tasks. *IEEE Transactions on Computers, Computers, IEEE Transactions on, IEEE Trans. Comput*, *50*(2), 111–130. https://0-doi-org.aupac.lib.athabascau.ca/10.1109/12.908988.

Babcock, J., Kramár, J., & Yampolskiy, R. V. (2019). Guidelines for artificial intelligence containment. *Next-Generation Ethics: Engineering a Better Society (Ed.) Ali. E. Abbas*, 90-112.

Bader, V., & Kaiser, S. (2019). Algorithmic decision-making? The user interface and its role for human involvement in decisions supported by artificial intelligence. *Organization*, *26*(5), 655-672.

Baker, T., Gill, J., & Solovay, R. (1975). Relativizations of the P=?NP question. SIAM Journal on computing, 4(4), 431-442.

Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., & Horvitz, E. (2019, October). Beyond accuracy: The role of mental models in human-AI team performance. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* (Vol. 7, pp. 2-11).

REFERENCES

Baum, S. D., Goertzel, B., & Goertzel, T. G. (2011). How long until human-level AI? Results from an expert assessment. *Technological Forecasting and Social Change*, 78(1), 185–195.

Bektaş, M., Pereira, J. K., Daams, F., & Van der Peet, D. L. (2024). ChatGPT in surgery: A revolutionary innovation? *Surgery Today*, *54*(8), 964-71. https://doi.org/10.1007/s00595-024-02800-6.

Bellamy, R.K., Dey, K., Hind, M., Hoffman, S.C., Houde, S., Kannan, K., Lohia, P.K., Martino, J., Mehta, S., Mojsilovic, A., Nagar, S., Ramamurthy, K.N., Richards, J.T., Saha, D., Sattigeri, P., Singh, M., Varshney, K.R., & Zhang, Y. (2018). AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias. *IBM Journal of Research and Development*, *63*(4/5), 4-1. *ArXiv, abs/1810.01943*.

Bettis, R. A. (2017). Organizationally intractable decision problems and the intellectual virtues of heuristics. *Journal of Management*, 43, 2620-2637. https://doi.org/10.1177/0149206316679253.

Bettis, R. A., & Hu, S. (2018). Bounded Rationality, Heuristics, Computational Complexity, and Artificial Intelligence. *Advances in Strategic Management*, 39, 139-150. https://doi.org/10.1108/S0742-332220180000039010.

Bhasin, S. (2019a) Can Artificial Intelligence Formulate Strategy? A Literature Review. Unpublished manuscript.

Bhasin, S. (2019b). A Conjoint Analysis Design to Elicit Preferences of AI Technologies and Implementation Characteristics to Improve Organizational Decision Making. Unpublished manuscript.

Bhasin, S. (2020a). DDBA 804 – Field Interviews Notes and Transcripts. Unpublished manuscript.

Bhasin, S. (2020b). DDBA 804 – Field Observations Notes. Unpublished manuscript.

Bhasin, S. (2020c). DDBA 804 – Field Report. Unpublished manuscript.

Bhasin, S. (2020d). DDBA 804 – Qualitative Study Research Proposal. Unpublished manuscript.

Bhasin, S. (2020e). Can Artificial Intelligence Formulate Strategy? A Literature Review. Unpublished manuscript.

Bhasin, S. (2022). Dissertation Proposal: AI May Soon Emulate Picasso and Tolkien—Is Porter Next? An Examination of AI Technologies Applied to Organization Strategy. Unpublished manuscript.

Bhatt, U., Xiang, A., Sharma, S., Weller, A., Taly, A., Jia, Y., Ghosh, J., Puri, R., Moura, J.M., & Eckersley, P. (2019). Explainable machine learning in deployment. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (*pp. 648-657*)*.

Biggiero, L. (Ed.). (2016, January). NK Simulation Modeling. In *Relational Methodologies and Epistemology in Economics and Management Sciences* https://doi.org/10.4018/978-1-4666-9770-6.ch002.

Boote, D. N., & Beile, P. (2005). Scholars Before Researchers: On the Centrality of the Dissertation Literature Review in Research Preparation. *Educational Researcher*, (6), 3.

Bjork, R. C. (2016). The master algorithm: how the quest for the ultimate learning machine will remake our world. *Perspectives on Science and Christian Faith*, 68(3), 214–215.

Blockchainized Internet of Minds: A New Opportunity for Cyber–Physical–Social Systems. (2018). *IEEE Transactions on Computational Social Systems*, 5(4), 897. https://doi.org/10.1109/TCSS.2018.2881344.

Bose, I., & Mahapatra, R. K. (2001). Business data mining - a machine learning perspective. *Information & Management*, (3), 211. https://doi.org/10.1016/S0378-7206(01)00091-X.

Brezillon, P. (1999, January 1). Context in Artificial Intelligence: I. A Survey of the Literature. *COMPUTERS AND ARTIFICIAL INTELLIGENCE*, *18*(4), 321–340.

Brock, J. K. U., & Von Wangenheim, F. (2019). Demystifying AI: What digital transformation leaders can teach you about realistic artificial intelligence. *California Management Review*, *61*(4), 110-134.

Bronshtein, I. N., Semendyayev, K. A, Musiol, G., & Muehlig, H.. (2007). *Handbook of mathematics*. Springer Science & Business Media. https://doi.org/10.1007/978-3-540-72122-2

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. New York, NY: W W Norton & Co.

Brynjolfsson, E., & McAfee, A. (2017, July 1). The business of artificial intelligence. *Harvard Business Review*, 1-20.

Bucila, C., Caruana, R., & Niculescu-Mizil, A. (2006). Model Compression. *ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING*, *CONF 12*, 535–541.

Cai, W., Jiang, J., Wang, F., Tang, J., Kim, S., & Huang, J. (2024). A survey on mixture of experts. *arXiv preprint arXiv:2407.06204*, *1*.

Canhoto, A. I., & Clear, F. (2020). Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. *Business Horizons*, *63*(2), 183-193.

Chan, S. L., Ip, W. H., & Cho, V. (2010). A model for predicting customer value from perspectives of product attractiveness and marketing strategy. *Expert Systems With Applications*, 37, 1207–1215. https://doi.org/10.1016/j.eswa.2009.06.030

Chan, S. L., & Ip, W. H. (2011). A dynamic decision support system to predict the value of customer for new product development. *Decision Support Systems*, 52, 178–188. https://doi.org/10.1016/j.dss.2011.07.002.

Cheng, Y., Wang, D., Zhou, P., & Zhang, T. (2017). A survey of model compression and acceleration for deep neural networks. *arXiv preprint arXiv:1710.09282*.

Cho, J. H., & Hariharan, B. (2019). On the efficacy of knowledge distillation. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4794-4802).

Choudhary, T., Mishra, V., Goswami, A. (2020). A comprehensive survey on model compression and acceleration. *Artif Intell Rev* 53, 5113–5155 (2020). https://doi.org/10.1007/s10462-020-09816-7.

Christen, T., Hess, M., Grichnik, D., & Wincent, J. (2022). Value-based pricing in digital platforms: A machine learning approach to signaling beyond core product attributes in cross-platform settings. *Journal of Business Research*, *152*, 82-92.

Choraś, M., Pawlicki, M., Puchalski, D., & Kozik, R. (2020, June). Machine learning–the results are not the only thing that matters! what about security, explainability and fairness?. In *International Conference on Computational Science* (pp. 615-628). Springer, Cham.

Christian, B. (2012). *The Most Human Human: What Artificial Intelligence Teaches Us About Being Alive*. Penguin Random House.

Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., & Malhotra, S. (2018). Notes from the AI frontier: Insights from hundreds of use cases. *McKinsey Global Institute*, 28.

Clancy, K. (2020). Is the Brain a Useful Model for Artificial Intelligence? Retrieved May 31, 2020, from https://www.wired.com/story/brain-model-artificial-intelligence/

Cohen, M. (1999). Commentary on the *Organization Science* Special Issue on Complexity. Organization Science, 10(3), May–June 1999, 373–376.

Collins, J. C., & Porras, J. I. (1994). Built to Last: Successful Habits of Visionary Companies. New York: Harper Business.

Cook, S. A. (1971, May). The complexity of theorem-proving procedures. In *Proceedings of the third annual ACM symposium on Theory of computing* (pp. 151-158).

Cook, S. (2000). The P versus NP problem. Clay Mathematics Institute, 2(6), 3.

Dağlarli (2019) - Explainable Artificial Intelligence (xAI) Approaches and Deep Meta-Learning Models. *Advances and applications in deep learning*, *79*.

Dai, D., Deng, C., Zhao, C., Xu, R., Gao, H., Chen, D., Li, J., Zeng, W., Yu, X., Wu, Y., Xie, Z., Li, Y.K., Huang, P., Luo, F., Ruan, C., Sui, Z., & Liang, W. (2024). DeepSeekMoE: Towards Ultimate Expert Specialization in Mixture-of-Experts Language Models. *Annual Meeting of the Association for Computational Linguistics*. *arXiv preprint arXiv:2401.06066*.

Dase, R. K., & Pawar, D. D. (2010). Application of Artificial Neural Network For Stock Market Predictions: A Review Of Literature. *International Journal of Machine Intelligence*, 2(2), 14-17. https://doi.org/10.9735/0975-2927.2.2.14-17.

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world: Don’t start with moon shots. *Harvard Business Review*, 96(1), 108–116.

Deane, M. (2018). AI and the Future of Privacy. *Towards Data Science*.

Domingos, P. (2015). *The master algorithm: how the quest for the ultimate learning machine will remake our world*. New York: Basic Books.

Došilović, F. K., Brčić, M., & Hlupić, N. (2018, May). Explainable artificial intelligence: A survey. In *2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)* (pp. 0210-0215). IEEE.

Dunjko, V., & Briegel, H. J. (2017). Machine learning\& artificial intelligence in the quantum domain. *arXiv preprint arXiv:1709.02779*.

Eibich, M., Nagpal, S., & Fred-Ojala, A. (2024). ARAGOG: Advanced RAG output grading. arXiv preprint arXiv:2404.01037.

Eisenhardt, K. M. (1999). Strategy as Strategic Decision Making: Successful strategy emerges from a decision process in which executives develop collective intuition, accelerate constructive conflict, maintain decision pacing, and avoid politics. *Sloan Management Review*, (3), 65.

Fabbri, A. R., Kryściński, W., McCann, B., Xiong, C., Socher, R., & Radev, D. (2021). Summeval: Re-evaluating summarization evaluation. *Transactions of the Association for Computational Linguistics*, *9*, 391-409.

Faes, L., Liu, X., Wagner, S.K., Fu, D.J., Balaskas, K., Sim, D.A., Bachmann, L.M., Keane, P.A., & Denniston, A.K. (2020). A Clinician's Guide to Artificial Intelligence: How to Critically Appraise Machine Learning Studies. *Translational Vision Science & Technology,* 9(2), 7-7.

Fazelpour, M. (2016). Developing Unit Cell Design Guidelines for Meso-scale Periodic Cellular Materials. (PhD Dissertation, Clemson University, Graduate School).

REFERENCES

Ferràs-Hernández, X. (2018). The future of management in a world of electronic brains. *Journal of Management Inquiry*, 27(2), 260–263.

Franklin, S. (2007). A Foundational Architecture for Artificial General Intelligence. Frontiers in Artificial Intelligence and Applications, June 2007.

Foulds, J. R., Islam, R., Keya, K. N., & Pan, S. (2020, April). An intersectional definition of fairness. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)* (pp. 1918-1921). IEEE.

Fu, J., Ng S-K., Jiang, Z., Liu P. (2023). GPTScore: Evaluate as You Desire. https://doi.org/10.48550/arXiv.2302.04166.

Gandhi, M. (2020). What exactly is meant by explainability and interpretability of AI? Retrieved Aug 9, 2020, from https://medium.com/analytics-vidhya/what-exactly-is-meant-by-explainability-and-interpretability-of-ai-bcea30ca1e56.

Ganesh Kumar, P. G., Aruldoss Albert Victoire, T., Renukadevi, P., & Devaraj, D. (2012). Design of fuzzy expert system for microarray data classification using a novel. *Genetic Swarm Algorithm. Expert Systems With Applications*, 39, 1811–1821. https://doi.org/10.1016/j.eswa.2011.08.069.

García-Peñalvo, F. J. (2023). La percepción de la Inteligencia Artificial en contextos educativos tras el lanzamiento de ChatGPT: disrupción o pánico. Education in the Knowledge Society (EKS), 24, e31279-e31279.

Gerbert, P. (2018). AI and the 'Augmentation' Fallacy. Retrieved Feb 11, 2019, from https://sloanreview.mit.edu/article/ai-and-the-augmentation-fallacy.

Gerke, S., Babic, B., Evgeniou, T., & Cohen, I. G. (2020). The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. *NPJ digital medicine*, *3*(1), 1-4.

Ghemawat, P., Levinthal, D. A. (2008). Choice Interactions and Business Strategy. *Management Science*. 54 (9), 1638-1651. http://dx.doi.org/10.1287/mnsc.1080.0883.

Goebel, R., Chander, A., Holzinger, K., Lecue, F., Akata, Z., Stumpf, S., Kieseberg, P., & Holzinger, A. (2018). Explainable AI: The new 42? In *International Cross-Domain Conference, Proceedings* (pp. 295-303). Springer Verlag Heidelberg. https://doi.org/10.1007/978-3-319-99740-7_21.

Goertzel, B. (2018, July 22). From Here to Human-Level Artificial General Intelligence in Four (Not All That) Simple Steps. https://singularityhub.com/2018/07/22/from-here-to-human-level-artificial-general-intelligence-in-four-not-all-that-simple-steps.

Goertzel, B. & Pennachin, C. (Eds.) (2007). *Artificial General Intelligence*. Springer Berlin Heidelberg.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C., & Bengio, Y. (2014). Generative Adversarial Nets. *Neural Information Processing Systems*, 27.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020) Generative adversarial networks. Communications of the ACM, 63(11), 139-144.

Goldreich, O. (2010). *P, NP, and NP-Completeness: The basics of computational complexity*. Cambridge University Press.

Gordijn, B., & Have, H. T. (2023). ChatGPT: evolution or revolution?. Medicine, health care and philosophy, 26(1), 1-2.

Gou, J., Yu, B., Maybank, S. J., & Tao, D. (2021). Knowledge distillation: A survey. International Journal of Computer Vision, 129(6), 1789-1819.

Grab, B., Olaru, M., & Gavril, R. M. (2019). The Impact Of Digital Transformation On Strategic Business Management. *Ecoforum Journal*, 8(1), 18th ser.

