



**Stanford – Vienna
Transatlantic Technology Law Forum**

A joint initiative of
Stanford Law School and the University of Vienna School of Law



TTLF Working Papers

No. 163

**Artificial Intelligence Systems: Typologies,
Risks, and Cross-Border Governance and
Regulatory Frameworks in Europe and the
U.S.**

Leon Anidjar

2026

TTLF Working Papers

Editors: Siegfried Fina, Mark Lemley, and Roland Vogl

About the TTLF Working Papers

TTLF's Working Paper Series presents original research on technology, and business-related law and policy issues of the European Union and the US. The objective of TTLF's Working Paper Series is to share "work in progress". The authors of the papers are solely responsible for the content of their contributions and may use the citation standards of their home country. The TTLF Working Papers can be found at <http://ttlfs.stanford.edu>. Please also visit this website to learn more about TTLF's mission and activities.

If you should have any questions regarding the TTLF's Working Paper Series, please contact Vienna Law Professor Siegfried Fina, Stanford Law Professor Mark Lemley or Stanford LST Executive Director Roland Vogl at the

Stanford-Vienna Transatlantic Technology Law Forum
<http://ttlfs.stanford.edu>

Stanford Law School
Crown Quadrangle
559 Nathan Abbott Way
Stanford, CA 94305-8610

University of Vienna School of Law
Department of Business Law
Schottenbastei 10-16
1010 Vienna, Austria

About the Author

Leon Yehuda Anidjar is a Senior Fellow at the Owen Graduate School of Management, Vanderbilt University, and a Non-Residential Affiliate at the Transatlantic Technology Law Forum (TTLF), Stanford Law School. He is a Ph.D. candidate in the Department of Strategic Management and Entrepreneurship at the Rotterdam School of Management, Erasmus University, and an Affiliated Research Associate at the European Banking Institute. He previously served as Assistant Professor in Corporate Law at IE University in Madrid and held research positions at Goethe University Frankfurt and Harris Manchester College, Oxford. He holds an LL.B. (magna cum laude) and a Ph.D. in law from the Hebrew University of Jerusalem and an LL.M. from Tel Aviv University, and is a consultant and member of the Red Teaming Network at OpenAI. His work focuses on corporate governance, AI governance, and technology and financial regulation. He is the author of *Specific Performance in Contracts: Theory, Law & Reality* (Hebrew University Press, 2020).

General Note about the Content

The opinions expressed in this paper are those of the author and not necessarily those of the Transatlantic Technology Law Forum or any of its partner institutions, or the sponsors of this research project.

Suggested Citation

This TTTF Working Paper should be cited as:
Leon Anidjar, Artificial Intelligence Systems: Typologies, Risks, and Cross-Border Governance and Regulatory Frameworks in Europe and the U.S., Stanford-Vienna TTTF Working Paper No. 163, <http://ttlf.stanford.edu>.

Copyright

© 2026 Leon Anidjar

Abstract

AI acts as a catalyst for platform innovation, reshaping industries, and the broader ecosystem. This study delves into AI governance, balancing the need for rapid technological advancement with the necessity for stringent oversight. Our analysis presents AI governance as a complex array of regulatory strategies that address the dynamics between governmental oversight and market forces, encompassing centralized and decentralized controls, and spanning from mandatory to voluntary, rigid to flexible systems. We advocate for a governance approach that simultaneously fosters innovation and manages risk, while also addressing the challenges of collective action. The research indicates that effective AI governance requires more than just adherence to laws; it involves management structures and procedures that ensure AI evolves responsibly within a framework that is both legally sound and operationally adaptable. Moreover, the array of governance tools highlights a fundamental issue: to govern AI efficiently, one must confront the intrinsic complexities of the technology and unveil the 'black box' of AI decision-making processes.

TABLE OF CONTENTS

INTRODUCTION	2
I. ARTIFICIAL INTELLIGENCE SYSTEM TYPOLOGY	3
Capabilities	4
Spectrum of AI capabilities	4
Learning and Adaptation approach	6
Application domain	7
Data approach	7
Capacity to approximate human autonomy and judgment	10
II. ARTIFICIAL INTELLIGENCE AS A GRAND SOCIETAL CHALLENGE IN BUSINESS MANAGEMENT RESEARCH	12
AI risks at the society level	12
AI risks at the market level	14
AI risks at the firm level	16
Strategy	16
Organization	18
Governance	20
III. ARTIFICIAL INTELLIGENCE GOVERNANCE MODELS	24
Centralized mandatory and static regulations	24
Voluntary agreements and decentralized soft law	28
Market self-governance	30
Adaptive standards, specifications, and certifications	31
Assessment	33
IV. DISCUSSION AND FUTURE RESEARCH	34
Data infrastructures	34
Addressing the black box directly	35
Conclusion	36
References	38

INTRODUCTION

Over the last few years, the rise of the platform economy has not only reshaped commerce but has also redefined market dynamics on a global scale (De Reuver, Sørensen & Basole, 2018). Tech giants such as Airbnb, has disrupted traditional hotel industries; Uber has upended how we think about personal transportation; and Deliveroo has forged new paths in food delivery services. Scholar shave concentrated on the relationships between different market players, the cascading impact of network effects (McIntyre & Srinivasan, 2017), the nature of competition within these ecosystems (Rietveld & Schilling, 2021), and the governance arrangements among participants. This includes oversight of how platform owners collaborate and sometimes clash with their complementors—those who add value to the platform with their offerings (Chen, Tong, Tang & Han, 2022). For instance, Amazon bridged the gap between independent sellers and a global customer base, and Spotify created a link between artists and a worldwide audience, disrupting traditional music distribution models.

Platform owners are increasingly capitalizing on Artificial Intelligence (AI) to enhance user

interaction and streamline their operations (Jacobides, Brusoni, & Candelon, 2021). AI algorithms scrutinize user data to customize content and suggestions, thereby reinforcing user involvement and loyalty to the platform. In the dynamic landscape of digital platforms, the adoption of AI is not merely a trend but a necessity for innovation to meet consumer preferences and demands. Take, for instance, Replika—an AI-driven platform that furnishes users with a bespoke chatbot companion for conversation. This chatbot adapts its conversational patterns through machine learning, gradually aligning with the user's communication style. It enables users to discuss their daily experiences, emotions, or delve into more profound discussions, with the AI responding in a manner akin to human interaction.

Nonetheless, AI integration presents its own set of challenges, notably the magnification of societal biases. As AI is trained on vast datasets, it is susceptible to perpetuating any underlying biases present in the data, thus requiring vigilant governance to prevent the reinforcement of historical prejudices. Our study explores the role of AI in platform ecosystems, scrutinizes the array of AI technologies involved, and recognizes their capacity to amplify biases and cause damages. We evaluate a range of governance strategies designed to mitigate negative impacts and seek to achieve equilibrium between encouraging innovation in platform ecosystems and ensuring the ethical and equitable implementation of AI.

I. ARTIFICIAL INTELLIGENCE SYSTEM TYPOLOGY

AI encompasses a multitude of technologies, each designed to tackle distinct tasks and challenges. This spectrum of AI systems spans from the simplicity of rule-based systems, which follow explicit instructions and predefined rules, to the complexity of deep learning neural networks, particularly deep neural networks within the field of deep learning. We introduce a taxonomy of AI systems and distinguish between different AI systems based on the following categories:

Capabilities. We distinguish between AI systems with limited capabilities versus those with more general and human-like intelligence: *Weak AI systems* refer to applications designed for specific tasks or domains. These systems excel in the tasks they are trained for but lack the ability to understand or perform tasks beyond their predefined scope. They do not possess general intelligence and operate within a limited context (Uuk, Gutierrez & Tamkin, 2023; Voss & Jovanovic, 2023; Cao et al., 2023; Maham & Küspert, 2023).

Strong AI systems, on the other hand, represents AI systems that have the capability to understand, learn, and perform a wide range of tasks across different domains, much like human intelligence. Foundational models are large-scale machine learning models, predominantly within the realm of Natural Language Processing (NLP). They are trained on extensive datasets encompassing vast amounts of text, enabling them to understand and generate human language across a wide range of topics and domains (Bommasani et al., 2021; OpenAI, 2023).

Foundational models are strong AI, differ primarily from traditional AI systems in scope and capabilities. They are designed to possess human-like intelligence and the ability to understand, learn, and perform various tasks across different domains. They aim to emulate the breadth of human cognitive abilities and adaptability (Wu et al., 2023). These models like OpenAI's GPT-4, are central to numerous AI systems across different sectors. Not only does GPT-4 power ChatGPT, but it also underlies many third-party applications. The term "Generative AI" encompasses both these models and the applications built on them, highlighting their content generation capabilities.

Spectrum of AI capabilities. Artificial intelligence spans a range of systems, each showcasing distinct capabilities. Moving along a spectrum, we transition from specialized algorithms to the realm of potential human-like cognition (Russell & Norvig, 2016):

Rule-Based Systems are the most basic form of AI, where the machine operates based on a

predefined set of rules. Such systems don't have the capability to learn from data or experiences. Their decision-making process is based purely on the rules that have been programmed into them. Examples include early chatbots or decision tree algorithms.

Reactive Machines focus on specific tasks and are designed to respond to inputs. They don't have the ability to retain memory of past interactions, meaning they don't learn from previous experiences. Their decisions are made based solely on specific scenarios they are designed for. An iconic example of this is IBM's Deep Blue, a chess-playing system. It evaluated numerous potential moves but did not retain memory or learn from past games.

Limited Memory AI can store data or experiences for a short period, allowing them to make better decisions based on recent information. However, this storage is transient, and the AI does not evolve or learn long-term lessons from this data. A prime example of Limited Memory AI is the technology behind self-driving cars, where the AI might remember recent speed, distance, or obstacle data to make immediate navigation decisions.

Machine Learning & Deep Learning are AI systems learn and make decisions based on data. The more data they are exposed to, the better their decisions become. Deep learning, a subset of machine learning, utilizes neural networks to analyze various factors of data. Examples include voice recognition systems like Siri or Alexa, or image recognition software that can identify objects or faces in pictures.

Self-aware AI represents a hypothetical future stage of AI where machines possess their own consciousness or awareness. Such machines would have their own emotions, needs, beliefs, and potentially desires. This type of AI is currently theoretical and has been popularized by science fiction, but it's still a topic of debate among experts regarding its feasibility.

Artificial General Intelligence (AGI) refers to machines that can perform any intellectual task that

a human can do. Unlike other AI systems which are designed and trained for specific tasks, AGI would be versatile and able to learn and adapt to any new task, much like a human being. This spectrum moves from Rule-Based Systems to AGI, the systems become more complex, adaptable, and capable of independent thought and learning (Sternberg, 2020).

Figure 1: Spectrum of AI Capabilities



Learning and Adaptation approach. Foundational models are primarily trained on vast amounts of data across a wide range of topics using supervised learning techniques, allowing them to generate relevant responses across various subjects and contexts. After this Foundational training, they can be further optimized using reinforcement learning or be fine-tuned with additional supervised learning on more specific datasets to cater to tasks or domains.¹ This enhancement process equips them with the ability to handle specialized tasks while maintaining the broad knowledge base from the initial training (Cao, Sheng, McAuley & Yao, 2023). However, even with the adaptability that reinforcement learning provides, real-time learning from user interactions is limited; any significant new learning or adaptation would necessitate a deliberate retraining or fine-tuning phase (Zhou et al., 2023).

In contrast, Narrow AI systems are predominantly constructed using supervised learning. In this

¹ Reinforcement learning in Foundational models operates in a structured sequential manner. Initially, a base language model, referred to as Z, undergoes pre-training using vast amounts of data. Subsequent to this phase, a reward model is designed to align the model's outputs with human preferences. This model is then fine-tuned in an iterative manner utilizing the established reward function. During these iterations, human evaluators review and rank the responses produced by the model in response to specific prompts. Their feedback is instrumental in determining reward scores, which in turn, inform the ongoing refinement of the reward model. This process strives to optimize the alignment between the AI-generated content and human expectations (Ibarz et al., 2021; Wang et al., 2023).

paradigm, the AI is educated using a substantial dataset comprised of input-output pairs. By modifying its internal parameters, the system harnesses this data to yield precise predictions or classifications. A quintessential example of Narrow AI is a spam filter, which undergoes training on emails designated as either "spam" or "not spam." In the context of Narrow AI, the use of reinforcement learning is less prevalent because many Narrow tasks have a well-defined scope with available labeled data, making supervised learning a more straightforward choice.

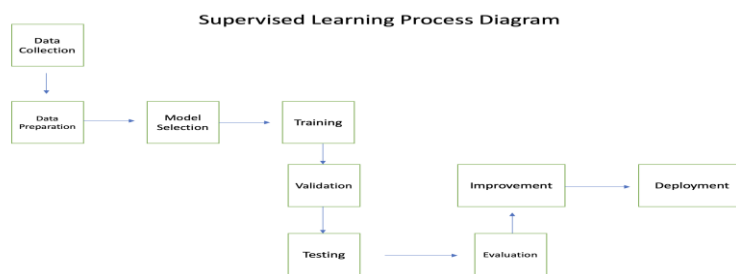
Application domain. Foundational models, endowed with their extensive training datasets and overarching capabilities, can serve diverse applications. These models have the proficiency to interpret and produce content spanning a wide spectrum of topics. As an illustration, GPT-4, a Foundational model, can seamlessly transition between varied tasks like holding discussions, crafting tales, or scripting code. Their modularity becomes even more evident when these models are accessed via APIs, allowing developers to integrate their vast knowledge into diverse applications (Patil, Zhang, Wang & Gonzalez, 2023). Furthermore, they can be fine-tuned for specialized endeavors. By layering a Foundational model pre-trained on general internet content with specific datasets, such as medical texts, it can be transformed into a valuable tool for healthcare professionals, aiding in diagnosis and therapeutic advice. In contrast, Narrow AI models are meticulously crafted for dedicated functions or a cluster of related tasks. These models excel when channeled for specific purposes, be it through direct application or APIs. Examples encompass voice assistants, honed for voice recognition; algorithms curated for specific outcomes like weather predictions; or systems designed for stock market analyses, fraud surveillance, and comprehensive risk assessments (Maple et al., 2023; Yeo et al., 2023; Kelly & Xiu, 2023).

Data approach. Structured and unstructured data represent the two primary categories of information that AI models work with. Structured data is well-organized, typically in tabular

format, with clearly defined relationships between variables. AI models tailored for structured data are adept at tasks that require analysis of this specific format. For example, an AI model might predict stock prices based on historical structured data or analyze customer purchase behavior from e-commerce transaction databases. Unlike structured data, unstructured data doesn't conform to a specific, predefined data model. It's often text-heavy and can be more challenging to analyze. Examples include emails, social media posts, books, articles, and more. Models rooted in Natural Language Processing (NLP) and deep learning (DL), such as GPT-4, are chiefly designed to handle and generate content. Their comprehensive training on a plethora of textual materials enables them to comprehend and formulate responses that pertain to such unstructured data (Yang et al., 2023). Moreover, there exist three fundamental types of data: input, training, and feedback. Together, these data categories power, refine, and strengthen machine learning algorithms, ensuring accuracy and efficiency across diverse applications and disciplines (Agrawal, Gans, & Goldfarb, 2018).

In Supervised Learning, the model is presented with input data accompanied by labels. Its primary task is to understand the association between this data and its respective labels. The training phase involves using a dataset filled with examples that have predetermined outcomes. Through this, the model fine-tunes its internal parameters to enhance its predictive capabilities. After prediction, feedback data emerges as the discrepancy between the model's predictions and the actual results.

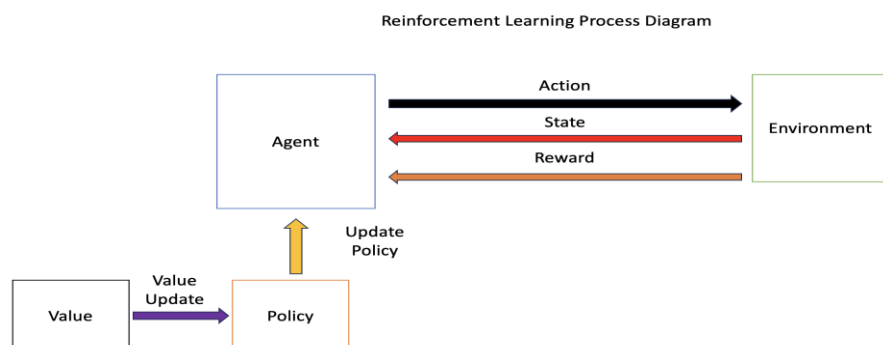
Figure 2: Machine Learning Cycle for Supervised Learning



Supervised Learning involves collecting raw data, preprocessing it for quality, dividing it into training and test sets, selecting and training a model, and tuning its parameters for optimal performance before deploying the refined model into its final application.

In Reinforcement Learning (RL), the entire learning process is centered around an agent interacting with an environment and trying to maximize a cumulative reward. Input data is primarily a representation of the current state of the environment. This state provides the agent with crucial information about its surroundings, enabling it to make informed decisions. Based on the perceived state from the input data, the agent decides on an action to take. The goal of this action is to obtain a positive reward or to avoid a negative consequence. As the agent interacts with the environment, it accumulates a collection of experiences. These experiences, typically presented in the form of state-action-reward sequences, form the training data in RL. This data is pivotal for the agent, as it allows it to understand the consequences of its actions in different states and to iteratively refine its strategy. Over time, the agent aims to develop a policy—a mapping of states to actions—that maximizes its overall reward. This feedback is vital for two reasons: firstly, it helps the agent evaluate the immediate benefit or cost of its action, and secondly, it guides the agent's learning process. If an action in each state consistently results in a positive reward, the agent learns to favor that action when encountering the state in the future. Conversely, if an action leads to a penalty, the agent learns to avoid that action in similar circumstances.