Grant, C. & Osanloo, A. (2016). Understanding, Selecting, and Integrating a Theoretical Framework in Dissertation Research: Creating the Blueprint for Your “House”. *Administration Issues Journal*, 4(2), 12-26.

Greene, D., Hoffmann, A. L., & Stark, L. (2019). Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning.

Gunning, D. (2017). *Explainable Artificial Intelligence (XAI)* [Slides]. Defense Advanced Research Projects Agency. https://www.darpa.mil/attachments/XAIIndustryDay_Final.pptx.

Gunning, D., & Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI magazine*, *40*(2), 44-58.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI—Explainable artificial intelligence. *Science robotics*, *4*(37), eaay7120.

Guo, W. (2020). Explainable artificial intelligence for 6G: Improving trust between human and machine. *IEEE Communications Magazine*, *58*(6), 39-45.

Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. BenchCouncil transactions on benchmarks, standards and evaluations, 2(4), 100089.

Hacker, P., Krestel, R., Grundmann, S., & Naumann, F. (2020). Explainable AI under contract and tort law: legal incentives and technical challenges. *Artificial Intelligence and Law*, *28*(4), 415-439.

Hambrick, D. C., & Fredrickson, J. W. (2005). Are you sure you have a strategy? *The Academy of Management Executive*, (4), 51. https://doi.org/10.5465/ame.2005.19417907.

Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California management review*, *61*(4), 5-14.

Handelman, G.S., Kok, H.K., Chandra, R.V., Razavi, A.H., Huang, S., Brooks, M., Lee, M., & Asadi, H. (2019). Peering Into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods. *American Journal of Roentgenology*, *212*(1), 38-43.

Hasperué, W. (2015). The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. *Journal of Computer Science & Technology* (JCS&T), 15(2), 157–158.

He, S., Fan, R-Z., Ding, L., Shen, L., Zhou, T., Tao, D. 2023. Merging Experts into One: Improving Computational Efficiency of Mixture of Experts.

He, X. O. (2024). Mixture of a million experts. *arXiv preprint arXiv:2407.04153*. https://doi.org/10.48550/arXiv.2407.04153.

Hinton. G. E. (2016). "On Radiology: Creative Destruction Lab: Machine Learning and the Market for Intelligence", Retrtieved June 17, 2019, from https://www.youtube.com/watch?v=2HMPRXstSvQ.

Hogendorn, C., & Frischmann, B. (2020). Infrastructure and general purpose technologies: a technology flow framework. *European Journal of Law and Economics*, 1-20.

Holzinger, A. (2018, August). From machine learning to explainable AI. In *2018 world symposium on digital intelligence for systems and machines (DISA)* (pp. 55-66). IEEE.

Holzinger, A., Kieseberg, P., Weippl, E., & Tjoa, A. M. (2018, August). Current advances, trends and challenges of machine learning and knowledge extraction: from machine learning to explainable AI. In *International Cross-Domain Conference for Machine Learning and Knowledge Extraction* (pp. 1-8). Springer, Cham.

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(4), e1312.

Hordijk, W., Kauffman, S. A., Stadler, P. F. (2020). Average Fitness Differences on NK Landscapes. *Theory in Biosciences*. 2020(139), 1-7. https://doi.org/10.1007/s12064-019-00296-0.

Horkoff, J. (2019, September). Non-functional requirements for machine learning: Challenges and new directions. In *2019 IEEE 27th international requirements engineering conference (RE)* (pp. 386-391). IEEE.

Huang, M. H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, 21(2), 155–172.

Huang, M. H., Rust, R., & Maksimovic, V. (2019). The feeling economy: Managing in the next generation of artificial intelligence (AI). *California Management Review*, *61*(4), 43-65.

Ip, J. (2025, February 23). LLM Evaluation Metrics: The Ultimate LLM Evaluation Guide. https://www.confident-ai.com/blog/llm-evaluation-metrics-everything-you-need-for-llm-evaluation.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation, 3*(1), 79–87. https://doi.org/10.1162/neco.1991.3.1.79

Jain, G. P., Gurupur, V. P., Schroeder, J. L., & Faulkenberry, E. D. (2014). Artificial intelligence-based student learning evaluation: a concept map-based approach for analyzing a student's understanding of a topic. *IEEE Transactions on Learning Technologies*, *7*(3), 267-279.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Chen, D., Dai, W., Madotto, A., & Fung, P. (2024). Survey of Hallucination in Natural Language Generation. arXiv.org. https://arxiv.org/abs/2202.03629.

Jiang, A.Q., Sablayrolles, A., Roux, A., Mensch, A., Savary, B., Bamford, C., Chaplot, D.S., Casas, D.D., Hanna, E.B., Bressand, F., Lengyel, G., Bour, G., Lample, G., Lavaud, L.R., Saulnier, L., Lachaux, M., Stock, P., Subramanian, S., Yang, S., ... Sayed, W.E. (2024). Mixtral of Experts. *ArXiv, abs/2401.04088*.

Kalla, D., Smith, N., Samaah, F., & Kuraku, S. (2023). Study and analysis of chat GPT and its impact on different fields of study. International journal of innovative science and research technology, 8(3).

Kannan, S., Subbaram, K., Ali, S., & Kannan, H. (2020). The role of artificial intelligence and machine learning techniques: Race for covid-19 vaccine. *Archives of Clinical Infectious Diseases*, *15*(2).

Karp, R. M. (1972). Reducibility among Combinatorial Problems. In: Miller, R. E., Thatcher, J. W., Bohlinger, J. D. (eds) Complexity of Computer Computations. The IBM Research Symposia Series. Springer, Boston, MA. https://doi.org/10.1007/978-1-4684-2001-2_9.

REFERENCES

Kauffman, S. A. (1988). Cambrian explosion and Permian quiescence: implications of rugged fitness landscapes. *Evolutionary Ecology*, 1989(3), 274-281.

Kauffman, S. A. (1993). *The Origins of Order: Self-Organization and Selection in Evolution*. New York, NY: Oxford University Press.

Kauffman, S. A., & Levin, S. (1987). Toward a General Theory of Adaptive Walks on Rugged Landscapes. *Journal of Theoretical Biology*. 1987(128), 11-45.

Kauffman S. A., & Weinberger (1989) - The NK Model of Rugged Fitness Landscapes And Its Application to Maturation of the Immune Response. *Journal of Theoretical Biology*. 1989(141), 211-245.

Khaleghi, B. (2019). The What of Explainable AI. Retrieved Aug 9, 2020, from https://www.elementai.com/news/2019/the-what-of-explainable-ai.

Kibria, M. G., Nguyen, K., Villardi, G. P., Zhao, O., Ishizu, K., & Kojima, F. (2018). Big data analytics, machine learning, and artificial intelligence in next-generation wireless networks. *IEEE access*, *6*, 32328-32338.

Kim, Y., & Rush, A. M. (2016, November). Sequence-level knowledge distillation. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 1317-1327).

Kinnunen, H., Holm, E., Nordman, A.-M., & Roschier, S. (2018). Academic consulting –fEarly Stop income stream, impact and brand building. *International Journal of Innovation Science*, *10*(2), 143. https://doi.org/10.1108/IJIS-08-2017-0075.

Knight, W. (2020). Watson's Creator Wants to Teach AI a New Trick: Common Sense. Retrieved May 31, 2020, from https://www.wired.com/story/watsons-creator-teach-ai-new-trick-common-sense.

Knight, W. (2020). It's Called Artificial Intelligence—but What Is Intelligence. Retrieved May 31, 2020, from https://www.wired.com/story/its-called-artificial-intelligence-but-what-is-intelligence.

Kolbjørnsrud, V., Amico, R., & Thomas, R. J. (2016). How artificial intelligence will redefine management. *Harvard Business Review*, 2–6.

Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine*, *23*(1), 89-109.

Kristianto, Y., Gunasekaran, A., Helo, P., & Sandhu, M. (2012). A decision support system for integrating manufacturing and product design into the reconfiguration of the supply chain networks. *Decision Support Systems*, 52, 790–801. https://doi.org/10.1016/j.dss.2011.11.014.

Kuhn, R., & Kacker, R. (2019). An application of combinatorial methods for explainability in artificial intelligence and machine learning (draft) (pp. 7-7). National Institute of Standards and Technology.

Kumar, V., Rajan, B., Venkatesan, R., & Lecinski, J. (2019). Understanding the role of artificial intelligence in personalized engagement marketing. *California Management Review*, *61*(4), 135-155.

Kumar, M., & Roy, S. (2018). Max-min fuzzy programming approach for compromise farming: a case study. *Agricultural Economics Research Review*, 31(2), 291–297. https://doi.org/10.5958/0974-0279.2018.00046.0.

Kundu, S., Basu, K., Sadi, M., Titirsha, T., Song, S., Das, A., & Guin, U. (2021, April). Special session: Reliability analysis for ai/ml hardware. In *2021 IEEE 39th VLSI Test Symposium (VTS)* (pp. 1-10). IEEE.

Kureishy, A., Meley, C., & Mackenzie, B. (2019). Achieving Real Business Outcomes from Artificial Intelligence. Boston: O'Reilly.

Lafley, A. G., & Martin, R. L. (2013). *Playing to win: How strategy really works*. Harvard Business Press.

Lalmuanawma, S., Hussain, J., & Chhakchhuak, L. (2020). Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. *Chaos, Solitons & Fractals*, *139*, 110059.

Lauterbach, A., & Bonime-Blanc, A. (2016). Artificial Intelligence: A Strategic Business And Governance Imperative. NACD Directorship, 42(5), 54–57.

Lee, H. W. (2017). When pursuing multiple goals, people prioritize the minimally acceptable level over the aspiration level.

Lei, N., & Moon, S. K. (2015). A Decision Support System for market-driven product positioning and design. *Decision Support Systems*, 69, 82–91. https://doi.org/10.1016/j.dss.2014.11.010.

Levinthal, D. A. (1997). Adaptation on Rugged Landscapes. *Management Science*. 43(7), 934-950. http://www.jstor.org/stable/2634336.

Levinthal, D. A. (2011). A Behavioral Approach to Strategy—what's the alternative? *Strategic Management Journal*, 32, 1517–1523. https://doi.org/10.1002/smj.963.

Levinthal, D. A., Warglien, M. (1999). Landscape Design: Designing for Local Action in Complex Worlds. *Organization Science*. 10(3), 342-357.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems, 793, 9459-9474. https://doi.org/10.48550/arXiv.2005.11401.

Li, M., Sun, J., Tan, X. 2024. Evaluating the effectiveness of large language models in abstract screening: a comparative analysis. *Systematic Reviews* (2024) 13:219. https://doi.org/10.1186/s13643-024-02609-x.

Li, S. (1999). The development of a hybrid intelligent system for developing marketing strategy. *Decision Support Systems*, 27(4), 395-409. https://doi.org/10.1016/S0167-9236(99)00061-5.

Li, S., & Li, J. Z. (2009). Hybridising human judgment, AHP, simulation and a fuzzy expert system for strategy formulation under uncertainty. *Expert Systems With Applications*, 36(Part 1), 5557–5564. https://doi.org/10.1016/j.eswa.2008.06.095.

Li, S., & Li, J. Z. (2010). WebInternational: Combining Web-based knowledge automation, fuzzy rules and on-line databases for international marketing planning. *Expert Systems With Applications*, 37, 7094–7100. https://doi.org/10.1016/j.eswa.2010.03.007.

Li, S., Zheng Li, J., He, H., Ward, P., & Davies, B. J. (2011). WebDigital: A Web-based hybrid intelligent knowledge automation system for developing digital marketing strategies. *Expert Systems With Applications*, 38, 10606–10613. https://doi.org/10.1016/j.eswa.2011.02.128.

Liao, S. (2005a). Expert system methodologies and applications—a decade review from 1995 to 2004. *Expert Systems With Applications*, 28, 93–103. https://doi.org/10.1016/j.eswa.2004.08.003.

Liao, S. (2005b). Technology management methodologies and applications. A literature review from 1995 to 2003. *Technovation*, 25, 381–393. https://doi.org/10.1016/j.technovation.2003.08.002.

Liu, S., & Vicente, L. N. (2022). Accuracy and fairness trade-offs in machine learning: A stochastic multi-objective approach. *Computational Management Science*, 1-25.

Liu, B., Ding, M., Shaham, S., Rahayu, W., Farokhi, F., & Lin, Z. (2021). When machine learning meets privacy: A survey and outlook. *ACM Computing Surveys (CSUR)*, *54*(2), 1-36.

Liu, Y., Iter, D., Xu, Y., Wang, S., Xu, R., Zhu, C. 2023. G-EVAL: NLG Evaluation using GPT-4 with Better Human Alignment.

Lo, K. M., Huang, Z., Qiu, Z., Wang, Z., & Fu, J. (2024). A closer look into mixture-of-experts in large language models. *arXiv preprint arXiv:2406.18219*.

London, A. J. (2019). Artificial intelligence and black-box medical decisions: accuracy versus explainability. *Hastings Center Report*, *49*(1), 15-21.

Lu, Z., & Shen, H. (2015, July). A security-assured accuracy-maximised privacy preserving collaborative filtering recommendation algorithm. In *Proceedings of the 19th International Database Engineering & Applications Symposium* (pp. 72-80).

Luger, G. F., & Stubblefield, W. A. (1998). *Artificial intelligence: structures and strategies for complex problem solving*. Don Mills, Ont.; Reading, Mass.: Addison-Wesley, Longman, c1998.

Luo, Y., Zhong, J., Liu, W. L., & Chen, W. N. (2022). Automatic Business Location Selection through Particle Swarm Optimization and Neural Network.

Magoulas, R., & Swoyer, S. (2020). *AI Adoption in the Enterprise*. Beijing: O´ Reilly.

Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures*, 90, 46–60.

Majeed, A.; Hwang, S.O. (2022). Data-Driven Analytics Leveraging Artificial Intelligence in the Era of COVID-19: An Insightful Review of Recent Developments. *Symmetry*, 14, 16. https://doi.org/10.3390/sym14010016.

Managing the ethical and risk implications of rapid advances in artificial intelligence: A literature review. (2016). 2016 Portland International Conference on Management of Engineering and Technology (PICMET), Management of Engineering and Technology (PICMET), 2016 Portland International Conference On, 682.

Manheim, K., & Kaplan, L. (2019). Artificial intelligence: Risks to privacy and democracy. *Yale JL & Tech.*, *21*, 106.

Markovic, M. (2019). Rise of the robot lawyers. *Ariz. L. Rev.*, *61*, 325.