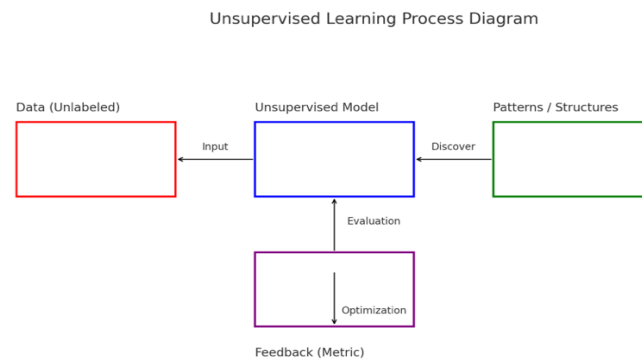
Figure 3: Machine Learning Cycle for Reinforcement Learning (RL)



Reinforcement Learning (RL) is a looped learning method where an agent learns to make decisions by performing actions in an environment, which result in new states and rewards or penalties. These outcomes inform the agent's value estimates for different actions, which then update its decision-making policy to maximize future rewards. This continuous feedback refines the agent's actions over time, aiming to optimize its long-term strategy.

In the context of Unsupervised Learning, the input data is presented without any accompanying labels. The primary challenge for the model is to unearth intrinsic patterns and structures within this data, such as clustering akin data points or simplifying data dimensions. As the model navigates through the data, it seeks to identify these patterns autonomously without the guidance of predetermined labels. Feedback in this domain often stems from external performance metrics, which serve to fine-tune the model, enhancing its ability to reveal concealed structures within the data. The following diagrams include describe data approaches in each of these ML applications.

Figure 4: Machine Learning Cycle for Reinforcement Learning (RL)



In unsupervised learning, a model sifts through unlabeled data to uncover patterns and structures on its own, clustering data, reducing complexity, or identifying relationships without any prior guidance. The model's success in revealing these patterns is measured and used to fine-tune its parameters, thereby improving its ability to detect these patterns. This iterative cycle of exploration, evaluation, and adjustment continues until the model's performance stabilizes or meets set success benchmarks.

Capacity to approximate human autonomy and judgment. Foundational models, drawing from a vast array of data, can approximate humans' complex decision-making processes. Their ability to process and weigh many factors against each other can resemble human judgment. For instance, when asked about a moral dilemma or a situational problem, Foundational models can generate answers that consider multiple perspectives. Moreover, these models, especially in interactive settings, can quickly create solutions or alternatives to posed problems, showcasing an ability like high cognitive function (Marjeh et al., 2023; Goldstein et al., 2022).

While these models can simulate judgment, they lack the emotional and ethical nuances inherent in human judgment. Humans often factor in empathy, personal experiences, cultural background, and moral considerations when making decisions, nuances that are challenging for a model to fully capture. Moreover, models' "originality" reflects the diverse data they have been trained on rather than a genuine spark of novelty. On the other hand, Narrow AI models are built to replicate certain human skills or decision-making abilities within a tightly defined scope. For example, a chess-playing AI is designed to mirror human strategy within the confines of chess. Narrow AI models offer consistent results within their specialized realm, uninfluenced by emotional biases or varying contexts. While it represents an advantage in terms of consistency, it also indicates a lack of comprehensive judgment. Therefore, their capacity for "judgment" or "reasoning," is bound exclusively to their predetermined purpose (See also, Dillion et al., 2023; Ganguli et al., 2023).

Table 1: Foundational and Narrow AI Systems – Main Distinctions

Category	Foundational Model (Including Generative AI)	Narrow AI
Capabilities	Broad range of tasks; can generate diverse outputs mimicking human cognition	Specific tasks or a set of closely related tasks
Learning & Adaptation	Trained on vast amounts of data and can be fine-tuned for specific tasks or domains	Typically trained on domain-specific data with limited adaptability outside that domain
Application	Highly versatile, can handle tasks ranging from general conversations to creating stories and even coding	Tailored specifically for a particular task or domain, such as voice recognition
Data	Can handle both structured and unstructured data, with capabilities in Natural Language Processing for handling unstructured data	Generally designed specifically for structured data scenarios, though there are exceptions that can handle unstructured data
Approximate Human Judgment	Generate outputs that mimic a broad spectrum of human cognitive function	Does not possess the breadth of human judgment and operates strictly within its designated task parameters.

II. ARTIFICIAL INTELLIGENCE AS A GRAND SOCIETAL CHALLENGE IN BUSINESS MANAGEMENT RESEARCH

The integration of AI and new technologies, such as blockchains and cryptocurrencies (Wang, Lumineau & Schilke, 2022) augmented reality (AR) and virtual reality (VR), edge computing, metaverse (Wang et al., 2023) and internet of things (IoT) (Li, Zhu & Darbandi, 2022) into the business landscape is undeniably revolutionizing operations, strategies, and competitive dynamics (Murray, Kim & Combs, 2023; Weking, Desouza, Fiert & Kowalkiewicz, 2023). The incorporation of AI technologies into the economic landscape goes beyond mere technological advancement; it signifies a fundamental change in perspective. Companies and institutions are undergoing profound alterations in multiple areas. This encompasses improvements in operational efficiency, the adoption of data-informed decision-making, and the evolution of sophisticated strategies and business models. This transformation emphasizes the importance for businesses to remain forward-thinking in fully leveraging AI while addressing its intrinsic challenges. Consequently, we view AI risks as grand societal challenges. Given AI's transformative impact, there is a pressing need to reevaluate and modify traditional management and governance practices to address AI risks (George, et al., 2016; Ferraro, Etzion & Gehman, 2015). We categorize these risks across three dimensions: *society, market, and the firm*, and further distinguishing between Foundational AI and Narrow AI, to offer a structured approach to understanding the multifaceted challenges and implications.

AI risks at the society level. Artificial Intelligence presents a myriad of ethical challenges that impact both societal structures and individual lives. Narrow AI, characterized by its specialization, excels in specific tasks. However, its restricted domain can sometimes magnify biases or prove inadequate in uncharted situations. For example, an AI system optimized to regulate urban traffic

signals may enhance daily flow and reduce congestion. Yet, during unforeseen events like emergency evacuations, its limitations might emerge, resulting in inefficiencies or even detrimental outcomes (Mallinson, Azevedo, Best, Robles & Wang, 2023).

In contrast, Foundational AI aims for a wider, almost human-like comprehension and capability. This expansive nature ushers in deep-seated ethical quandaries. Imagine a Foundational AI system responsible for urban traffic management. Confronted with a sudden accident, it may be tasked with immediate life-or-death decisions, like choosing which vehicle's passengers to prioritize for safety. Such scenarios evoke complex ethical dilemmas, akin to the renowned "trolley problem." These instances emphasize the imperative to align AI-driven decisions with human values, morals, and ethical standards (Shelby et al., 2023; Hendrycks, Mazeika & Woodside, 2023; Jenkins, Cerny & Hribek, 2022; Lai et al., 2021).

Furthermore, Foundational AI models, which are trained on vast datasets, can unintentionally perpetuate, and even magnify societal biases and stereotypes (Wachter, 2022; Weidinger et al., 2021). These models operate in a "black box" manner, meaning their decision-making processes are not easily understood. As they are used across numerous applications, the potential harmful effects of their biases can impact many people, potentially more than individual human biases (Ferrara, 2023).

Human biases have been long recognized, but the rise of Foundational AI can magnify the consequences of these biases. For instance, an AI model might, due to biased training data, associate certain racial or ethnic groups with higher qualifications, which could lead to unfair advantages or disadvantages (Hagerty & Rubinov, 2019; Liu et al., 2021). Such widespread use of biased AI can lead to reinforced societal biases against marginalized groups.

Moreover, AI models have the potential to amplify disinformation campaigns. For instance, during

the 2016 U.S. presidential election, false information and propaganda were spread online at an unprecedented scale (Lazer et al., 2023). With AI, these efforts could become cheaper, more scalable, and more convincing. Using AI in tools like chatbots can offer highly personalized interactions, potentially influencing users' views on controversial topics. This personalization can also lead users to share more personal information, which can further be used to tailor disinformation.

Additionally, AI model may produce or disseminate information that is not based on factual data or its training but appears plausible ("hallucinations") (Ye, et al., 2023). This can be compared to a human having a false memory or making a statement based on an incorrect recollection. When an AI model "hallucinates", it might generate content that sounds authoritative and accurate, even if it's entirely fabricated or lacks a basis.

Lastly, AI models can be harnessed for enhanced surveillance. Their capability to process and understand vast amounts of data in different formats (text, audio, image, video) can be employed to monitor public spaces and online content. Because AI models can understand nuanced content, such as detecting hate speech, enabling real-time monitoring of many individuals. This makes AI a potent tool for surveillance, lowering the previously high costs and complexities associated with such endeavors.

AI risks at the market level. Foundational AI models serve as the bedrock for a multitude of applications. This trend underscores a shift toward the concentration of economic power in the hands of a select few firms (Weidinger et al., 2022). As the development of Foundational AI often requires massive resources, data, and expertise, there's a possibility that only a few large corporations or entities can genuinely compete in this space. This concentration can lead to a monopolistic or oligopolistic market structure, potentially stifling competition, innovation, and

leading to higher prices for consumers (e.g., Korinek & Stiglitz, 2021).

Moreover, because data serves as the lifeblood for AI systems (Anidjar, Packin & Panezi, 2023) particularly for Foundational AI, possessing control over expansive data sets enables these entities to persistently improve their AI models and enhanced their capabilities. This phenomenon can engender a feedback loop: superior AI draws more users, leading to an influx of data, which then bolsters the AI even more. Therefore, data monopolies have the potential to centralize economic power, with many AI-generated profits funneled to those dominating the data sphere, amplifying concerns about escalating economic disparities (Acemoglu, 2021).

Furthermore, Foundational AI, by virtue of its general applicability, can be tailored to perform a wide range of tasks that were previously exclusive to human labor. This adaptability can lead to significant job displacements across various sectors, from manual tasks to even more complex cognitive roles (Webb, 2019). As Foundational models become more sophisticated, their capacity to automate a vast array of tasks across diverse sectors grows exponentially. While automation can undoubtedly enhance operational efficiency and overall productivity, it concurrently ignites concerns about possible job losses and displacements (Acemoglu & Restrepo, 2019; Eloundou et al., 2023). This transition could result in heightened unemployment rates and a profound misalignment of skills in the labor market (Autor, 2022).

Additionally, a few dominant companies, including Meta, Microsoft, OpenAI and Alphabet are at the forefront of developing the Foundational AI models that are the backbone for multiple value-generating applications. Developers working on applications downstream often require direct or indirect access to these models. The way the models are disseminated, be it through APIs or as open source, significantly determines the level of influence and authority the originating developers retain control over their use and adaptation (Mathijssen, Overeem & Jansen, 2020).

However, the benefits derived from these Foundational AI models, or the subsequent applications constructed upon them might not be evenly distributed across the board. Such an uneven distribution could amplify existing economic and social disparities, potentially leading to a more divided technological landscape where access and benefits are concentrated with a few technology giants (Weidinger et al., 2022). On the other hand, because Narrow AI systems are specialized tools built for singular tasks, be it translating languages, playing a specific game, or recognizing images, the development and commercialization of these tools can be distributed among numerous players. For instance, one company might excel in creating a Narrow AI for medical diagnoses, while another might specialize in financial forecasting.

AI risks at the firm level. The swift incorporation of AI into the business landscape has ushered in remarkable benefits, from optimizing processes to spearheading novel innovations.² Yet, with these progressions come palpable challenges that companies must address. These hurdles touch upon diverse business areas, such as strategy, organization, and governance protocols.

Strategy. The literature at the intersection of AI and business strategy has explored how AI can strategically enhance decision-making processes, with a particular focus on improving decision support and fostering stronger stakeholder relationships. Kitsios & Kamariotou's (2016) work sheds light on the growing importance of AI in the business landscape. By aligning decision support systems with strategic information systems planning principles, businesses can integrate AI solutions more harmoniously into their overarching strategies. For instance, AI can be unitized

² Other research has constructed a framework that synthesizes both the antecedents and consequences of adopting AI in the pursuit of various forms of innovation. These innovations include product, process, business model, incremental, radical, digital, social, sustainable, open, service, disruptive, market, and organizational innovations (Mariani, Machado & Nambisan, 2023; Mariani, Machado, Magrelli & Dwivedi, 2023). Other researchers have delved into the use of AI in areas such as marketing, technological progression, ethical considerations, distribution channels, strategic marketing, performance metrics, and the STP model (Segmentation, Targeting, and Positioning) (Vlacic et al., 2021; Davenport et al, 2020; Mariani, Perez-Vega & Wirtz, 2022). Lastly, researchers have investigated how AI can shape production and consumption behaviors to promote sustainable resource utilization and sustainable business practices (Vaio, Palladino, Hassan & Escobar, 2020).

to pinpoint the most economically advantageous IT infrastructure configurations that align with the overarching business strategy (Thompson, Ekman, Selby & Whitaker, 2014).

However, the implementation of AI-driven decision support systems can be a double-edged sword. While they offer the potential for enhanced efficiency and insights, the costs associated with their implementation and maintenance can be burdensome for small to medium-sized businesses.

Furthermore, while AI can analyze vast datasets and make predictions based on patterns, it lacks the human touch. Decision-making in business is not always black and white. Many situations require intuition, emotional intelligence, cultural understanding, and ethical considerations – traits inherently human and not easily replicated by machines. Over-relying on AI tools might lead businesses to overlook these vital aspects of decision-making, which could be detrimental in the long run (Passi & Vorvoreanu, 2022; Schoeffer, De-Arteaga & Kuehl, 2022). Relatedly, automation bias is the tendency for individuals to overly depend on automated recommendations and occurs when users neglect to act because an automated alert system doesn't notify them (Vered et al., 2023; Gilovich, Griffin & Kahneman, 2002; Mosier & Skitka, 2018).

Several other studies have delved into the strategic integration of AI to enhance customer and prospective employee relationship (Black & van Esch, 2020), offering the potential to elevate customer experiences through personalization, improved service quality, and enhanced convenience. Through continuous learning and the ability to process feedback in real-time, AI-driven systems can identify areas of improvement and refine service offerings. This ensures that customer queries are addressed promptly and efficiently.

For example, Sujata et al. (2019) introduced a conceptual model designed to align IT strategy with business strategy, emphasizing the strategic deployment of AI in various applications, including sentiment analysis, emotion detection, virtual assistants, chatbots, and content curation. While AI

holds promise in revolutionizing customer relationships through personalization, improved service quality, and enhanced convenience, it also presents some challenges. AI-driven service tools, such as chatbots, may misinterpret complex queries, can lead to customer frustration, or can make experiences impersonal, inconvenient, and even time consuming (Borges et al., 2021).

Organization. The concept of bounded rationality suggests that humans have a limited capacity to process information, leading to decision-making that might not always be optimal, as they can only consider a certain amount of data and variables (Simon, 1955). However, AI has the potential to revolutionize this because of its ability to consider a broader range of information, scenarios, and variables when making decisions or predictions that are more 'rational' in the sense that they are based on a wider array of information and a more comprehensive understanding of potential outcomes (i.e., Csaszar & Steinberger, 2022).

Studies have explored how strategic alignment and organizational design impact an organization's capability to integrate AI effectively. Companies that prioritize ongoing learning and encourage innovation have a smoother transition to AI. This is largely due to their employees being more receptive to technological advancements and having an eye for AI's potential applications (Mikalef & Gupta, 2021; Enholm, Papagiannidis, Mikalef & Krogstie, 2021). This is largely due to their employees being more receptive to technological advancements and having an eye for AI's potential applications. Such proactive organizations are better equipped to weave AI seamlessly into their workflow.

The effects of AI technologies have been recognized to focus on HRM strategies, specifically in areas like job substitution, collaboration between humans and robots/AI, decision-making and learning prospects. Additionally, they influence HRM activities such as recruitment, training, and job performance (Vrontis et al., 2022). In a study analyzing the impact of a generative AI

conversational assistant on 5,179 customer support agents, it was found that the AI tool boosted productivity by 14% on average. Notably, novice and low-skilled workers saw a 35% improvement, while experienced workers had minimal changes (Brynjolfsson, Li & Raymond, 2023). Another study revealed that Foundational AI isn't just about boosting efficiency; it's transformative. The technology cut down time usage by 25% and led to a 40% enhancement in practical management tasks (Dell'Acqua et al., 2023).

Nonetheless, for Foundational AI to be successfully integrated, it is crucial that employees have faith in its decision-making abilities. The challenge here stems from the fact that AI lacks the kind of intuitive reasoning that humans possess, creating difficulties in merging human insight with data produced by AI. Therefore, building trust in AI requires consistent and dependable performance, transparency in its operations, and an ongoing commitment to employee training and education (Makarius et al., 2020).