Martin, R. L., & Riel, J. (2019) - The One Thing You Need to Know About Managing Functions. *Harvard Business Review*. https://hbr.org/2019/07/the-one-thing-you-need-to-know-about-managing-functions

Martin, R. L. (2021). Strategy vs. Planning: Complements not Substitutes. Retrieved from https://rogermartin.medium.com/strategy-vs-planning-complements-not-substitutes-ea08e56809d6.

Martínez-López, F. J., & Casillas, J. (2013). Artificial intelligence-based systems applied in industrial marketing: An historical overview, current and future insights. *Industrial Marketing Management*, 42, 489–495. https://doi.org/10.1016/j.indmarman.2013.03.001.

Mazeh, I., & Shmueli, E. (2019). A personal data store approach for recommender systems: enhancing privacy without sacrificing accuracy. *Expert Systems with Applications*, *139*, 112858.

McGuire, J. (2023). *Empirical evaluations of transformers* (Master's thesis, Rutgers The State University of New Jersey, School of Graduate Studies).

McKelvey, B. (1999). Avoiding Complexity Catastrophe in Coevolutionary Pockets: Strategies for Rugged Landscapes. *Organization Science*, 10, 1999, 294-321.

McKelvey, B. (2000). Toward A Model-Centered Strategy Science: More Experiments, Less History Published in Research in Competence-Based Management. *Research in Competence-Based Management*—Part C, R. Sanchez and A. Heene (eds), Stamford CT: JAI Press, 217–253.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, *54*(6), 1-35.

Meinhart, W. A. (1966). Artificial Intelligence, Computer Simulation of Human Cognitive and Social Processes, and Management Thought. *The Academy of Management Journal*, 9(4), 294. https://doi.org/10.2307/254948.

Metcalf, L., Askay, D. A., & Rosenberg, L. B. (2019). Keeping humans in the loop: pooling knowledge through artificial swarm intelligence to improve business decision making. *California Management Review*, *61*(4), 84-109.

Microsoft. (2020). *Visual Studio Code* (Version 1.45.1) [Computer software].

Min, H. (2010). Artificial intelligence in supply chain management: theory and applications. *International Journal of Logistics*, (1), 13. https://doi.org/10.1080/13675560902736537.

Mintzberg, H. (1987a). Crafting strategy. *Harvard Business Review*, 65(4), 66.

Mintzberg, H. (1987b). The Strategy Concept I: Five Ps For Strategy. *California Management Review*, 30(1), 11. https://doi.org/10.2307/41165263.

Mintzberg, H. (1987c). The Strategy Concept II: Another Look at Why Organizations Need Strategies. California Management Review, 30(1), 25. https://doi.org/10.2307/41165264.

Mitchell, T. M. (1997). Does machine learning really work?. *AI magazine*, *18*(3), 11-11.

Mittelstadt, B., Russell, C., & Wachter, S. (2019, January). Explaining explanations in AI. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 279-288).

McCarthy, J., & Hayes, P. J. (1981). Some Philosophical Problems from the Standpoint of Artificial Intelligence. Readings in Artificial Intelligence, 431–450. https://doi.org/10.1016/B978-0-934613-03-3.50033-7.

Moldoveanu, M. C. (2009). Thinking strategically about thinking strategically: The computational structure and dynamics of managerial problem selection and formulation. *Strategic Management Journal*, 30: 737–763. https://doi.org/10.1002/smj.757.

Moldoveanu, M. C. (2011). *Inside Man*: *The Discipline of Modeling Human Ways of Being.* Stanford: Business Books Stanford.

Moldoveanu, M. (2016). Algorithmic Foundations for Business Strategy. Working Paper 17-036. Harvard Business School.

Moldoveanu, M. (2019a). Intelligent Artificiality: Algorithmic Microfoundations for Strategic Problem Solving. Working Paper 19-072. Harvard Business School.

Moldoveanu, M. C. (2019b) - 'Organizing', 'Innovating' and 'Managing' in Complexity Space. Working Paper 19-128. Harvard Business School.

Moldoveanu, M. (2019c). AI_Enabled_CXO_June_2019. Unpublished manuscript.

Moldoveanu, M. (2019d). AI_Activity_Chain_July_12_2019. Unpublished manuscript.

Moldoveanu, M. C. (2020). AI_Value_Chain_Tradeoffs. Unpublished manuscript.

Moldoveanu, M. (2021). Probably Approximately Ethical: The Tradeoffs Every Executive Needs to Understand to Make ML Work at Scale. Unpublished manuscript.

Moldoveanu, M. (2022). Problem Scoping for Enterprise-Wide ML Deployment. Unpublished manuscript.

Moldoveanu, M. C., Bauer, R. M. (2004). On the Relationship Between Organizational Complexity and Organizational Structuration. *Organization Science*, 15(1), 98-118. https://doi.org/10.1287/orsc.1030.0058.

Moldoveanu, M. C., & Baum, J. A. C. (2014). Epinets: The Epistemic Structure and Dynamics of Social Networks. Stanford: Stanford University Press.

Moldoveanu, M. C., Leclerc, O. P. (2015). *The Design of Insight: How to Solve Any Business Problem*. Stanford, California: Stanford University Press.

Molnar, C. (2020). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Leanpub.

Mosele, J. (2018). *Artificial Intelligence in the Sports Industry* (Unpublished master's thesis). Politecnico Di Milano, Milan, Italy.

Namaki, M. S. S. E. (2018). How Companies are Applying AI to the Business Strategy Formulation. *Scholedge International Journal of Business Policy & Governance*, 5(8), 77–82. https://doi.org/10.19085/journal.sijbpg050801.

Näsi, J. (1999). Information systems and strategy design. The knowledge creation function in three modes of strategy-making. *Decision Support Systems*, 26, 137-149. https://doi.org/10.1016/S0167-9236(99)00025-1.

Negnevitsky, M. (1997). The history of artificial intelligence or from the 'dark ages' to the knowledge-based systems, Applications of Artificial Intelligence in Engineering XII, Computational Mechanics Publications, R.A.Adey, G.Rzevski and R.Teti (eds), Boston, pp. 571-591. ISBN 9-85312-471-0 (1997) [Research Book Chapter]

Negnevitsky, M. (2005). *Artificial Intelligence: A guide to intelligent systems*. Harlow: Pearson.

Ng, Andrew (2016). What Artificial Intelligence Can and Can't Do Right Now. Retrieved March 8, 2019, from https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now.

Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V.D., López García, Á., Heredia, I., Malík, P., & Hluchý, L. (2019). Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. *Artificial Intelligence Review,* 52(1), 77-124.

Nomiks. (2023). An introduction to objective functions. Retrieved Jan 20, 2024, from https://medium.com/@Nomiks/an-introduction-to-objective-functions-bd43e44bff68.

OECD (2024), Explanatory memorandum on the updated OECD definition of an AI system. *OECD Artificial Intelligence Papers*, No. 8. OECD Publishing, Paris. https://doi.org/10.1787/623da898-en.

Oke, S. A. (2008). A Literature Review on Artificial Intelligence. *International Journal of Information and Management*, (4), 535.

Oneto, L., & Chiappa, S. (2020). Fairness in machine learning. In *Recent Trends in Learning From Data* (pp. 155-196). Springer, Cham.

Overgoor, G., Chica, M., Rand, W., & Weishampel, A. (2019). Letting the computers take over: Using AI to solve marketing problems. *California Management Review*, *61*(4), 156-185.

Papadimitriou, C. H. 1977. The Euclidean travelling salesman problem is NP-complete. *Theoretical Computer Science*, 4(3), 237-244.

Papernot, N. (2017, November). Security and Privacy in Machine Learning. Pennsylvania State University & Google Brain. Retrieved July 3, 2021, from https://www.microsoft.com/en-us/research/uploads/prod/2018/03/Nicolas-Papernot-Security-and-Privacy-in-Machine-Learning.pdf.

Papernot, N., McDaniel, P., Sinha, A., & Wellman, M. P. (2016). Sok: Security and privacy in machine learning. In *2018 IEEE European Symposium on Security and Privacy (EuroS&P)* (pp. 399-414). IEEE.

Park, W., Kim, D., Lu, Y., & Cho, M. (2019). Relational knowledge distillation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3967-3976).

Partington, D. (Ed.). (2002). Essential skills for management research. Sage.

Patriarca, R., Di Gravio, G., Cioponea, R., & Licu, A. (2022). Democratizing business intelligence and machine learning for air traffic management safety. *Safety Science*, *146*. https://doi.org/10.1016/j.ssci.2021.105530.

Pessach, D., Shmueli, E. (2022). A Review on Fairness in Machine Learning. ACM Computing Survey. 55, 3, Article 51 (February 2022). https://doi.org/10.1145/3494672.

Phan, L., Gatti, A., Han, Z., Li, N., Hu, J., Zhang, H., Shaaban, M., Ling, J., Shi, S., Choi, M., Agrawal, A., Chopra, A., Khoja, A., Kim, R., Ren, R., Hausenloy, J., Zhang, O., Mazeika, M., Anderson, D., ... Hendrycks, D. (2025). Humanity's Last Exam. arXiv preprint arXiv:2501.14249.

Piano, S. (2020). Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. *Humanities and Social Sciences Communications*, *7*(1), 1-7.

Pinson, S. D., Louca, J. A., & Moraitis, P. (1997). A distributed decision support system for strategic planning. *Decision Support Systems*, (1), 35. https://doi.org/10.1016/S0167-9236(96)00074-7.

Porter, M. E. (1979). How competitive forces shape strategy. *Harvard Business Review*, *57*(2), 137-145.

Porter, M. E. (1980a). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.

Porter, M. E. (1980b). How competitive forces shape strategy. *McKinsey Quarterly*, Spring 1980, 34-50.

Porter, M. E. (1985), Competitive Advantage, New York: Free Press.

Porter, M. E. (1991). Towards a Dynamic Theory of Strategy. *Strategic Management Journal*, 12, 95-117.

Porter, M. E. (1996). What is Strategy. *Harvard Business Review*, 74(6), 61–78.

Porter, M. E., Millar, V. E. (1985) - How information gives you competitive advantage. *Harvard Business Review*, July-August 1975, 149-160.

Prabhakar, S., Pankanti, S., & Jain, A. K. (2003). Biometric recognition: Security and privacy concerns. *IEEE security & privacy*, *1*(2), 33-42.

Prahalad, C. K., Hamel, G. (1990). The core competence of the corporation. *Harvard Business Review,* May-June 1990, 71-91.

Python (Version 2.7.16) [Computer software]. Python.org.

Python (Version 3.12.1) [Computer software]. Python.org.

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137-141.

Rashidi, H. H., Tran, N. K., Betts, E. V., Howell, L. P., & Green, R. (2019). Artificial intelligence and machine learning in pathology: the present landscape of supervised methods. *Academic pathology*, *6*, 2374289519873088.

Raybaut, P. (2009). Spyder-documentation. *Available Online at: Spyder-IDE.org*.

Reeves, M., & Moldoveaunu, M. (2017). Artificial intelligence: the gap between promise and practice. Scientific America.

Reinsel, D., Gantz, J., & Rydning, J. (2018). The digitization of the world from edge to core. IDC white paper, 16, 1-28.

Rezende, D. J., Mohamed, S., & Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models. In International conference on machine learning (pp. 1278-1286). PMLR.

Rivkin, J. W. (2000). Imitation of Complex Strategies. *Management Science*. 46(6), 824-844.

Rivkin, J. W., Siggelkow, N. (2003). Balancing Search and Stability/ Interdependencies Among Elements of Organizational Design. *Management Science*. 49(3), 290-311.

Rivkin, J. W., Siggelkow, N. (2007). Patterned Interactions in Complex Systems: Implications for Exploration. *Management Science*. 53(7), 1068-1085.

Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable machine learning for scientific insights and discoveries. *Ieee Access*, *8*, 42200-42216.

Roy, R., Raiman, J., Kant, N., Elkin, I., Kirby, R., Siu, M., Oberman, S.F., Godil, S., & Catanzaro, B. (2021). PrefixRL: Optimization of Parallel Prefix Circuits using Deep Reinforcement Learning. *2021 58th ACM/IEEE Design Automation Conference (DAC)*, 853-858.

Royal Society, The (2019). Explainable AI: the basics, Policy briefing. Retrieved August 9, 2020, from https://royalsociety.org/ai-interpretability.

Roychowdhury, S., Soman, S., Ranjani, H. G., Gunda, N., Chhabra, V., & Bala, S. K. (2024). Evaluation of rag metrics for question answering in the telecom domain. arXiv preprint arXiv:2407.12873.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206-215.

Russell, S. J., & Norvig, P. (1995). *Artificial intelligence: A Modern Approach*. New Jersey: Prentice Hall.

Russell, S. & Norvig, P. (2020). Artificial Intelligence: A Modern Approach (4th edition). Pearson.

Schmelzer, R. (2019). Understanding Explainable AI. Retrieved Aug 9, 2020, from https://www.forbes.com/sites/cognitiveworld/2019/07/23/understanding-explainable-ai.

Schneider, S., & Leyer, M. (2019). Me or information technology? Adoption of artificial intelligence in the delegation of personal strategic decisions. *Managerial and Decision Economics*, 40(3), 223-231.

Scotland, J. (2012) Exploring the Philosophical Underpinnings of Research: Relating Ontology and Epistemology to the Methodology and Methods of the Scientific, Interpretive, and Critical Research Paradigms. English Language Teaching; Vol. 5, No. 9; 2012. doi:10.5539/elt.v5n9p9.

Sheikh, H., Prins, C., & Schrijvers, E. (2023). Artificial intelligence: definition and background. In *Mission AI: The new system technology* (pp. 15-41). Cham: Springer International Publishing.

Shokri, R., & Shmatikov, V. (2015, October). Privacy-preserving deep learning. In *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security* (pp. 1310-1321).

Shrestha, Y. R., Ben-Menahem, S. M., & Von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, *61*(4), 66-83.

Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. *Cutter business technology journal*, *31*(2), 47-53.

Singh, J., Flaherty, K., Sohi, R.S., Deeter-Schmelz, D., Habel, J., Le Meunier-Fitzhugh, K., Malshe, A., Mullins, R., & Onyemah, V. (2019). Sales profession and professionals in the age of digitization and artificial intelligence technologies: concepts, priorities, and questions. *Journal of Personal Selling & Sales Management,* 39(1), 2-22.

Siggelkow, N. (2003). Why focus? A study of intra-industry focus effects. *The Journal of Industrial Economics*, *51*(2), 121-150.

Siggelkow, N. (2007). Persuasion with case studies. *Academy of management journal*, *50*(1), 20-24.

Siggelkow, N., Levinthal, D. A. (2003). Temporarily Divide to Conquer: Centralized, Decentralized, and Reintegrated Organizational Approaches to Exploration and Adaptation. *Organization Science*. 14(6), 650-669.