Furthermore, employees and job candidates occasionally express doubts regarding human expertise in handling AI systems. While there is a prevailing belief that such systems could potentially rival or surpass human performance, people often question the fairness of AI-automated decisions compared, especially when it comes to Narrow AI. Thus, a human-focused perspective, advocated by academics and policymakers, emphasizes the importance of considering individuals' perceptions of fairness during the design and execution phases of algorithmic decision-making (Starke, Baleis, Keller & Marcinkowski, 2021; Langer & Landers, 2021). Moreover, organizations aiming to reap the advantages of AI adoption need to focus beyond mere technical assets, and to invest in non-technical elements, such as human skills, leadership, team collaboration, a culture of innovation, and governance methods to allow an effective AI-employee collaboration (Chowdhury et al., 2023).

Governance. The board of directors is paramount in a corporation, serving as its main governing body. Entrusted with the responsibility of oversight, the board plays a crucial role in monitoring the company's operations, ensuring that management practices align with stakeholders' interests. In addition to their monitoring role, the board of directors also contributes to the service and strategic direction of the company. By leveraging their collective expertise, insights, and industry knowledge, board members provide invaluable guidance to executive management (Boivie, Withers, Graffin & Corley, 2021). In the boardroom, certain decisions can shape the future direction of an entire organization. Traditional decision-making relied heavily on human expertise, experience, and intuition. However, with the advent of AI, vast amounts of data can be analyzed swiftly to derive previously inaccessible insights or are too time-consuming to obtain. When board members employ AI-driven analytics, they can make more informed choices about market trends, risks, and efficiencies (Cihon, Schuett, & Baum, 2021). The synergy between board functions and AI models can be reflected in one of the following ways (Hilb, 2020):

Table 2: AI and Board of Directors Synergy

Type of Human-AI Synergy	Implications for the Board of Directors
Assisted Intelligence	The board primarily guides the decision-making process, leveraging AI for specific tasks such as translations or converting voice to text
Augmented Intelligence	In this framework, even though the board is the ultimate decision-maker, AI serves a pivotal role. It provides insights that go beyond human capacities. A notable example is AI's ability to quickly analyze vast amounts of data to identify anomalies, proving itself invaluable
Amplified Intelligence	This model embodies a synergy between the board's intellect and AI capabilities. While AI can suggest potential courses of action, it's the directors who have the final say, particularly in intricate scenarios that demand nuanced judgement
Amplified Intelligence	This model embodies a synergy between the board's intellect and AI capabilities. While AI can suggest potential courses of action, it's the directors who have the final say, particularly in intricate scenarios that demand nuanced judgement
Autopoietic Intelligence	This forward-thinking AI concept envisions systems that not only autonomously make decisions but also self-improve over time.

The synergies between board functions and AI present several challenges, especially when considering the difference between Narrow AI and Foundational AI. *First*, because Narrow AI can

only excel in a designed function and lacks the capability to perform tasks outside of its specialization, board might become overly reliant on AI for certain functions, potentially overlooking human expertise, which might be more suitable in nuanced or complex situations. This “may lead to a diminution in the skills and competence required for human governance. Over-dependency on AI may lead to a temptation to defer to decisions to AI as a default rule, but the credibility and effectiveness of such systems still need to be tested over a long period of time.” (Zhao, 2022: 689). Moreover, even with Foundational AI, there is the potential for outcomes that are inaccurate, incomplete, or biased. While such AI can be instrumental in generating ideas and offering readily available factual data, directors must verify its reliability and trustworthiness to uphold their fiduciary responsibilities (Cunningham et al., 2023; Hickman & Petrin, 2021). While big data and AI offer useful insights, they still need skilled interpretation for decision-making. AI requires knowledgeable users to maximize its potential. Therefore, although technology enhances board functions, it cannot replace the expertise often needed for such functions (Picciau, 2021).

Second, AI tools operate based on programming instructions, not common human incentives, or broad standards of conduct. Loyalty obligations are inapplicable to algorithms. While AI may be unbiased, its operations depend on its programming. Algorithms lack the capability for broad interpretations and cannot be held directly accountable for poor decisions. As a result, any accountability arising from AI's decisions falls on the legal entities using them or the board members overseeing them (Armour & Eindenmüller, 2020. For the opposite view, see Li, 2023).

Third, A main concern with Foundational AI is whether it can attain consciousness and becomes vulnerable to exploitation (Butlin et al., 2023; Bauer, Liebich, Hinz & Kosfeld, 2023). Yet, corporate law stipulates that board members must be human (Gramitto Ricci, 2020).

The following table summarizes the primary AI risks covered so far and discusses potential research questions for future exploration.

Table 3: Research Agenda for AI Risks and Harms

	Society Level	Market Level	Firm Level
Ethical Concerns	Life-or-death decisions related to traffic management and the "trolley problem."	AI models influencing market trends based on biased or fabricated data.	Alignment of AI decisions with company strategy and ethics.
<i>Future Research Questions</i>	<p>How might AI-driven decisions redefine societal understanding of moral dilemmas?</p> <p>What mechanisms can ensure that AI respects diverse societal ethics across regions?</p> <p>Can societal ethical frameworks be harmonized for global AI governance?</p>	<p>How can markets ensure fairness in AI-driven economic predictions?</p> <p>What checks and balances can be put in place to prevent AI misuse in market analytics?</p> <p>How might AI ethics shape consumer trust in various market segments?</p>	<p>How are AI ethics integrated into corporate governance structures?</p> <p>What ethical challenges do firms face when balancing stakeholder interests in AI deployments?</p> <p>Can ethics training be effectively incorporated into AI development cycles within firms?</p>
Bias & Stereotype	Magnification of societal biases and stereotypes by AI models.	Potential for biased market predictions and investment strategies.	Internal biases affecting corporate decisions, operations, and organizational efficiencies
<i>Future Research Questions</i>	<p>How can AI-driven systems be designed to mitigate, rather than amplify, systemic societal biases?</p> <p>What societal institutions are most at risk of being influenced by AI biases, and how can they be safeguarded?</p> <p>How does AI bias influence societal decision-making in critical areas like justice system?</p>	<p>How do AI biases influence market value propositions and competitive dynamics?</p> <p>Can AI-driven market solutions lead to new forms of bias, and how can they be preemptively addressed?</p> <p>What role does market regulation play in monitoring and rectifying AI biases?</p>	<p>How can firms develop internal metrics to continuously monitor and correct AI-driven biases?</p> <p>What role does organizational culture play in perpetuating or mitigating AI biases?</p> <p>How can firms foster transparency and accountability in addressing AI biases in their operations?</p> <p>How can firms integrate fairness tools into their AI workflows?</p> <p>Can firms be held accountable for unintentional biases in their AI models?</p>
Disinformation & Surveillance	Spread of false information using AI at an unprecedented scale. Enhanced surveillance in online spaces.	AI-driven market forecasts that may be based on "hallucinations."	Use of AI-generated data that may not be entirely factual.

<i>Future Research Questions</i>	<p>How can societies create and enforce standards to counteract AI-driven disinformation?</p> <p>What is the societal toll of widespread AI surveillance on individual privacy and freedoms?</p> <p>Can there be a balance between AI-driven surveillance for safety and individual rights?</p>	<p>How does AI-driven disinformation impact market perceptions and investor behaviors?</p> <p>What mechanisms can markets employ to quickly verify or refute AI-generated information?</p> <p>How can markets ensure transparency in AI-powered surveillance tools, ensuring they are not misused?</p>	<p>How are firms navigating the challenges of AI-driven disinformation within their industries?</p> <p>What internal measures can firms take to ensure that AI tools respect user privacy while optimizing operations?</p> <p>Can firms create a balance between AI-driven data collection and individual privacy rights of employees and customers?</p>
Black Box Issues	Difficulty in understanding the decision-making processes of AI.	Challenges in validating the reliability of AI market models.	Difficulty in deciphering AI decisions within firm operations and organization.
<i>Future Research Questions</i>	<p>How can societies promote understanding and transparency in complex AI-driven processes?</p> <p>What societal measures can ensure that non-transparent AI systems are not misused in critical sectors like healthcare?</p> <p>How can public trust be rebuilt in sectors tainted by non-transparent AI mishaps?</p>	<p>What market mechanisms can enhance transparency in AI-driven trading or investment strategies?</p> <p>How can markets ensure that black box AI models do not lead to unwarranted financial risks?</p> <p>What are the market implications of a widespread move towards explainable AI models??</p>	<p>How can firms enhance internal understanding of complex AI-driven processes among employees and stakeholders?</p> <p>What measures can firms implement to ensure AI transparency in consumer-facing products?</p> <p>How does the black box nature of AI influence firm risk management and mitigation strategies?</p>

III. ARTIFICIAL INTELLIGENCE GOVERNANCE MODELS

We have conceptualized AI systems along a continuum that spans from Narrow to Strong AI. Each point on this spectrum holds unique repercussions for society, markets, and businesses. In response to the myriad challenges AI presents, AI governance refers to the development of AI systems, the societal, ethical, and economic consequences of AI utilization, and the competition challenges posed by a limited number of tech giants controlling the global AI ecosystem (Veale, Matus & Gorwa, 2023; Maas & Villalobos, 2023; Roberts, Hine, Taddeo & Floridi, 2023). When developing governance frameworks, it's essential to consider AI variations, highlighting how AI heterogeneity affects governance in these domains. We distinguish between various forms of governance: hard law regulations, international soft law agreements and codes, stakeholder-oriented industry self-governance and management standards and specifications. These types of governance reflect different distinctions, such as state v. market dynamics, centralized v. decentralized, mandatory v. voluntary and adaptive v. static regulatory arrangements. We examine each separately and discuss its respective weaknesses (i.e., Ulnicane et al., 2021; Taeihagh, 2021; Petit & De Cooman, 2020).

Centralized mandatory and static regulations. The EU Artificial Intelligence Act (AIA) proposal focuses on managing the risks associated with AI systems used within the EU, emphasizing that risks are its primary concern (Veale & Borgesius, 2021).³ The AIA classifies AI systems based on four distinct risk levels: unacceptable, high, limited, and minimal/none, with primary regulatory attention directed at the unacceptable and high-risk categories. Systems with

³ Recital 14 articulates this risk-based approach, stating: "To establish a balanced and effective framework for AI systems, a distinct risk-based strategy is essential. This strategy should align the nature and details of regulations with the potential risks posed by the AI systems. As such, it's vital to ban certain AI practices, set requirements for high-risk AI systems, define responsibilities for the associated operators, and mandate transparency provisions for specific AI systems."

an "unacceptable risk" are characterized by their significant potential for manipulation, often targeting vulnerable groups such as the elderly, disabled, or those of certain socioeconomic standings, as seen in social scoring systems.⁴

High-risk systems can be split into two subcategories: (i) those that function as safety components or products already under safety regulations, like toys and medical equipment; and (ii) those used for sensitive tasks, encompassing areas like biometrics, critical infrastructure, education, workforce management, vital services, law enforcement, migration oversight, and judicial processes.

For high-risk AI systems, developers are mandated to ensure their solutions neither jeopardize health, safety, nor infringe upon fundamental rights. This encompasses a broad spectrum of measures from risk management, data governance, and monitoring to transparency, human oversight,⁵ and cybersecurity benchmarks (Articles 6–51). On another note, AI systems in the "limited risk" category have constraints primarily centered on transparency, mandating users be informed when interacting with such an AI and highlighting any artificial or tampered content (Chamberlain, 2023).⁶

The EU Parliament also suggested that general-purpose AI providers should adhere to obligations resembling those for high-risk AI systems. It further delineates responsibilities across the AI value chain. Specifically, providers of foundational models must share data and system development details with downstream developers to ensure AIA compliance. Generative AI system providers must meet comprehensive transparency standards, strive to prevent the creation of unlawful content, and provide a summary of copyrighted material usage in their training data.

⁴ Explanatory memorandum of the regulation, 5.2.2.

⁵ Article 13 & 14(1).

⁶ Article 52

However, recent study of Bommasani et al (2023) presented an index that uses 100 social and technical metrics to assess the transparency of developers during the creation and application of foundational models by firms like OpenAI, Google, and Meta. The research revealed that none of these developers offer comprehensive data on the broader effects of their main models, including affected user count, impacted market sectors, or harm mitigation strategies. Interestingly, "open" developers, those sharing model details and potential data, outshine their "closed" peers, such as API providers, in transparency.

Furthermore, AIA indirectly enhances the investigative capabilities of European competition agencies. Article 59 mandates each Member State to establish a national supervisory authority, which will act both as a notifying and market surveillance entity. These authorities will oversee the procedures for assessing and monitoring conformity assessment bodies. Furthermore, Article 63 stipulates that any relevant information discovered during market surveillance, which might be pertinent to Union competition laws, must be promptly reported by the national supervisory authority to the Commission and relevant national competition agencies. This ensures that competition-relevant data found by supervisory authorities is shared with competition agencies for enforcement purposes (Schrepel, 2023).

Weaknesses. "Risk" in regulatory contexts is often viewed as a multifaceted concept that comprises two primary dimensions: the probability that a certain harm will occur and the scale or impact of that harm when it does. Risk based regulation is therefore about "focusing on *outcomes* rather than *specific rules and process* as the goal of regulation" (OECD, 2021, chapter 6). In risk-based regulatory governance, the emphasis shifts from viewing risk as merely a target of regulation to seeing it as a strategic approach to such governance (Van Der Heijden, 2019), and it can be described as "a method grounded in evidence to direct resource use and prioritize the most

significant risks within a clear, methodical, and justified structure" (Black & Baldwin, 2010: 181).

When applying risk frameworks for regulating AI technologies, two challenges must be addressed:

First, advanced AI models, particularly deep neural networks, are often characterized as "black boxes." This term refers to the opacity of these systems; while they can produce remarkable and accurate outcomes, understanding the intricate pathways and relationships within them that lead to a given decision remains elusive. In simpler models, the correlation between an input and its corresponding output can be direct and apparent. However, in sophisticated Foundational AI architectures, the multitude of layers and connections is so complex that even experts might struggle to interpret them fully. This inherent opacity challenges the application of traditional risk frameworks to AI technologies.

Second, risks associated with AI are **not** static. They can fluctuate based on technological advancements, shifts in usage patterns, and changing societal values. For instance, what is deemed a low-risk application today might evolve into a high-risk scenario in the future due to unforeseen consequences or new use cases.

Third, risks in AI systems might not always manifest immediately. Instead, they could accumulate over time, leading to larger systemic problem. For example, a slight bias in a machine learning model might seem negligible initially but can compound over time as the model process more biased data, leading to significant discriminatory outcomes. The AIA translation of general legal values into rigid obligations limits adaptive risk evaluation (Grozdanovski & De Cooman, 2022). Considering generative AI systems like Google Bard and ChatGPT highlights some limitations in the AIA's risk-focused approach. Foundational AI differs from the 'conventional' AI models that the AIA predominantly addresses, especially regarding their versatility and widespread application. This differentiation presents several challenges (Hacker, Engel & Mauer, 2023):

Firstly, the classification system set forth by the AI Act primarily bases its assessment on the intended purpose of the AI, as specified by the provider. Nevertheless, in the case of versatile AI systems like generative AI, the responsibility to determine the usage often falls on the professional user. This situation could lead to a potential gap in risk assessment, as the current framework may not necessarily encourage AI system providers to thoroughly evaluate inherent biases or the possibilities of misuse. Providers could potentially circumvent classifications that would label their product as high-risk.

Secondly, generative AI platforms differ significantly from more traditional forms of AI. Their adaptable nature and wide array of potential applications set them apart. The current AI Act places a strong emphasis on data accuracy and the minimization of biases. Yet, as generative AI continues to evolve and pull from increasingly diverse data sources, there are growing concerns surrounding "data equity." This brings to the forefront the ethical considerations related to not fully acknowledging or protecting the rights of the original content creators. Consequently, trying to pigeonhole foundational AI models into specific categories like High-Risk or Prohibited unveils a challenge. It reflects a delicate balance between avoiding under-regulation, which might allow for unchecked biases and misuse, and over-regulation, which could stifle innovation and the development of AI technologies (Novelli et al., 2023; Helberger & Diakopoulos, 2023).

Voluntary agreements and decentralized soft law. Political alignment institutions strive to promote agreement among countries in addressing global policy matters (Cihon, Matthijs & Luke, 2020). Even in the absence of formal structures, these institutions can create international forums for establishing norms on AI related regulatory issues (Erdelyi & Goldsmith, 2018, 2022). While these norms are not enforceable, they can still influence behavior and gain traction. Entities focused on AI-related political consensus-building suggest incorporating AI discussions into the

schedules of existing, renowned consensus-building groups like the OECD (2019), or UNESCO (2021), Notable norm-defining bodies include the International Organization for Standardization (ISO), the International Electrotechnical Commission (IEC), the International Telecommunication Union (ITU), and the UNCITRAL's Working Group on Electronic Commerce.

The U.S. federal government has embraced a governance approach for AI that is decentralized and based on risk assessment. The AI Bill of Rights (AIBoR) Blueprint underscores the possible adverse effects that AI may have on economic stability and civil rights. Nevertheless, because of its advisory nature, the real-world effectiveness of the AIBoR depends significantly on the actions pursued by distinct federal agencies, resulting in an implementation process that is somewhat uncertain. This pattern was reinforced by the recent Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (October 2023). It directs specific agencies to oversee AI risks within their respective sectors, with the Federal Trade Commission assigned to boost AI development competition, while the Department of Health and Human Services is tasked with regulating AI in healthcare. Moreover, the order compels AI companies involved in national security or critical infrastructure to submit safety test results and relevant data to the government. While agencies are developing voluntary standards and best practices for AI, their application is limited to government agencies. The private sector, however, has the discretion to follow these voluntary standards as they choose.