Siggelkow, N., & Rivkin, J. W. (2005). Speed and search: Designing organizations for turbulence and complexity. *Organization Science*, *16*(2), 101-122.

Sipser, M. (1992, July). The history and status of the P versus NP question. In *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing* (pp. 603-618).

Song, H. F., Yang, G. R., & Wang, X. J. (2017). Reward-based training of recurrent neural networks for cognitive and value-based tasks. *Elife*, *6*, e21492.

Skillicorn, D. B., & Talia, D. (1998). Models and languages for parallel computation. *Acm Computing Surveys (Csur)*, *30*(2), 123-169.

REFERENCES

Smith, C., McGuire, B., Huang, T., and Yang, G. (2006). The History of Artificial Intelligence. History of Computing CSEP 590A University of Washington.

Sparks, K. (2018, February 5). AND, OR, XOR — Bitwise Operators. Retrieved Nov 26, 2023, from https://krissparks.medium.com/and-or-xor-bitwise-operators-2ae3fbcfa5bc.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), 1929-1958.

Stein, S. (2020). The weird, frightening future of AI may not be what you think. Retrieved June 10, 2020, from https://www.cnet.com/science/the-weird-frightening-future-of-ai-may-not-be-what-you-think.

Stonehouse G., Snowdon W. (2007). Competitive Advantage Revisited: Michael Porter on Strategy and Competitiveness. *Journal of Management Inquiry,* 16(3), 256-273. https://doi.org/10.1177/1056492607306333.

Syed, A. B., & Zoga, A. C. (2018). Artificial Intelligence in Radiology: Current Technology and Future Directions. *Seminars in Musculoskeletal Radiology*, *22*(5), 540–545. https://doi.org/10.1055/s-0038-1673383.

Szychter, A., Ameur, H., Kung, A., & Daussin, H. (2018). The impact of artificial intelligence on security: a dual perspective.

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, *61*(4), 15-42.

Tjoa, E., & Guan, C. (2020). A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE transactions on neural networks and learning systems*, *32*(11), 4793-4813.

Tolan, S., Miron, M., Gómez, E., & Castillo, C. (2019, June). Why machine learning may lead to unfairness: Evidence from risk assessment for juvenile justice in catalonia. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law* (pp. 83-92).

Trajtenberg, M. (2018). AI as the Next GPT: A Political-Economy Perspective. National Bureau of Economic Research. *Working Paper Series*. https://doi.org/10.3386/w24245.

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind* 49, 433-460. https://doi.org/10.1093/mind/LIX.236.433.

The role of artificial intelligence in understanding the strategic decision-making process. (1991). *IEEE Transactions on Knowledge and Data Engineering*, 3(2), 149-159. https://doi.org/10.1109/69.87995.

US Food and Drug Administration (2019). Proposed regulatory framework for modifications to Artificial Intelligence/Machine Learning (AI/ML)-based Software as a Medical Device (SaMD).

Valdivia, A., Sánchez-Monedero, J., & Casillas, J. (2021). How fair can we go in machine learning? Assessing the boundaries of accuracy and fairness. *International Journal of Intelligent Systems*, *36*(4), 1619-1643.

Valera, I., Singla, A., & Gomez Rodriguez, M. (2018). Enhancing the accuracy and fairness of human decision making. *Advances in Neural Information Processing Systems*, *31*.

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Varian, H. (2018). Artificial intelligence, economics, and industrial organization. In *The economics of artificial intelligence: an agenda* (pp. 399-419). University of Chicago Press.

Vats, A., Raja, R., Jain, V., & Chadha, A. (2024). The evolution of mixture of experts: A survey from basics to breakthroughs. *Preprints (August 2024)*.

Visual Studio Code (Version 1.45.1) [Computer software]. Microsoft.

Vellido, A. (2019). Societal issues concerning the application of artificial intelligence in medicine. *Kidney Diseases*, *5*(1), 11-17.

Vollmer, S.J., Mateen, B.A., Bohner, G., Király, F.J., Ghani, R., Jónsson, P., Cumbers, S., Jonas, A., McAllister, K.S., Myles, P., Granger, D., Birse, M., Branson, R., Moons, K.G., Collins, G.S., Ioannidis, J.P., Holmes, C., & Hemingway, H. (2018). Machine learning and AI research for Patient Benefit: 20 Critical Questions on Transparency, Replicability, Ethics and Effectiveness. *ArXiv, abs/1812.10404*.

Wang, F-Y., Miao, Q., Li, X., Wang, X., Lin, Y. (2023). What Does ChatGPT Say: The DAO from Algorithmic Intelligence to Linguistic Intelligence. IEEE/CAA Journal of Automatica Sinica, 10(3).

Wardat, Y., Tashtoush, M. A., AlAli, R., Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *EURASIA Journal of Mathematics, Science and Technology Education*, 2023, 19(7), em2286. https://doi.org/10.29333/ejmste/13272.

Weinberger, E. (1990). Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biological cybernetics*, *63*(5), 325-336.

Weinberger, E. D. (1996). NP completeness of Kauffman's NK model, a tuneably rugged fitness landscape. Santa Fe Institute Technical Reports.

Williams, G. (2021, July). *Global Corporate Capex Survey 2021*. S&P Global. https://www.spglobal.com/_assets/documents/ratings/research/100288873.pdf.

Wigderson, A. 2006. P, NP and mathematics – a computational complexity perspective.

Williamson, S. (2006). *Notes on macroeconomic theory*. University in St. Louis. Department of Economics.

Wilson, J. (2014). Essentials of business research: A guide to doing your research project. *Essentials of Business Research*, 1-376.

Wojczynski, M. K., & Tiwari, H. K. (2008). Definition of phenotype. *Advances in genetics*, *60*, 75-105.

Wolpert, D. H., & Macready, W. G. (1996). The Mathematics of Search.

Wolpert, D. H., Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, 1997. https://doi.org/10.1109/4235.585893.

Wright, S. (1931), Evolution in Mendelian Populations. *Genetics*, 16, 97-159.

Wright, S. (1932). The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution. In D. F. Jones (Ed.), *Intl. Proceedings of the Sixth International Congress on Genetics*, 1, 356-366. Brooklyn, NY: Brooklyn Botanic Garden,

Wu, H., Han, H., Wang, X., & Sun, S. (2020). Research on artificial intelligence enhancing internet of things security: A survey. *Ieee Access*, *8*, 153826-153848.

Xu, Z., & Liu, H. (2021). Knowledge accumulating: The general pattern of learning. *arXiv preprint arXiv:2108.03988*.

Xu, R., Baracaldo, N., & Joshi, J. (2021). Privacy-preserving machine learning: Methods, challenges and directions. *arXiv preprint arXiv:2108.04417*.

Yan, Z. (2024). Evaluating the Effectiveness of LLM-Evaluators (aka LLM-as-Judge). eugeneyan.com.

Yang, Y., Fu, C., Chen, Y.-W., Xu, D.-L., & Yang, S.-L. (2016). A belief rule based expert system for predicting consumer preference in new product development. *Knowledge-Based Systems*, 94, 105-113. https://doi.org/10.1016/j.knosys.2015.11.012.

Yun, L., Zhuang, Y., Fu, Y., Xing, E. P., & Zhang, H. (2024). Toward inference-optimal mixture-of-expert large language models. arXiv preprint arXiv:2404.02852.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Computation*, 8, 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X.

Zadeh, L. A. (1975). Fuzzy Logic and Approximate Reasoning (In Memory of Grigore Moisil). Synthese, 30(3/4), 407. https://doi.org/10.1007/BF00485052.

Zadeh, L. A. (1976). A fuzzy-algorithmic approach to the definition of complex or imprecise concepts. *International Journal of Man-Machine Studies*, 8(3), 249-291. https://doi.org/10.1016/S0020-7373(76)80001-6.

Zadeh, L. A. (1983). The role of fuzzy logic in the management of uncertainty in expert systems. *Fuzzy Sets and Systems*, 11(1–3), 199. https://doi.org/10.1016/S0165-0114(83)80081-5.

Zadeh, L. A. (1996). Fuzzy logic = computing with words. *IEEE Transactions on Fuzzy Systems*, 4(2), 103. https://doi.org/10.1109/91.493904.

Zadeh, L. A. (1997). Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets and Systems*, (2), 111. https://doi.org/10.1016/S0165-0114(97)00077-8.

Zadeh, L. A. (2015). Fuzzy logic—a personal perspective. Fuzzy Sets and Systems, 281, 4-20. https://doi.org/10.1016/j.fss.2015.05.009.

Zeng, W., Ren, X., Su, T., Wang, H., Liao, Y., Wang, Z., Jiang, X., Yang, Z., Wang, K., Zhang, X., Li, C., Gong, Z., Yao, Y., Huang, X., Wang, J., Yu, J., Guo, Q., Yu, Y., Zhang, Y., ... Tian, Y. (2021). PanGu-α: Large-scale Autoregressive Pretrained Chinese Language Models with Auto-parallel Computation. *ArXiv, abs/2104.12369*.

Zhou, N., Zhang, Z., Nair, V. N., Singhal, H., Chen, J., & Sudjianto, A. (2021). Bias, Fairness, and Accountability with AI and ML Algorithms. *arXiv preprint arXiv:2105.06558*.

Zhou, Y., Lei, T., Liu, H., Du, N., Huang, Y., Zhao, V., Dai, A.M., Chen, Z., Le, Q., & Laudon, J. (2022). Mixture-of-Experts with Expert Choice Routing. 36th Conference on Neural Information Processing Systems (NeurIPS 2022). *ArXiv, abs/2202.09368*.

Zimmermann, H. J. (1985). *Fuzzy set theory and its applications*. International series in management science/Operations research. Dordrecht: Kluwer-Nijhoff Publishing. https://doi.org/10.1007/978-94-015-7153-1.

Zimmermann, H.-J. (1993). *Fuzzy Sets, Decision Making, and Expert Systems*. Retrieved Feb 9, 2019, from https://books.google.ca/books?hl=en&lr=&id=-PfnCAAAQBAJ.

Zimmermann, H.-J. (2012). Fuzzy sets, decision making, and expert systems (Vol. 10)). Berlin: Springer Science & Business Media.

Zliobaite, I. (2015). On the relation between accuracy and fairness in binary classification. *arXiv preprint arXiv:1505.05723*.

European Parliament and Council of the European Union (2016). General Data Protection Retrieved from https://gdpr-info.eu.

State of California Department of Justice (2018). California Consumer Privacy Act. Retrieved from https://oag.ca.gov/privacy/ccpa.

AI Cases, grouped by topic, para-cited:

Agriculture

Tiwari et al. (2018). Analytics Empowering Agriculture: Jayalaxmi Agro Tech

Aviation

Makinen and Burgelman. (2018). Noodle AI

Pah et al. (2018). Evaluating the Cognitive Analytics Frontier

Youngdahl (2017) JetBlue and Gladly: Onmichannel Customer Service

Banking & Financial Services

Armacord Incorporated: Combatting Money-laundering Using Data Analytics

Chandrasekhar (2017). RBC Social Network Analysis

Cohen et al. (2018). Artificial Intelligence and the Machine Learning Revolution in Finance

Datar and Bowler (2018). Lending Club (A,B,C)

Hou et al. (2016). Alibaba's Growth Frenzy

Johnston and Peterson (2018). Commercial International Bank

Khachatryan, D. (2014).

Kumar et al. (2016). Predicting Earnings Manipulation by Indian Firms

Liang and Zheng (2016). Data Analytics at DBS Group Audit

Meng and Zhu (2019). A Chairman's Decision: Launching a Robo-Advisor in CCB Principal Asset Management Company

Niessing and Plassmann (2017). Telenor - Revolutionizing Retail Banking in Serbia

Ovchinnikov (2018). Private Banking Advisers at BCB Edmonton (A,B,C)

Schifrin and Shaw (2017). RBC - Using People Strategy and Analytics to Drive Employee Performance (A,B)

Wang and Thomas (2018). New Constructs: Disrupting Fundamental Analysis with Robo-Analysts

Wendell, C. (2016). Bringing Quick Loans to the Unbankable (A), (B)

Communications

Kar (2015). Growing the Online Portal of Business Fundas

Kerr and Moloney (2018). Vodafone - Managing Advanced Technologies & Artificial Intelligence

Customer Service

Avery and Steenburgh (2018). Hubspot & Motion AI

Brandwein and Niessing (2019). WeWork - Service Excellence through Business Model Innovation

Seth and Kumar (2017). Improving Customer Engagement Through Analytics

Education & Coaching

Sawhney and Goodman (2018). VMock - Pivoting to Succeed and Scale

Fashion

Bussgang and Kelley (2018). Choosy

Food and Nutrition

Amraham et al. (2016). Customer Analytics at Bigbasket – Product Recommendations

Israeli A., Lane D., (2019). DayTwo: Going to market with Gut Microbiome

Pan et al. (2019). Seng Hua Hng Foodstuffs: Cracking the Camel Nit for Growth, Using Data Analytics

Vimala et al. (2018). Nestlé: Developing a Digital Nutrition Platform for Japan

General

Applegate et al. (2017). Podium Data: Harnessing the Power of Big Data Analytics

Kerr et al. (2019). Autonomous Vehicles: The Rubber Hits the Road...but When?

Mohasseb et al. (2019). Two Brothers, Two Methods: “Happiness Index” vs. “Data & Analytics”

Ovchinnikov, A. (2018). Retention Modeling at Scholastic Travel Company (A)

Ovchinnikov, A. (2018). Retention Modeling at Scholastic Travel Company (B)

Health Services

Aggarwal, R. (2017). PatientsLikeMe: Using Social Network Health Data to Improve Patient Care

Greenstein S., Gulick S. (2018). Zebra Medical Vision

Ma et al. (2016). Data Analytics at Alexandra Health System: A New Journey in the Healthcare Industry

Quelch, J. A., Rodriguez, M. L. (2015). Carolinas HealthCare System: Consumer Analytics

Rahul et al. (2017). Predicting Net Promotor Score (NOS) to Improve Patient Experience at Manipal Hospitals

Home Services

Cui et al. (2018). Dami Technology Co. Ltd.: A Smart Kitchen Ecosystem

Lal, R. Johnson, S. (2018). Amazon, Google, and Apple: Smart Speakers and the Battle for the Connected Home

Walker R. (2019). Alexa: A Pandora's Box of Risks

Yoffie et al. (2018). Voice War: Hey Google vs. Alexa vs. Siri

Human Resources

Bernstein et al. (2017). GROW: Using Artificial Intelligence to Screen Human Intelligence

Buche, I. (2016). Recruit Japan: Harnessing Data to Create Value (Abridged)

Buche, I. (2016). Recruit Japan: Harnessing Data to Create Value

Dutta, D., Venkatagiri, S. (2017). Edge Networks: Making Hr Intelligent

Guetta, C. D. (2019). An Introduction to AI for Text Mining: A Companion to the Evisort Case

Guetta, C. D. (2018). Evisort: An A.I.-Powered Start-up Uses Text Mining to Become Google for Contracts

Joseph et al. (2018). Evie.AI: The Rise of Artificial Intelligence, and the Future of Work

Lakhani et al. (2016). Aspiring Minds

Mark, K. (2017). Agoda: People Analytics and Business Culture (A)

Mark, K. (2017). Agoda: People Analytics and Business Culture (B)

Rahul, K., Kumar, U. D. (2016). HR Analytics at ScaleneWorks – Behavioral Modeling to Predict Renege

Innovation

Davis et al. (2019). ByteDance Beyond China: Leveraging Consumer Artificial Intelligence (AI) from Toutiao to Musical.ly and TikTok

Duke et al. (2019). Nvidia: Winning the Deep-Learning Leadership Battle

Farronato et al. (2018). Innovation at Uber: The Launch of Express POOL

Orleans, A., Leslie, M. (2019). Cerebras: A Tale of Dreams and Risks

Yoffie et al. (2018). Numenta: Inventing and (or) Commercializing AI

Youngdahl, W. E., Hunsaker, B. T. (2018). Coda Cofee and bext360 Supply Chain: Machine Vision, AI, IoT, and Blockchain

Youngdahl, W. E., Hunsaker, B. T. (2018). SingularityNET: Blockchain-Driven AI Marketplace and Quest

Xie et al. (2018). SenseTime: World's Most Valuable Artificial Intelligence Startup

Manufacturing

Chandrasekhar, R. (2017). Dow Chemical Co.: Big Data in Manufacturing

Hall, O. P., Ko, K. (2015). Trionym Systems: Investment Decision-Making Using Prescriptive Analytics

Lehman, C., Siegel, R. (2017). General Electric in 2017: Naming and Claiming the Industrial Internet

Markoff, M., Seifert, R. W. (2018). Tetra Pak: A Digitally enabled Supple Chain as a Competitive Advantage

Sathe, P. T. (2017). Fast-Tracking Friction Plate Validation Testing: BorgWarner Improves Efficiency with Machine Learning Methodology

Zettelmeyer, F. (2014). CDK Digital Marketing: Addressing Channel Conflict with Data Analytics

Marketing

Avery, J. (2018). Tailor Brands: Artificial Intelligence-Driven Branding

Music

Elberse et al. (2006). Polyphonic HMI: Mixing Music and Math

Public Sector, Government, and Defense

Hwarng, H. B., and Ran, J. (2018). Jiuzhai Valley National Park: Data-Driven Economic Growth and Ecological Preservation

Kontokosta et al. (2017). NYC311

Pan G., Chan, C. W. (2019). Accountant-General's Department: Empowering Public Sector Finance through Data Analytics in Singapore

Stettin, J., Petriglieri, G. (2018). Google and Project Maven (A): Big Tech, Government and the AI Arms Race

Stettin, J., Petriglieri, G. (2018). Google and Project Maven (B): An Eventful Week in June

Xue et al. (2019). City of London Water: Predicting Electricity Prices and Optimizing Operations

Retail

Datar, S. M., Bowler, C. N. (2018). Predicting Purchasing Behavior at PriceMart (A)

Datar, S. M., Bowler, C. N. (2018). Predicting Purchasing Behavior at PriceMart (B)

Deighton, J., Kornfeld, L. (2016). Target Stores: The Hunt for "Unvolunteered Truths"

Israeli, A., Avery, J. (2018). Predicting Consumer Tastes with Big Data at Gap

Kiran et al. (2017). Machine Learning Algorithms to Drive CRM in the Online E-Commerce Site at VMWare

Kotha, R. R., Bhattacharya, L. (2018). Data Analytics-Driven E-commerce for the Small Seller: Zilingo

Xie, R. S., Sethi, V. (2018). FlipKart VS. Amazon VS. SnapDeal: Winner-Take-All Battle in India

Zhou, K., Lau, J. (2018). Pinduoduo: 300+ Million Shoppers Teaming for Good Deals

Technology

Bhandari et al. (2015). Marketing Head's Conundrum

Hunsaker, B. T., Youngdahl, W. E. (2018). Intel: AI and Industry 4.0 Strategy

Makinen, J., Siegel. R. E. (2019). C3: Enabling Digital Industrial Transformation

Nelley, T. Keller, J. T. (2019). From Globalization to Dual Digital Transformation: CEO Thierry Breton Leading Atos Into "Digital Shockwaves" (B)

Nelley et al. (2019). From Globalization to Dual Digital Transformation: CEO Thierry Breton Leading Atos Into "Digital Shockwaves" (A)

Park, J. E. (2016). Gamaya: Taking Farming Into The 21st Century

Srujana et al. (2015). Science of Social Influence – HP BrandClout

Various

Antoni et al. (2017). Deep Technology Applications for Developing Economies: Four Vignettes

Hall, O. Ko, K. (2018). Equuleus Car Sharing Inc.: Revenue Management

Mathaisel, D. F. X. (2018). Analytics for the Sustainability of Unique Products: The Case of Hemp

Niessing et al. (2018). Jaguar Land Rover: Towards a Customer-Centric Organisation. Leveraging Customer Intelligence and Data Analytics for Sustainable Growth

Qi et al. (2018). Mobike: A Smart Bike-Sharing Service Platform

Toffel et al. (2018). Improving Worker Safety in the Era of Machine Learning (A)

Toffel et al. (2018). Improving Worker Safety in the Era of Machine Learning (B)

Appendix A: Glossary and Definitions of Terms and Acronyms[68]

Algorithm is a set of mathematical or computer-coded "instructions or rules that, especially if given to a computer, will help to calculate an answer to a problem" (https://dictionary.cambridge.org/dictionary/english/algorithm, para. 1, retrieved Apr 13, 2019).

Analytic Hierarchy Process (AHP) is a decision analysis technique aiming at assessing multi-attribute alternatives (S. Li & Li, 2009, p. 5559).

Analytics is "a process in which a computer examines information using mathematical methods in order to find useful patterns" (https://dictionary.cambridge.org/dictionary/english/analytics, para. 1, retrieved Apr 13, 2019); or "the analysis of data, typically large sets of business data, by the use of mathematics, statistics, and computer software" (https://www.dictionary.com/browse/analytics, para. 2, retrieved Apr 13, 2019).

Artificial Intelligence (AI), is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment. (OECD, 2024).

---

[68] These terms are defined within the context of AI technologies and organization strategy, and not necessarily in line with their use in unrelated conversational or written language. Various sources are cited for these definitions, including: referenced literature, industry accepted definitions, dictionaries, and online databases. Definitions which are not cited may include commonly-used industry definitions from the author's own professional experiences.

AI ⊃ ML ⊃ DL Deep learning is a subset of machine learning, which is a subset of artificial intelligence. (Goodfellow et al., 2016, as cited in Holzinger et al., 2018).

AI Technique is the utilization of one or more underlying AI technologies applied via computational system(s).

AI Technology refers to an underlying technology, such as ML, DL, etc., achieved via one or more MLAs.

Artificial General Intelligence (AGI) is non-domain-specific artificial intelligence, theoretically "capable of surpassing human capacity" (Meek et al., 2016, pp. 683, 686) and "modeled after the neural networks of the human brain" (Lauterback & Bonime-Blanc, 2016, p. 54).

Artificial Neural Network (ANN) is "a computer system or a type of computer program that is designed to copy the way in which the human brain operates" (https://dictionary.cambridge.org/dictionary/english/neural-network, para. 1, retrieved Apr 13, 2019).

Classification is "an algorithm that finds functions that help divide the dataset into classes based on various parameters. When using a Classification algorithm, a computer program gets taught on the training dataset and categorizes the data into various categories depending on what it learned. Classification algorithms find the mapping function to map the "x" input to "y" discrete output. The algorithms estimate discrete values (in other words, binary values such as 0 and 1, yes and no, true or false, based on a particular set of independent variables. To put it another, more straightforward way, classification algorithms predict an event occurrence probability by fitting data to a logit function. Classification algorithms are used for things like email and spam classification, predicting the willingness of bank customers to pay their loans, and identifying cancer tumor cells. (https://www.simplilearn.com/regression-vs-classification-in-machine-learning-article, retrieved June 1, 2024)

Types of Classification algorithms typically used in machine learning contexts, include:

- Decision Tree Classification: This type divides a dataset into segments based on particular feature variables. The divisions' threshold values are typically the mean or mode of the feature variable in question if they happen to be numerical.

- K-Nearest Neighbors: This Classification type identifies the K nearest neighbors to a given observation point. It then uses K points to evaluate the proportions of each type of target variable and predicts the target variable that has the highest ratio.

- Logistic Regression: This classification type isn't complex so it can be easily adopted with minimal training. It predicts the probability of Y being associated with the X input variable.

- Naïve Bayes: This classifier is one of the most effective yet simplest algorithms. It's based on Bayes' theorem, which describes how event probability is evaluated based on the previous knowledge of conditions that could be related to the event.

- Random Forest Classification: Random forest processes many decision trees, each one predicting a value for target variable probability. You then arrive at the final output by averaging the probabilities.

- Support Vector Machines: This algorithm employs support vector classifiers with an exciting change, making it ideal for evaluating non-linear decision boundaries. This process is possible by enlarging feature variable space by employing special functions known as kernels.

Cloud or *Cloud Computing* is the use of services, computer programs, etc. that are on the internet rather than ones that you buy and put on your computer (https://dictionary.cambridge.org/dictionary/english/cloud-computing, para. 1, retrieved Apr 13, 2019).

Competitive Advantage is a term used to describe abilities gain through attributes and/or resources to perform at a higher level than others in the same industry, market, or competitive space. It is achieved, gained, or improved through generic strategies in one or more of cost leadership, differentiation, and focus (Porter, 1985).

Decision Support System (DSS) is "a computer program that can arrange and sort large amounts of data, and that is used to help people in companies and organizations make important decisions based on the data" (https://dictionary.cambridge.org/dictionary/english/decision-support-system?q=Decision+Support+System, para. 1, retrieved Apr 13, 2019).

Deep Computing or *deep learning* is a branch within and "type of artificial intelligence that uses algorithms based on the way the human brain operates" (https://dictionary.cambridge.org/dictionary/english/deep-learning, para. 2, retrieved Apr 13, 2019). It is a type of machine learning.

Deep Neural Network is an artificial neural network with usually 3 or more layers: input, hidden and output (https://www.sciencedirect.com/topics/engineering/three-layered-neural-network, retrieved Apr 13, 2019).

Distributed Decision Support System (DDSS) is a multi-layer DSS distributed within systems at different locations or different hierarchical levels within an architecture.

Distributed Intelligence (DI) is a "function of strategically relevant human and social capital assets—the networked intellectual capabilities of human agents" (McKelvey, 2000, p. 1).

Distributed Strategic Decision Support System (DSDSS) is a "general framework which integrates advances in distributed decision making and distributed artificial intelligence" (Pinson et al., 1997, p. 35).

Expert System (ES) is "a computer system that asks questions and gives answers that have been thought of by a human expert" and has been developed for a specific application and expected outcome, such as inventory management within a factory floor (https://dictionary.cambridge.org/dictionary/english/expert-system, para. 1, retrieved Apr 13, 2019; O'Leary, n.d.).

Explainability in the context of AI, is "about verification, or providing justifications for the model's outputs, often after it makes its predictions. Explainable AI (XAI) is used to identify the factors that led to the results. Various explainability methods can be used to present the models in ways that make their complex processes and underlying data science clear to a human being using natural language."

(https://www.ibm.com/think/topics/explainable-ai, https://www.ibm.com/think/topics/interpretability; all retrieved Dec 11, 2024).

Fuzzy Logic is computational logic and operators, e.g. addition, subtraction, comparison, sequencing, applied to fuzzy sets to gain approximate reasoning (Zadeh, 1975).

Fuzzy Set is a set of data or information which have a non-discrete definition or comparators, such that discrete values cannot be assigned to them for computation (Zadeh, 1965).

GFlowNetworks or GflowNets (GFNs) is an algorithmic architecture to model probabilistic strategies, helping the system explore a variety of strategic options and providing insight into their potential future outcomes (Bengio et al., 2021; Bengio et al., 2023).

Generative Adversarial Network (GAN) is "a type of generative AI model that utilizes two neural networks in a unique and adversarial way to generate new data that resembles the training data. Some highly technical use cases, such as modeling probabilistic distributions or sampling from an arbitrary distribution, may be better suited for other types of generative AI models like Variational Autoencoders (VAEs) or Generative Stochastic Networks (GSNs). However, most of the popular generative AI applications under use are performed by GAN. (https://research.aimultiple.com/gan-use-cases/, retrieved Sep 7, 2024). GANS can simulate market events, generate synthetic competitors, or create diverse future scenarios to inform decision-making.

Hamming distance is the distance between two strings of equal length is the number of positions at which the corresponding symbols are different. In other words, it measures the minimum number of *substitutions* required to change one string into the other, or the minimum number of *errors* that could have transformed one string into the other. In a more general context, the Hamming distance is one of several string metrics for measuring the edit distance between two sequences. The Hamming distance is often used in machine learning when comparing different strings or binary vectors. More specifically with binary strings, the Hamming distance is a metric for comparing two binary data strings. While comparing two binary strings of equal length, Hamming distance is the number of bit positions in which the two bits are different. (https://en.wikipedia.org/wiki/Hamming_distance, retrieved Mar 18, 2022; https://datagy.io/python-hamming-distance/, retrieved Mar 18, 2022; https://www.tutorialspoint.com/what-is-hamming-distance, retrieved Mar 18, 2022.)

Hyperlearning is a process whereby machines learn in "virtualized (at least partially) environments with neither human training nor data input from the real world", allowing "systems to learn at machine speed and develop novel solutions in specific settings, frequently involving unsupervised learning and reinforcement learning algorithms" (Gerbert, 2018).

Intellectual Property Analytics "is the data science of analysing large amount of intellectual property information, to discover relationships, trends and patterns in the data for decision making. It is a multidisciplinary approach that makes use of mathematics, statistics, computer programming, and operations research to gain valuable knowledge from data, to support decision making rooted in the business context" (Aristodemou & Tietze, 2018, p. 38).