Recently, the US's National Security Commission on Artificial Intelligence introduced the idea of creating an Emerging Technology Coalition. This group is set to promote the growth and application of novel technologies in harmony with democratic ideals and standards. Comparably, the Council of Europe (CoE) has been a leader, especially known for its treaties on human rights, cybercrime, and data privacy. With AI's growing importance, the CoE created the Ad Hoc

Committee on AI (CAHAI) in 2019 to consider comprehensive legal framework for AI.

Weaknesses. Global governance via international soft law arrangements encounters several shortcomings. At its core, soft law lacks a legally binding nature. This means that even if states or entities agree to specific norms or standards, there is no formal mechanism to penalize non-compliance. In multilateral settings, the dynamics become even more intricate. The process of shaping soft laws often involves multiple stakeholders from different countries and organizations. This setup can inadvertently give dominant or influential entities an undue advantage, allowing them to shape the guidelines in ways that primarily serve their interests. Lastly, the inclusion or exclusion of major players in the soft law negotiation process can significantly impact its success. For instance, excluding a global player like China from such discussions could jeopardize the entire negotiation process (Maas & Villalobos, 2023).

Market self-governance. Several industry organizations and stakeholders have introduced voluntary standards, guidelines, and codes of conduct, largely driven by major technology companies or their close affiliates (Veale et al., 2023). Companies like Microsoft, with its Office of Responsible AI, and Google's Advanced Technology External Advisory Council, play a central role. Although these entities emphasize their intention to foster collaboration among AI industry stakeholders, it's not guaranteed that their main goal is to shape a socially responsible AI governance. These leading companies control a vast majority of essential resources, including hardware, expertise, and data, giving them a unique edge in AI innovation. Their resource dominance positions them to have a major impact on the direction of AI policy, research, and development (Hemphill, 2016; Taeihagh, 2021).

In 2022, top AI firms like AI21 Labs and OpenAI stressed the importance of transparency in AI operations by releasing a joint "Recommendation for Language Model Deployment." They urged

developers to provide usage guidelines, acknowledge model weaknesses, and share safety measures. Google's AI guidelines also highlight "Transparency Artifacts" to shed light on their AI technology's functionality, decision-making processes, and the need for a comprehensive perspective that includes technology, people, and the overall environment.

Weaknesses. The incorporation of corporate agendas into AI's regulatory environment has sparked concerns about the potential overshadowing of broader societal objective. The dominant influence of leading tech companies, combined with the intricate nature of AI algorithms, threatens to public involvement. For example, “major technology companies have exerted significant influence over the framing of AI issues and the formulation of AI policy, such as through lobbying efforts and their inclusion in AI expert groups formed by government” (Taeihagh, 2021: 145; Jobin, Ienca & Vayena, 2019). This situation results in a precedence to the views of a select few, reflecting a regulatory capture, at the expense of the broader public's interest and the core values of trustworthiness and fairness (Guihot, Matthew & Suzor, 2017).

Adaptive standards, specifications, and certifications. Standardization is key to addressing information asymmetries in AI by aiding in the selection and monitoring of providers, ensuring they meet ethical and auditing goals (Goodman & Trehu, 2023). The process of certification not only assists users identify and trust reputable AI services but also encourages providers to adopt and adhere to high standards, as it can be used to showcase AI providers' commitment to ethics and societal values (Cihon, Kleinaltenkamp, Schuett & Bauu, 2021). Historically, global computing governance has relied heavily on self-regulatory bodies to set engineering standards crucial for the seamless functioning of networked technologies. ISO/IEC JTC 1 formed the Subcommittee 42 dedicated to Artificial Intelligence back in 2017. Prior to this, in 2016, IEEE initiated the Global Initiative on Ethics of Autonomous and Intelligent Systems, which has been

working on the P7000 series of standards, focusing on the ethical aspects of AI and intelligent systems. Based on AIA's Article 9, the European Commission has presented a preliminary standardization request to CEN and CENELEC to develop safe and trustworthy artificial intelligence standards that will lay the groundwork for creating European harmonized adaptive AI management framework aligned with AIA's goals (Schuett, 2023). In a similar token, the U.S. proactively establishes infrastructures to manage AI risks, with significant efforts led by the National Institute of Standards and Technology (NIST). NIST's AI Risk Management Framework (RMF), which was finalized in January 2023, is a voluntary framework that provides detailed guidance on managing AI risks across its lifecycle. Moreover, NIST is creating an AI RMF Playbook that will offer practical implementation guidance, supplemented with case studies illustrating risk mitigation in specific AI scenarios (Engler, 2023).

Weaknesses. The standardization of AI presents challenges that are exacerbated by the lack of harmonized international AI mandatory legal framework. At the core is the need for definitive and repeatable standards that secure privacy, safety, and accessibility, while also aligning with the broader values and fundamental rights upheld by the EU. These standards must carefully differentiate between the technical details and broader legal and ethical contexts. Their development requires a collaborative approach that includes a wide array of stakeholders from diverse backgrounds, ensuring that the standards are shaped by a multiplicity of insights and concerns. Furthermore, in the European context, AIA outlines only high-level directives for high-risk AI applications, the detailed technical standards for compliance will be delineated by harmonized standards, a process that is indicative of where substantive policymaking is likely to take place. This delegation of detailed rulemaking to standardization bodies raises critical concerns about the legitimacy of such processes, particularly in an international context where there is no

agreed-upon regulatory consensus, emphasizing the need for standards that genuinely embody the AIA norms of protecting fundamental rights while being sensitive to the global absence of uniform AI regulations (Tartaro, 2023).

Assessment. Navigating the governance of AI involves a complex interplay between the rigidity of established legal frameworks and the swift progression of technological advancements. Hard laws provide a stable, predictable foundation, setting a baseline for safety and ethical standards in the rapidly evolving realm of AI. However, their relative rigidity can lead to mismatches with the ever-changing tech landscape, sometimes hindering innovation or leaving novel AI applications in a regulatory gray area. On the other hand, soft law strategies and industry-led governance introduce a degree of flexibility to the regulatory environment. These approaches adapt more fluidly to the changing landscape of AI technologies, promoting rapid, innovation-focused responses that do not necessarily require revising rigid legal frameworks. In a similar vein, standardization efforts attempt to bridge this gap, aiming to establish universal benchmarks and protocols applicable to AI technologies. These efforts are designed to align the precision of legal standards with the continuous progress of technological development. Yet, the nascent framework for AI governance is compromised by AI's critical position in global competition, the ineffectiveness of international bodies, and disagreements on AI policy priorities worldwide, creating substantial obstacles to effective collective action (Roberts, Hine, Taddeo & Floridi, 2023). The governance of AI requires a synergistic approach that combines the certainty of law with the dynamism of management. Legal structures provide the necessary safety and ethical guidelines, while management practices offer the agility to adapt to technological changes. This balance ensures AI develops responsibly within a framework that is both legally sound and operationally efficient.

IV. DISCUSSION AND FUTURE RESEARCH

The governance landscape for artificial intelligence is diverse, capturing a spectrum of regulatory approaches that include command-and-control regulation, network governance, self-regulation, co-regulation, technical standardization, and accreditation (Baldwin, Cave & Lodge, 2010). These approaches reflect the multifaceted nature of AI risks and benefits. In the absence of a global consensus on AI regulatory norms, this diversity is not only inevitable but also advantageous (OECD, 2023). Each governance method—whether hard law, soft law, industry self-regulation, or standardization—brings its own strengths and can mitigate the limitations of the others. Thus, the heterogeneity of AI governance mirrors the complex challenges posed by AI technologies, necessitating a composite approach that leverages the full array of regulatory tools to manage the risks and maximize the benefits of AI advancements (Brownsword, Scotford & Yeung, 2016). However, the multitude of governance tools for regulating AI systems suggests a significant underlying deficiency. The wide range of governance methods used to regulate AI, encompassing everything from stringent legal frameworks to flexible industry standards and self-regulation, highlights a significant underlying problem: the intrinsic complexity and lack of transparency in AI systems make them resistant to simple regulatory approaches. The so-called 'black box' nature of AI, involving intricate data processing and decision-making layers that lack transparency and predictability, poses substantial obstacles to the development of efficient and cost-effective governance measures. To tackle this issue, we call policymakers and scholars to concentrate their research efforts on investigating two principal paths:

Data infrastructures. States may serve as the primary source of finance required to support the initial and critical phases of AI system development. While private entrepreneurial entities are critical for developing novel technologies for commercialization purposes, their business model

relies heavily on rapid results and expects excessively high returns. As a result, they would be unmotivated to adopt voluntary governance measures or internalize external ones under significantly underdeveloped enforcement environment (Mazzucato et al., 2022). Mariana Mazzucato (2014) argued that the government plays a primary role in driving disruptive innovation that revolutionizes industries and economic growth (Mazzucato, Kattel & Ryan-Collins, 2022); it must design a policy that puts an end to the “*era of light-touch self-regulation*” (Ramos & Mazzucato, 2022). To bridge this divide, governance frameworks ought to concentrate on the regulation of data assets, mandating essential data handling techniques both before and after data processing. Preprocessing of data necessitates a thorough examination of correlations and distributions within datasets, along with the mitigation of biases through methods like reweighting and sampling (Wu, 2022; Liu et al., 2023). In dealing with datasets that have complex or multiple features, there is a challenge in discerning the fundamental data patterns. Researchers recommend the creation of standards or markers that identify specific data attributes, thus aligning them with societal values (Bartolo et al., 2020). Furthermore, post-processing approaches such as recalibration, enrichment, and synthetic data generation can be utilized following the training process to ensure model outputs are in sync with overarching policy objectives (Yang & Stoyanovich, 2016; van Breugel et al., 2021; Gal & Lynskey, 2023). These approaches enhance the resilience of machine learning systems by introducing them to a broader spectrum of data scenarios (Manousakas & Aydınoğlu, 2023; Liang et al., 2022).