Interpretability in the context of AI "focuses on understanding the inner workings of an AI model while AI explainability aims to provide reasons for the model's outputs. Interpretability is about transparency, allowing users to comprehend the model's architecture, the features it uses and how it combines them to deliver predictions. An interpretable model's decision-making processes are easily understood by humans. Greater interpretability requires greater disclosure of its internal operations." (https://www.ibm.com/think/topics/interpretability, retrieved Dec 11, 2024)

Intractable refers to a type of problem that can be solved in theory (e.g. given large but finite resources, especially time), but for which in practice any solution takes too many resources to be useful (en.wikipedia.org/wiki/Computational_complexity_theory, retrieved Feb 22, 2025).

Language Models (LMs) are the general terms used to refer to deep-learning based language models of varying size, e.g. large, small, etc. LMs are critical in generating insights, processing unstructured data, and summarizing vast amounts of information.

Large Language Models (LLMs) are a deep-learning models where MLAs are used to train the model on data such as but not limited to, language, text, prose, media, mathematics, software code, and hardware design. This pre-training data then gets transformed and stored into constructs like vector database(s) or similar, for rapid accessibility in conversational or other interaction use-cases. The models are designed to *sense-make* in that the sequence of the elements in any generated response appear to make sense to the user. They are not truth-engines nor OOTB designed to optimize the dimension of accuracy. Other techniques like RAG can be *added* to improve dimensions such as accuracy.

Machine Learning (ML) is “the process by which a computer is able to improve its own performance (as in analyzing image files) by continuously incorporating new data into an existing statistical model” (https://www.merriam-webster.com/dictionary/machine%20learning, para. 1, retrieved Apr 13, 2019) to perform functions such as classify, predict, recommend, etc.

Machine Learning Algorithm (MLA) is an algorithm designed to produce the effect of ML.

Mixture of Experts (MoE) is a transformer architecture “transformer architecture allows increasing the model’s size and in effect the output quality and inference speed while keeping the compute cost fixed. Instead of dense transformer layers, an MoE uses sparse Feedforward Neural Network (FFN) layers (aka experts) and a gate network (router) to determine which tokens are sent to each expert, usually, top-k experts are selected. Mixture of Experts can scale to larger sizes for better inference quality without significantly raising the compute costs. This is because only a subset of experts (top-k) are activated for each token regardless of the total number of experts. Also, since MoEs have a limited number of active parameters during inference when compared to a dense transformer which activates all parameters, it’s relatively more computationally economical.” (https://wandb.ai/zaiinn440/hybridMoe/reports/MoE-vs-Dense-vs-Hybrid-LLM-architectures--Vmlldzo3NzYwNzAw, retrieved Jan 4, 2025).

Nondeterministic-Polynomial Time functions are a set of functions (y) of time (x) of at minimum, exponential $O(y^x)$, and at most, undefined $O(\infty)$ time to solve.

NP or NP-Class Problems are a class of 'nondeterministic polynomial-time problems, having a nondeterministic polynomial-time-hard problem, having solution algorithms that require a number of operations that is a greater than any polynomial (e.g., exponential) function of the problem's variables and constraints.' (Moldoveanu, 2009, p. 737)

NP-Complete Problem is 'a class of problem where the solution cannot be found in polynomial time, but the solution can be verified in polynomial time'.

NP-Hard Problem is 'a class of problem where neither the solution nor solution-verification can be completed in polynomial time'; 'a nondeterministic polynomial-time-hard problem, having solution algorithms that require a number of operations that is a greater than any polynomial (e.g., exponential) function of the problem's variables and constraints.' (Moldoveanu, 2009, p. 737).

Objective function can vary, but in all cases, it attempts to achieve a desired outcome or end-goal, e.g. maximize a gain *non-exclusive or* minimize a loss—either or both conditions being met may satisfy the objective—based on a set of constraints and the relationship between one or more decision variables.

Optimization is a general term to refer to maximizations, minimizations, or other desired outcomes of most organization problems or goals. Forms of optimization include:

Stochastic: is a method used to optimize systems that have random variables. These variables can be modeled as probability distributions. The goal of stochastic optimization is to find the best solution that maximizes the expected value of the objective function. This method is based on probability theory, and it is often used in decision making problems, such as portfolio management and operations research.

Robust: is a method used to optimize systems that have uncertain or unknown parameters. Based on optimization and robust control theory, the goal of robust optimization is to find the best solution that is robust to the uncertainty. It is often used in engineering and management problems, such as supply chain management and transportation. In summary, stochastic optimization is used to optimize systems with random variables, whereas robust optimization is used to optimize systems with uncertain or unknown parameters.

P, *P-Class* or *P-Hard Problems* are a class of problems that can be solved and their solution can be verified in polynomial time, and 'polynomial-time hard problems, have solution algorithms that require a number of operations that is at most a polynomial function of the number of independent variables and constraints' (Moldoveanu, 2009, p. 737).

Polynomial Time functions are a set of functions (y) of time (x) of at minimum, a constant O(1), and at most, a polynomial O(x) time to solve.

Probabilistic-Based Inferential Systems answer problems of classification or other optimizations using probabilities and self-refining rules based on existing outputs and training data. Examples include machine learning, artificial neural networks, and deep learning systems.

Production frontier is the mapping from input(s) to output(s). From economics, the technological relation between quantities of physical inputs and quantities of output of goods. Used to define marginal product and to distinguish allocative efficiency. One important purpose of the production function is to address allocative efficiency in the use of factor inputs in production and a desired output or outcome.

Python is a programming language and is well suited for coding many things due to its ease of learning and flexibility, as well as capabilities.

Reinforcement Learning (RL) is an algorithmic approach to dynamically adjust a computed solution as real-world feedback comes in, optimizing the organization's decisioning over time.

Retrieval Augmented Generation (RAG) is "a technique for enhancing the accuracy and reliability of generative AI models with facts fetched from external sources" or other corpi of data. (https://blogs.nvidia.com/blog/what-is-retrieval-augmented-generation/, retrieved Sep 7, 2024), originally coined by Lewis et al. in 2020. RAG utilizes retrieval of knowledge from internal and external databases, making sure the system bases decisions on the latest and most relevant information.

Regression is mathematical techniques that "finds correlations between dependent and independent variables. Therefore, regression algorithms help predict continuous variables such as house prices, market trends, weather patterns, oil and gas prices (a critical task these days!), etc. The Regression algorithm's task is finding the mapping function so we can map the input variable of 'x' to the continuous output variable of 'y'."

(https://www.simplilearn.com/regression-vs-classification-in-machine-learning-article, retrieved June 1, 2024)

Types of Regression algorithms typically used in machine learning contexts, include[69]:

- Linear Regression: This type is the least complicated form of regression, where the dependent variable is continuous.
- Decision Tree Regression: The primary purpose of this regression is to divide the dataset into smaller subsets. These subsets are created to plot the value of any data point connecting to the problem statement.
- Principal Components Regression: This regression technique is widely used. There are many independent variables, or multicollinearity exists in your data.
- Polynomial Regression: This type fits a non-linear equation by using the polynomial functions of an independent variable.
- Random Forest Regression: Random Forest regression is heavily used in Machine Learning. It uses multiple decision trees to predict the output. Random data points are chosen from the given dataset and used to build a decision tree via this algorithm.
- Support Vector Regression: This regression type solves both linear and non-linear models. It uses non-linear kernel functions, like polynomials, to find an optimal solution for non-linear models.

[69] https://arunp77.medium.com/regression-algorithms-29f112797724, retrieved Jun 1, 2024.

Reliable means the extent to which *something* is stable and consistent, i.e. a consistent result may be deemed to be reliable.

Rule-Based Deterministic Systems are expected to yield same output for the same input—based on a set of programmed rules. Examples include fuzzy logic, expert systems.

Stochastic processes are random processes, "widely used in mathematical models of systems or phenomena that appear to vary in a random manner" Examples include the Bernoulli process of a sequence of independent and identically distributed (iid) random variables of value 0 or 1, and random walks defined as sums of iid variables or vectors in some state space (https://en.wikipedia.org/wiki/Stochastic_process, retrieved Nov 16, 2024).

Stochastic Optimization is a method of generating and using random variables to represent an optimization problem to produce more suitable and consistent results. Stochastic optimization often represents real-world problems more accurately by introducing some uncertainty into the problem definition or result, reflecting the variability of inputs to and/or outputs from the optimization process. (https://c3.ai/glossary/artificial-intelligence/stochastic-optimization/, retrieved Oct 24, 2021)

Strategic is a prefix adjective to describe a verb or action in support or furtherance of a strategy, e.g. strategic decision-making.

Strategy Formulation or Development is an inward and outward examination via approaches, processes and methods to develop, define, or articulate a Strategy.

Support-Vector Clustering (SVC) is a non-supervised machine learning through clustering of unlabelled data into natural groups using SVM statistical models, proceeded by mapping new data into the formed groups. (https://en.wikipedia.org/wiki/Support-vector_machine, retrieved Jul 22, 2022)

Support-Vector Machines (SVMs) are supervised machine learning models and algorithms that analyze data and perform classification for regression analysis, by first classifying training examples into one of two categories, thereby efficiently and robustly mapping new data into a category and expanding the width of the gap between the two categories. SVMs are typically non-probabilistic but can be probabilistic, using Platt Scaling to transform the classification model outputs into a probability distribution over classes. (https://en.wikipedia.org/wiki/Support-vector_machine, retrieved Jul 22, 2022; https://en.wikipedia.org/wiki/Platt_scaling, retrieved Jul 22, 2022)

Tractable refers to a type of problem that can be solved in practice ().

Transparency "means that an AI model is understandable by itself" (https://www.ericsson.com/en/reports-and-papers/white-papers/explainable-ai--how-humans-can-trust-ai, retrieved Oct 24, 2021)

Understandability "means that AI model is able to make humans understand its functionalities." (https://www.ericsson.com/en/reports-and-papers/white-papers/explainable-ai--how-humans-can-trust-ai, retrieved Oct 24, 2021)

Valid means the extent to which *something* can be generalized to other cases/settings, i.e. the deductive application of a theory to multiple empirical subjects may be considered valid.

Appendix B: Problem Complexity Classes

This is a brief introduction to complexity concepts, for readers who may be unfamiliar with the meaning and inter-relationships of P, NP, NP-Complete, and NP-Hard. Given the spectrum of complexity, measures of time-to-solve, and time-to-verify, contained within both P-class and NP-class problems, this introduction highlights the differences of the most complex scenarios of each, but this need not be the case.

Polynomial Time or P-class problems have solutions or solution sets which at most grow in polynomial time as the problem variables increase, e.g. arithmetic, sorting or generalized Checkers (a defined size such as an 8x8 board). Karp (1972) builds on Cook (1971), and proves that P-class problems are tractable and solvable in at most, polynomial time. I posit that measures of complexity and time-to-solve are important measures in computational systems as they go directly to considerations such as system design and resources such as processing chips, memory, network, and architectures.

Nondeterministic Polynomial Time or NP-class problems have solutions or solution sets which grow at greater than polynomial time, e.g. exponentially or higher, as the problem variables increase. Examples include non-generalized Sudoku or non-generalized Chess (i.e. a non-defined or NxN board) (Wigderson, 2006). NP-class problems are further classified into, in order of increasing complexity and time-to-solve/verify: NP-complete, and NP-Hard. NP-complete means that the solution can be verified in P as the problem complexity increases, but cannot be solved in P. Examples of this are generalized NxN Sudoku, or the Travelling Salesman Problem (TSP) Search (Karp, 1972). NP-hard problems are at least as hard as NP-complete, but neither the solution nor its verification can be completed in P, e.g. protein structures, generalized

NxN Chess, or TSP Optimization (TSP search optimized for the shortest total distance) (Arora, 2003; Tannenbaum, 2014). Synthesizing interrelationships contained in Goldreich (2010) and Tannenbaum, 2014, all NP-Complete problems are both NP-Hard and in NP, but not all NP-Hard problems are NP-Complete, meaning that they may not have solutions that can be verified in polynomial (P) time. Lastly, NP-Hard problems may be outside the class of problems that have a solution, or require a decision, i.e. they are not *decision problems*[70]. Below is a summary table of the different complexity classes I've introduced, with some algorithm examples.

If P = NP[71]:

$$\{P = NP = NP\_Complete\} \subset NP\_Hard$$

If P ≠ NP[1]:

$$P \subset NP$$
$$NP\_Complete \subset NP$$
$$P \neq NP$$
$$NP \subset NP\_Hard$$

Problems of P-class can be:	Constant time	$O(1)$
	Logarithmic time	$O(\log_2(x))$
	Linear time	$O(x)$
	Quadratic time	$O(x^2)$
	Polynomial time	$O(x^y)$
Problems of NP-class can be:	Exponential time	$O(y^x)$
	Factorial time	$O(x!)$
	Undefined time	$O(\infty)$

[70] https://en.wikipedia.org/wiki/NP-completeness, retrieved Feb 22, 2025.

[71] The analysis of whether P=NP or P≠NP is beyond the scope of this study, but for my purposes here, I will adopt the widely supported belief that P≠NP. relying on P≠NP is helpful for an analysis of problem complexity, and computational resources applied to same. This research study does not turn on adopting this widely supported notion.

Appendix C: Brief Survey of Generative AI Techniques

While the terms and technologies associated with 'Chatbot' and 'GenAI' are related and often intermingled, there are some important fundamental differences between them[72], in both approach-to-produce and underlying architecture and MLAs. In short, the former is largely based on training data to make sense via a conversational channel such as a prompt, while the latter is typically the former augmented with other deep-learning approaches and/or externally-sourced information[73]. A non-exhaustive summary of different generative capabilities, and underlying computational models and approaches is provided below in Table 25. I've intentionally left out names of any products or tools, such as ChatGPT because the field and offerings are moving too fast to be of value in this context.

Table 25

Non-Exhaustive Summary of ML/AI-based Products and Tools categorized by Industry use and Capabilities[74]

Industry	Capabilities	Computational Models/Approaches
Business and Commerce	Chatbots	Foundation models
	T**ext/**Prose generation	Machine learning algorithms
	Text and content synthesis	Detection, Classification, Segmentation, Prediction
	Trend Analysis	
		AI assistants

[72] https://research.aimultiple.com/conversational-ai-vs-generative-ai/, retrieved Sep 7, 2024.

[73] https://aimultiple.com/hub/generative-ai#market-map, retrieved Sep 7, 2024.

[74] Integrated and expanded from https://www.fastcompany.com/90856183/30-ai-tools-you-can-try-for-free (retrieved Dec 22, 2023), https://research.aimultiple.com/generative-ai-applications/ (retrieved Dec 22, 2023), and https://www.upwork.com/resources/generative-ai-applications (retrieved Dec 22, 2023).

Industry	Capabilities	Computational Models/Approaches
	Sentiment analysis/text classification	LMs
	Fraud detection	GANs
	Risk management	AI agents
	Customer service communications	Sub-tasking
Healthcare	Medical imaging	Detection, Classification, Segmentation, Prediction
	Drug/molecule discovery	Rules-based inferential models
	Data Synthesis	
	Patient communication	Deep Learning
	Drug manufacturing	LMs
		Robotics
Software Development	Code generation	Foundation Models
	Code debugging	LLMs
Hardware Development	Chip/Semiconductor design	Deep Reinforcement Learning[75]
	Circuit optimization	RAG[17]
Education	Course design and personalization	LLMs
	Content creation	
	Tutoring	
Manufacturing, Supply Chain and Logistics	Production planning	ES
	Inventory management	

[75] Roy et al. (2021); https://developer.nvidia.com/blog/designing-arithmetic-circuits-with-deep-reinforcement-learning/, retrieved Sep 7, 2024.

Industry	**Capabilities**	**Computational Models/Approaches**
	Quality control	Detection, Classification, Segmentation, Prediction
	Predictive maintenance	Rules-based inferential models
		Foundation models
Sound, Art and Visual	Video generation	GAN
	Image generation	Trained Conditional-GANs
	Image-to-Photo Translation	FAE-GAN
	Image-to-Image Conversion	SP-GAN
	3D Shape Generation	
	Text to still image	
	Text to video	

Delving deeper into these generative concepts, key developments that enhance the abilities of LLM-based tools and technologies to cross into the realm of generative tools, is through approaches such as retrieval-augmented generation (RAG[76,77]) and generative adversarial networks (GANs). RAG in-effect takes the responses from a priori *or posteriori* trained LLMs, and grounds or augments these with vast corpuses of data to validate or verify the responses, which collectively become embedded with the LLM-generated data in the vector database (or other storage construct). In simple terms, if none exist there is increased likelihood

[76] https://research.ibm.com/blog/retrieval-augmented-generation-RAG, retrieved Aug 24, 2024.

[77] https://cloud.google.com/use-cases/retrieval-augmented-generation, retrieved Aug 24, 2024.

that the LLM-generated response may be incorrect, while the existence of external validators does the opposite and increases the relevance and quality of generated/outputted content. This form of augmentation has value on its own, but can also be coupled with other techniques such as deep learning. This is where GAN approaches come in, which takes generated data, such as outputs of LLM-based systems, and feeds this information *back into* the system as new training data for two neural networks: one generator creating data and a discriminator to evaluate its authenticity. The concept of adversarial comes into the picture via opposing DL models that in-effect compete to find the more correct answer to the query—the sample based on training data *or* the generated sample—the overall system self-learning and self-refining by cycling through this iterative approach.

Appendix D: Python Functions and [Quasi-]Code

Cubic Production Function with Single Input and Output Variables as a Trade-off

```
import numpy as npy
import matplotlib.pyplot as plt

def cubic_production_function(input_variable, a=1,
b=0, c=0, d=0):
    """
    Cubic production function.

    Parameters:
    - input_variable: Input variable
    - a, b, c, d: Coefficients of the cubic function
(default values create a simple cubic function)

    Returns:
    - output_variable: Output variable (product)
    """

    output_variable = a * input_variable**3 + b *
input_variable**2 + c * input_variable + d
    return output_variable

# Example usage with coefficients:
input_values_cubic = npy.linspace(-5, 5, 100)
a_coefficient = 1
b_coefficient = -2
c_coefficient = 0
d_coefficient = 5

outputs_cubic = [cubic_production_function(I,
a_coefficient, b_coefficient, c_coefficient, d_coefficient)
for I in input_values_cubic]

# Plotting the cubic production function
plt.plot(input_values_cubic, outputs_cubic)
plt.title('Cubic Production Function')
plt.xlabel('Input Variable')
plt.ylabel('Output Variable')
plt.show()
```

Basic Positive Concave Down Function

```
import numpy as npy
import matplotlib.pyplot as plt

# Define values for e and i
e_values = npy.linspace(0, 1, 100)
I_values = npy.arange(1, 10)

# Plot the equation for different values of i
plt.figure(figsize=(10, 6))
for I in I_values:
    O_values = -(1 - e_values) ** I
    plt.plot(e_values, O_values, label=f'I = {I}')

plt.title('Accuracy vs. Privacy Loss')
plt.xlabel(' Privacy Loss (e)')
plt.ylabel('Accuracy (I)')
plt.grid(True)
plt.show()
```

Provision for Enhancement via M(*e,I*) ML Options and N(*e.I*) Computational Resources

```
import numpy as npy
import matplotlib.pyplot as plt

e_values = npy.linspace(0, 1, 100)
I_values = npy.arange(1, 10)

# Define functions for M(e, I) and N(e, I)
def M(e, I):
    # Example function for machine learning options
    return 1.0  # Placeholder, replace with actual
function

def N(e, I):
    # Example function for computational resources
    return 1.0  # Placeholder, replace with actual
function

plt.figure(figsize=(10, 6))
for I in I_values:
```

```
        O_values = (1 – (1 - e_values) ** I) * M(e_values,
I) * N(e_values, I)
     plt.plot(e_values, O_values, label=f'I = {I}')
     plt.title('Accuracy vs. Privacy Loss')
     plt.xlabel(' Privacy Loss (e)')
     plt.ylabel('Accuracy (I)')
     plt.legend()
     plt.grid(True)
     plt.show()
```

Inflection Points

```
   import numpy as npy
   import matplotlib.pyplot as plt
   from scipy.misc import derivative

   # Define values for e and set I to a fixed value
   e_values = npy.linspace(0, 1, 100)
   I_values = npy.arange(1, 11)

   # Define functions for M(e, I) and N(e, I)
   def M(e, I):
       return 1.0

   def N(e, I):
       return 1.0

   # Function to calculate the second derivative of O(e, I)
   def second_derivative_O(e, I):
       def O(e):
           return (1 - (1 - e) ** I) * M(e, I) * N(e, I)
       return derivative(O, e, n=2)

   # Compute the location of the inflection point for each
value of i
   inflection_points = []
   for I in I_values:
       e_inflection = None
       for e in e_values:
           if second_derivative_O(e, I) *
second_derivative_O(e + 0.01, I) < 0:
               e_inflection = e
               break
       inflection_points.append(e_inflection)
```

```
   # Plot the risk of identification vs. probability of
error
   plt.figure(figsize=(10, 6))
   for I in I_values:
       O_values = (1 - (1 - e_values) ** I) * M(e_values,
I) * N(e_values, I)
       plt.plot(e_values, O_values, label=f'I = {I}')

   # Plot the location of the inflection point vs. Accuracy
(I)
   inflection_points_filtered = [inflection_point for
inflection_point in inflection_points if inflection_point
is not None]
   plt.plot(inflection_points_filtered, [(1 - (1 - e) **
10) * M(e, 10) * N(e, 10) for e in
inflection_points_filtered], 'ro', label='Inflection
Point')

   plt.title('Accuracy vs. Privacy Loss')
   plt.xlabel(' Privacy Loss (e)')
   plt.ylabel('Accuracy (I)')
   plt.legend(title='Legend')
   plt.grid(True)
   plt.show()
```

Appendix E: Analysis of AI Case Studies (Harvard Business School Library, 2020)

Table 26

Categorization of AI Case Studies by Industry/Function, ML Problem, Computational Resources, and Set of Seven Trade-off Dimensions

			INDEPENDENT DIMENSIONS	OBJECTIVE FUNCTION DIMENSIONS (ALGORITHM) (↑=Gain, ↓=Loss, ↔=Same, •=Relevant)						
Paper/Case	**Industry/ Function**	**ML Problem(s)**	**Computational Resources (CPU-Memory-Speed In/Out)**	**Accuracy**	**Reliability**	**Security**	**Fairness**	**Privacy**	**Explainability**	**Speed**
1. Tiwari et al. (2018). Analytics Empowering Agriculture: Jayalaxmi Agro Tech	Agriculture	Prediction	**CPU** ↔ **Memory** ↔ **Speed** ↑	↑	↑	↓	↑	↓		↑
2. Pah et al. (2018). Evaluating the Cognitive Analytics Frontier	Aviation	Classification	**CPU** ↔ **Memory** ↔ **Speed** ↑	↑	↑	↔↑			↑	↑
3. Youngdahl (2017) JetBlue and Gladly: Onmichannel Customer Service	Aviation	Prediction	**CPU** ↓ **Memory** ↓ **Speed** ↑	↑			↑			↑
4. Makinen and Burgelman. (2018). Noodle AI	Aviation Consulting	Prediction Classification Optimization	**CPU** ↓ **Memory** ↓ **Speed** ↑	↑	↑				↑	↑
5. Meng and Zhu (2019). A Chairman's Decision: Launching a Robo-Advisor in CCB Principal Asset Management Company	Banking & Financial Services	Prediction Optimization	**CPU** **Memory** ↓ **Speed**	↑		↑				
6. Cohen et al. (2018). Artificial Intelligence and the Machine Learning Revolution in Finance	Banking & Financial Services	Prediction Classification	**CPU** **Memory** **Speed**	↑						↑
7. Hou et al. (2016). Alibaba's Growth Frenzy	Banking & Financial Services	N/A	**CPU** **Memory** **Speed**			↑				
8. Khachatryan, D. (2014). 9. Armacord Incorporated:	Banking & Financial Services	Classification	**CPU** ↔ **Memory** ↔			↑				↑

			INDEPENDENT DIMENSIONS	**OBJECTIVE FUNCTION DIMENSIONS (ALGORITHM)** (↑=Gain, ↓=Loss, ↔=Same, •=Relevant)						
Paper/Case	**Industry/ Function**	**ML Problem(s)**	**Computational Resources (CPU-Memory-Speed In/Out)**	**Accuracy**	**Reliability**	**Security**	**Fairness**	**Privacy**	**Explainability**	**Speed**
Combatting Money-laundering Using Data Analytics			**Speed** ↑							
10. Wendell, C. (2016). Bringing Quick Loans to the Unbankable (A)	Banking & Financial Services	Prediction	**CPU** ↓ **Memory** ↓ **Speed** ↑	↑						
11. Wendell, C. (2016). Bringing Quick Loans to the Unbankable (A)	Banking & Financial Services	Prediction	**CPU** ↓ **Memory** ↓ **Speed** ↑	↑						
12. Johnston and Peterson (2018). Commercial International Bank	Banking & Financial Services	Prediction Classification	**CPU** ↓ **Memory** ↓ **Speed** ↑							
13. Liang and Zheng (2016). Data Analytics at DBS Group Audit	Banking & Financial Services	Prediction	**CPU** ↔ **Memory** ↔ **Speed** ↑	↑						↑
14. Datar and Bowler (2018). Lending Club (A)	Banking & Financial Services	Prediction	**CPU** ↓ **Memory** ↓ **Speed** ↔	↑						↑
15. Datar and Bowler (2018). Lending Club (B)	Banking & Financial Services	Prediction	**CPU** ↓ **Memory** ↓ **Speed** ↔	↑						↑
16. Datar and Bowler (2018). Lending Club (C)	Banking & Financial Services	Prediction	**CPU** ↓ **Memory** ↓ **Speed** ↔	↑						↑
17. Wang and Thomas (2018). New Constructs: Disrupting Fundamental Analysis with Robo-Analysts	Banking & Financial Services	Classification	**CPU** ↔ **Memory** ↔ **Speed** ↑	↑						↑
18. Kumar et al. (2016). Predicting Earnings Manipulation by Indian Firms	Banking & Financial Services	Prediction	**CPU** **Memory** **Speed**	↑						
19. Ovchinnikov (2018). Private Banking Advisers at BCB Edmonton (A)	Banking & Financial Services	Optimization	**CPU** ↓ **Memory** ↓ **Speed** ↔							

			INDEPENDENT DIMENSIONS	OBJECTIVE FUNCTION DIMENSIONS (ALGORITHM) (↑=Gain, ↓=Loss, ↔=Same, ●=Relevant)						
Paper/Case	**Industry/ Function**	**ML Problem(s)**	**Computational Resources (CPU-Memory-Speed In/Out)**	**Accuracy**	**Reliability**	**Security**	**Fairness**	**Privacy**	**Explainability**	**Speed**
20. Ovchinnikov (2018). Private Banking Advisers at BCB Edmonton (B)	Banking & Financial Services	Optimization	**CPU** ↓ **Memory** ↓ **Speed** ↔							
21. Ovchinnikov (2018). Private Banking Advisers at BCB Edmonton (c)	Banking & Financial Services	Optimization	**CPU** ↓ **Memory** ↓ **Speed** ↔							
22. Schifrin and Shaw (2017). RBC - Using People Strategy and Analytics to Drive Employee Performance (A)	Banking & Financial Services	Optimization	**CPU** ↔ **Memory** ↔ **Speed** ↑							
23. Schifrin and Shaw (2017). RBC - Using People Strategy and Analytics to Drive Employee Performance (B)	Banking & Financial Services	Optimization	**CPU** ↔ **Memory** ↔ **Speed** ↑							
24. Chandrasekhar (2017). RBC Social Network Analysis	Banking & Financial Services	Prediction	**CPU** ↔ **Memory** ↔ **Speed** ↑	↑		↑				↓
25. Niessing and Plassmann (2017). Telenor - Revolutionizing Retail Banking in Serbia	Banking & Financial Services	Prediction Classification	**CPU** ↑ **Memory** ↑ **Speed** ↔			●				
26. Kar (2015). Growing the Online Portal of Business Fundas	Communications		**CPU** **Memory** **Speed**							
27. Kerr and Moloney (2018). Vodafone - Managing Advanced Technologies & Artificial Intelligence	Communications		**CPU** **Memory** **Speed**							
28. Avery and Steenburgh (2018). Hubspot & Motion AI	Customer Service		**CPU** **Memory** **Speed**							
29. Seth and Kumar (2017). Improving Customer Engagement Through Analytics	Customer Service		**CPU** **Memory** **Speed**							

			INDEPENDENT DIMENSIONS	OBJECTIVE FUNCTION DIMENSIONS (ALGORITHM) (↑=Gain, ↓=Loss, ↔=Same, •=Relevant)						
Paper/Case	Industry/ Function	ML Problem(s)	Computational Resources (CPU-Memory-Speed In/Out)	Accuracy	Reliability	Security	Fairness	Privacy	Explainability	Speed
30. Brandwein and Niessing (2019). WeWork - Service Excellence through Business Model Innovation	Customer Service		**CPU Memory Speed**							
31. Sawhney and Goodman (2018). VMock - Pivoting to Succeed and Scale	Education		**CPU Memory Speed**							
32. Bussgang and Kelley (2018). Choosy	Fashion		**CPU Memory Speed**							
33. Amraham et al. (2016). Customer Analytics at Bigbasket – Product Recommendations	Food and Nutrition		**CPU Memory Speed**							
34. Israeli A., Lane D., (2019). DayTwo: Going to market with Gut Microbiome	Food and Nutrition		**CPU Memory Speed**							
35. Vimala et al. (2018). Nestlé: Developing a Digital Nutrition Platform for Japan	Food and Nutrition		**CPU Memory Speed**							
36. Pan et al. (2019). Seng Hua Hng Foodstuffs: Cracking the Camel Nit for Growth, Using Data Analytics	Food and Nutrition		**CPU Memory Speed**							
37. Kerr et al. (2019). Autonomous Vehicles: The Rubber Hits the Road...but When?	General		**CPU Memory Speed**							
38. Mohasseb et al. (2019). Two Brothers, Two Methods: "Happiness Index" vs. "Data & Analytics"	General		**CPU Memory Speed**							

			INDEPENDENT DIMENSIONS	OBJECTIVE FUNCTION DIMENSIONS (ALGORITHM) (↑=Gain, ↓=Loss, ↔=Same, •=Relevant)						
Paper/Case	Industry/ Function	ML Problem(s)	Computational Resources (CPU-Memory-Speed In/Out)	Accuracy	Reliability	Security	Fairness	Privacy	Explainability	Speed
39. Applegate et al. (2017). Podium Data: Harnessing the Power of Big Data Analytics	General		CPU Memory Speed							
40. Ovchinnikov, A. (2018). Retention Modeling at Scholastic Travel Company (A)	General		CPU Memory Speed							
41. Ovchinnikov, A. (2018). Retention Modeling at Scholastic Travel Company (B)	General		CPU Memory Speed							
42. Quelch, J. A., Rodriguez, M. L. (2015). Carolinas HealthCare System: Consumer Analytics	Health Services		CPU Memory Speed							
43. Ma et al. (2016). Data Analytics at Alexandra Health System: A New Journey in the Healthcare Industry	Health Services		CPU Memory Speed							
44. Aggarwal, R. (2017). PatientsLikeMe: Using Social Network Health Data to Improve Patient Care	Health Services		CPU Memory Speed							
45. Rahul et al. (2017). Predicting Net Promotor Score (NOS) to Improve Patient Experience at Manipal Hospitals	Health Services		CPU Memory Speed							
46. Greenstein S., Gulick S. (2018). Zebra Medical Vision	Health Services		CPU Memory Speed							
47. Walker R. (2019). Alexa: A Pandora's Box of Risks	Home Services		CPU Memory Speed							

			INDEPENDENT DIMENSIONS	OBJECTIVE FUNCTION DIMENSIONS (ALGORITHM) (↑=Gain, ↓=Loss, ↔=Same, •=Relevant)						
Paper/Case	**Industry/ Function**	**ML Problem(s)**	**Computational Resources (CPU-Memory-Speed In/Out)**	**Accuracy**	**Reliability**	**Security**	**Fairness**	**Privacy**	**Explainability**	**Speed**
48. Lal, R. Johnson, S. (2018). Amazon, Google, and Apple: Smart Speakers and the Battle for the Connected Home	Home Services		**CPU Memory Speed**							
49. Cui et al. (2018). Dami Technology Co. Ltd.: A Smart Kitchen Ecosystem	Home Services		**CPU Memory Speed**							
50. Yoffie et al. (2018). Voice War: Hey Google vs. Alexa vs. Siri	Home Services		**CPU Memory Speed**							
51. Mark, K. (2017). Agoda: People Analytics and Business Culture (A)	Human Resources		**CPU Memory Speed**							
52. Mark, K. (2017). Agoda: People Analytics and Business Culture (B)	Human Resources		**CPU Memory Speed**							
53. Lakhani et al. (2016). Aspiring Minds	Human Resources		**CPU Memory Speed**							
54. Joseph et al. (2018). Evie.AI: The Rise of Artificial Intelligence, and the Future of Work	Human Resources		**CPU Memory Speed**							
55. Guetta, C. D. (2019). An Introduction to AI for Text Mining: A Companion to the Evisort Case	Human Resources		**CPU Memory Speed**							
56. Guetta, C. D. (2018). Evisort: An A.I.-Powered Start-up Uses Text Mining to Become Google for Contracts	Human Resources		**CPU Memory Speed**							
57. Bernstein et al. (2017). GROW: Using Artificial Intelligence to Screen Human Intelligence	Human Resources		**CPU Memory Speed**							

			INDEPENDENT DIMENSIONS	OBJECTIVE FUNCTION DIMENSIONS (ALGORITHM) (↑=Gain, ↓=Loss, ↔=Same, •=Relevant)						
Paper/Case	Industry/ Function	ML Problem(s)	Computational Resources (CPU-Memory-Speed In/Out)	Accuracy	Reliability	Security	Fairness	Privacy	Explainability	Speed
58. Rahul, K., Kumar, U. D. (2016). HR Analytics at ScaleneWorks – Behavioral Modeling to Predict Renege	Human Resources		**CPU Memory Speed**							
59. Dutta, D., Venkatagiri, S. (2017). Edge Networks: Making Hr Intelligent	Human Resources		**CPU Memory Speed**							
60. Buche, I. (2016). Recruit Japan: Harnessing Data to Create Value (Abridged)	Human Resources		**CPU Memory Speed**							
61. Buche, I. (2016). Recruit Japan: Harnessing Data to Create Value	Human Resources		**CPU Memory Speed**							
62. Orleans, A., Leslie, M. (2019). Cerebras: A Tale of Dreams and Risks	Innovation		**CPU Memory Speed**							
63. Youngdahl, W. E., Hunsaker, B. T. (2018). Coda Cofee and bext360 Supply Chain: Machine Vision, AI, IoT, and Blockchain	Innovation		**CPU Memory Speed**							
64. Farronato et al. (2018). Innovation at Uber: The Launch of Express POOL	Innovation		**CPU Memory Speed**							
65. Davis et al. (2019). ByteDance Beyond China: Leveraging Consumer Artificial Intelligence (AI) from Toutiao to Musical.ly and TikTok	Innovation		**CPU Memory Speed**							
66. Yoffie et al. (2018). Numenta: Inventing and (or) Commercializing AI	Innovation		**CPU Memory Speed**							
67. Duke et al. (2019). Nvidia: Winning the Deep-	Innovation		**CPU Memory Speed**							

			INDEPENDENT DIMENSIONS	OBJECTIVE FUNCTION DIMENSIONS (ALGORITHM) (↑=Gain, ↓=Loss, ↔=Same, •=Relevant)						
Paper/Case	**Industry/ Function**	**ML Problem(s)**	**Computational Resources (CPU-Memory-Speed In/Out)**	**Accuracy**	**Reliability**	**Security**	**Fairness**	**Privacy**	**Explainability**	**Speed**
Learning Leadership Battle										
68. Xie et al. (2018). SenseTime: World's Most Valuable Artificial Intelligence Startup	Innovation		**CPU Memory Speed**							
69. Youngdahl, W. E., Hunsaker, B. T. (2018). SingularityNET: Blockchain-Driven AI Marketplace and Quest	Innovation		**CPU Memory Speed**							
70. Zettelmeyer, F. (2014). CDK Digital Marketing: Addressing Channel Conflict with Data Analytics	Manufacturing		**CPU Memory Speed**							
71. Chandrasekhar, R. (2017). Dow Chemical Co.: Big Data in Manufacturing	Manufacturing		**CPU Memory Speed**							
72. Sathe, P. T. (2017). Fast-Tracking Friction Plate Validation Testing: BorgWarner Improves Efficiency with Machine Learning Methodology	Manufacturing		**CPU Memory Speed**							
73. Lehman, C., Siegel, R. (2017). General Electric in 2017: Naming and Claiming the Industrial Internet	Manufacturing		**CPU Memory Speed**							
74. Markoff, M., Seifert, R. W. (2018). Tetra Pak: A Digitally enabled Supple Chain as a Competitive Advantage	Manufacturing		**CPU Memory Speed**							
75. Hall, O. P., Ko, K. (2015). Trionym Systems: Investment Decision-	Manufacturing		**CPU Memory Speed**							

			INDEPENDENT DIMENSIONS	OBJECTIVE FUNCTION DIMENSIONS (ALGORITHM) (↑=Gain, ↓=Loss, ↔=Same, •=Relevant)						
Paper/Case	**Industry/ Function**	**ML Problem(s)**	**Computational Resources (CPU-Memory-Speed In/Out)**	**Accuracy**	**Reliability**	**Security**	**Fairness**	**Privacy**	**Explainability**	**Speed**
Making Using Prescriptive Analytics										
76. Avery, J. (2018). Tailor Brands: Artificial Intelligence-Driven Branding	Marketing		**CPU Memory Speed**							
77. Elberse et al. (2006). Polyphonic HMI: Mixing Music and Math	Music		**CPU Memory Speed**							
78. Xue et al. (2019). City of London Water: Predicting Electricity Prices and Optimizing Operations	Public Sector, Government, and Defense		**CPU Memory Speed**							
79. Pan G., Chan, C. W. (2019). Accountant-General's Department: Empowering Public Sector Finance through Data Analytics in Singapore	Public Sector, Government, and Defense		**CPU Memory Speed**							
80. Stettin, J., Petriglieri, G. (2018). Google and Project Maven (A): Big Tech, Government and the AI Arma Race	Public Sector, Government, and Defense		**CPU Memory Speed**							
81. Stettin, J., Petriglieri, G. (2018). Google and Project Maven (B): An Eventful Week in June	Public Sector, Government, and Defense		**CPU Memory Speed**							
82. Hwarng, H. B., and Ran, J.. (2018). Jiuzhai Valley National Park: Data-Driven Economic Growth and Ecological Preservation	Public Sector, Government, and Defense		**CPU Memory Speed**							
83. Kontokosta et al. (2017). NYC311	Public Sector, Government, and Defense		**CPU Memory Speed**							

			INDEPENDENT DIMENSIONS	OBJECTIVE FUNCTION DIMENSIONS (ALGORITHM) (↑=Gain, ↓=Loss, ↔=Same, •=Relevant)						
Paper/Case	**Industry/ Function**	**ML Problem(s)**	**Computational Resources (CPU-Memory-Speed In/Out)**	**Accuracy**	**Reliability**	**Security**	**Fairness**	**Privacy**	**Explainability**	**Speed**
84. Kotha, R. R., Bhattacharya, L. (2018). Data Analytics-Driven E-commerce for the Small Seller: Zilingo	Retail		**CPU Memory Speed**							
85. Xie, R. S., Sethi, V. (2018). FlipKart VS. Amazon VS. SnapDeal: Winner-Take-All Battle in India	Retail		**CPU Memory Speed**							
86. Kiran et al. (2017). Machine Learning Algorithms to Drive CRM in the Online E-Commerce Site at VMWare	Retail		**CPU Memory Speed**							
87. Zhou, K., Lau, J. (2018). Pinduoduo: 300+ Million Shoppers Teaming for Good Deals	Retail		**CPU Memory Speed**							
88. Israeli, A., Avery, J. (2018). Predicting Consumer Tastes with Big Data at Gap	Retail		**CPU Memory Speed**							
89. Datar, S. M., Bowler, C. N. (2018). Predicting Purchasing Behavior at PriceMart (A)	Retail		**CPU Memory Speed**							
90. Datar, S. M., Bowler, C. N. (2018). Predicting Purchasing Behavior at PriceMart (B)	Retail		**CPU Memory Speed**							
91. Deighton, J., Kornfeld, L. (2016). Target Stores: The Hunt for "Unvolunteered Truths"	Retail		**CPU Memory Speed**							
92. Makinen, J., Siegel. R. E. (2019). C3: Enabling	Technology		**CPU Memory**							

			INDEPENDENT DIMENSIONS	OBJECTIVE FUNCTION DIMENSIONS (ALGORITHM) (↑=Gain, ↓=Loss, ↔=Same, •=Relevant)						
Paper/Case	**Industry/ Function**	**ML Problem(s)**	**Computational Resources (CPU-Memory-Speed In/Out)**	**Accuracy**	**Reliability**	**Security**	**Fairness**	**Privacy**	**Explainability**	**Speed**
Digital Industrial Transformation			**Speed**							
93. Park, J. E. (2016). Gamaya: Taking Farming Into The 21st Century	Technology		**CPU Memory Speed**							
94. Hunsaker, B. T., Youngdahl, W. E. (2018). Intel: AI and Industry 4.0 Strategy	Technology		**CPU Memory Speed**							
95. Bhandari et al. (2015). Marketing Head's Conundrum	Technology		**CPU Memory Speed**							
96. Srujana et al. (2015). Science of Social Influence – HP BrandClout	Technology		**CPU Memory Speed**							
97. Nelley, T. Keller, J. T. (2019). From Globalization to Dual Digital Transformation: CEO Thierry Breton Leading Atos Into "Digital Shockwaves" (B)	Technology		**CPU Memory Speed**							
98. Nelley et al. (2019). From Globalization to Dual Digital Transformation: CEO Thierry Breton Leading Atos Into "Digital Shockwaves" (A)	Technology		**CPU Memory Speed**							
99. Mathaisel, D. F. X. (2018). Analytics for the Sustainability of Unique Products: The Case of Hemp	Various		**CPU Memory Speed**							
100. Antoni et al. (2017). Deep Technology Applications for Developing	Various		**CPU Memory Speed**							

			INDEPENDENT DIMENSIONS	OBJECTIVE FUNCTION DIMENSIONS (ALGORITHM) (↑=Gain, ↓=Loss, ↔=Same, •=Relevant)						
Paper/Case	**Industry/ Function**	**ML Problem(s)**	**Computational Resources (CPU-Memory-Speed In/Out)**	**Accuracy**	**Reliability**	**Security**	**Fairness**	**Privacy**	**Explainability**	**Speed**
Economies: Four Vignettes										
101. Hall, O. Ko, K. (2018). Equuleus Car Sharing Inc.: Revenue Management	Various		**CPU Memory Speed**							
102. Toffel et al. (2018). Improving Worker Safety in the Era of Machine Learning (A)	Various		**CPU Memory Speed**							
103. Toffel et al. (2018). Improving Worker Safety in the Era of Machine Learning (B)	Various		**CPU Memory Speed**							
104. Niessing et al. (2018). Jaguar Land Rover: Towards a Customer-Centric Organisation. Leveraging Customer Intelligence and Data Analytics for Sustainable Growth	Various		**CPU Memory Speed**							
105. Qi et al. (2018). Mobike: A Smart Bike-Sharing Service Platform	Various		**CPU Memory Speed**							