Addressing the black box directly. Advocating for ethical AI necessitates a thorough understanding of its core—the learning algorithm. The widespread deployment of AI in society requires more than just legal transparency regarding its outputs and data or the establishment of technical infrastructure standards. These approaches alone are inadequate for addressing the

fundamental biases that are often embedded in AI systems. It is the unobscured visibility into the learning algorithms at the heart of AI that is critical for recognizing and correcting these biases, thereby ensuring the responsible application and maintenance of AI's integrity (von Eschenbach, 2021). Engineered for superior predictive performance, deep learning algorithms surpass traditional methods in areas such as image and speech recognition, along with forecasting consumer behavior. Their design, which often involves intricate network structures capable of sifting through complex datasets, inherently sacrifices transparency for accuracy. Consequently, algorithms' internal decision-making processes remain obscure to human observers.

In response, the field has seen the emergence of Explainable AI (XAI), a suite of techniques developed to unravel the complexity of these models. XAI works by simplifying AI's elaborate decision paths into forms more intelligible to humans, effectively converting "black-box" models into "glass-box" models for greater transparency. These XAI methods differ in their application; some are embedded within the model's architecture, making them model-specific, while others are model-agnostic, deriving explanations from the model's input-output relations. They also vary in the breadth of explanation, providing either a comprehensive overview of the model's logic or insights into individual predictions. The goal of XAI is to bolster confidence in AI applications by clarifying how AI systems function and decide, thus fostering a greater degree of trust among users and developers (Das & Rad, 2020; Rai, 2020).

Conclusion. AI governance must balance predictability with the adaptability required for fast-paced technological innovation. Hard laws establish essential safety and ethical benchmarks but risk obsolescence, whereas soft laws and industry self-regulation deliver the flexibility to evolve with technological progress, with standardization efforts seeking to synchronize these varying regulatory mechanisms. Nonetheless, AI governance is challenged by global competition,

ineffective international coordination, and conflicting policy priorities. The diverse regulatory tools reflect the complexity of AI systems, which resist simple regulation due to their opaque, 'black box' nature. Thus, effective AI governance requires a nuanced approach that balances legal certainty with operational adaptability to navigate the intricate landscape of AI technology and its global implications.

References

- Acemoglu, D. (2021). *Harms of AI* (No. w29247) (p. w29247). Cambridge, MA: National Bureau of Economic Research. Retrieved November 3, 2023, from <http://www.nber.org/papers/w29247.pdf>
- Acemoglu, D., & Restrepo, P. (2019). Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives*, 33(2), 3–30. Retrieved November 3, 2023, from <https://pubs.aeaweb.org/doi/10.1257/jep.33.2.3>
- Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction, judgment, and complexity: A theory of decision-making and artificial intelligence. *The economics of artificial intelligence: An agenda* (pp. 89–110). University of Chicago Press.
- Anidjar, L. Y., Packin, N. G., & Panezi, A. (2023). The Matrix of Privacy: Data Infrastructure in the AI-Powered Metaverse. *SSRN Electronic Journal*. Retrieved November 3, 2023, from <https://www.ssrn.com/abstract=4363208>
- Armour, J., & Eidenmueller, H. G. M. (2019). Self-Driving Corporations? *SSRN Electronic Journal*. Retrieved November 3, 2023, from <https://www.ssrn.com/abstract=3442447>
- Autor, D. (2022). *The Labor Market Impacts of Technological Change: From Unbridled Enthusiasm to Qualified Optimism to Vast Uncertainty* (No. w30074) (p. w30074). Cambridge, MA: National Bureau of Economic Research. Retrieved November 3, 2023, from <http://www.nber.org/papers/w30074.pdf>
- Baldwin, R., Cave, M., & Lodge, M. (Eds.). (2010). *The Oxford Handbook of Regulation* (1st ed.). Oxford University Press. Retrieved November 3, 2023, from <https://academic.oup.com/edited-volume/34523>
- Bartolo, M., Roberts, A., Welbl, J., Riedel, S., & Stenetorp, P. (2020). Beat the AI: Investigating Adversarial Human Annotation for Reading Comprehension. *Transactions of the Association for Computational Linguistics*, 8, 662–678. Retrieved November 3, 2023, from <https://direct.mit.edu/tacl/article/96474>
- Bauer, K., Liebich, L., Hinz, O., & Kosfeld, M. (2023). Decoding GPT’s hidden “rationality” of cooperation. *SSRN Electronic Journal*. Retrieved November 3, 2023, from <https://www.ssrn.com/abstract=4576036>
- Black, J. S., & Van Esch, P. (2020). AI-enabled recruiting: What is it and how should a manager use it? *Business Horizons*, 63(2), 215–226. Retrieved November 3, 2023, from <https://linkinghub.elsevier.com/retrieve/pii/S0007681319301612>
- Black, J., & Baldwin, R. (2010). Really Responsive Risk-Based Regulation. *Law & Policy*, 32(2), 181–213. Retrieved November 3, 2023, from <https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9930.2010.00318.x>
- Boivie, S., Withers, M. C., Graffin, S. D., & Corley, K. G. (2021). Corporate directors’ implicit theories of the roles and duties of boards. *Strategic Management Journal*, 42(9), 1662–1695. Retrieved November 3, 2023, from <https://onlinelibrary.wiley.com/doi/10.1002/smj.3320>
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., et al. (2022, July 12). On the Opportunities and Risks of Foundation Models. arXiv. Retrieved November 3, 2023, from <http://arxiv.org/abs/2108.07258>
- Bommasani, R., Klyman, K., Longpre, S., Kapoor, S., Maslej, N., Xiong, B., Zhang, D., et al. (2023). The Foundation Model Transparency Index. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2310.12941>
- Borges, A. F. S., Laurindo, F. J. B., Spínola, M. M., Gonçalves, R. F., & Mattos, C. A. (2021). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. *International Journal of Information Management*, 57, 102225. Retrieved November 3, 2023, from <https://linkinghub.elsevier.com/retrieve/pii/S0268401219317906>
- Brownsword, R., Scotford, E., & Yeung, K. (Eds.). (2016). *The Oxford Handbook of Law, Regulation and*

- Technology* (Vol. 1). Oxford University Press. Retrieved November 3, 2023, from <https://academic.oup.com/edited-volume/27999>
- Brynjolfsson, E., Li, D., & Raymond, L. (2023). *Generative AI at Work* (No. w31161) (p. w31161). Cambridge, MA: National Bureau of Economic Research. Retrieved November 3, 2023, from <http://www.nber.org/papers/w31161.pdf>
- Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., Deane, G., et al. (2023). Consciousness in Artificial Intelligence: Insights from the Science of Consciousness. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2308.08708>
- Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023, March 7). A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT. arXiv. Retrieved November 3, 2023, from <http://arxiv.org/abs/2303.04226>
- Cao, Y., Sheng, Q. Z., McAuley, J., & Yao, L. (2023). Reinforcement Learning for Generative AI: A Survey. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2308.14328>
- Chamberlain, J. (2023). The Risk-Based Approach of the European Union's Proposed Artificial Intelligence Regulation: Some Comments from a Tort Law Perspective. *European Journal of Risk Regulation*, 14(1), 1–13. Retrieved November 3, 2023, from https://www.cambridge.org/core/product/identifier/S1867299X22000381/type/journal_article
- Chen, L., Tong, T. W., Tang, S., & Han, N. (2022). Governance and Design of Digital Platforms: A Review and Future Research Directions on a Meta-Organization. *Journal of Management*, 48(1), 147–184. Retrieved November 6, 2023, from <http://journals.sagepub.com/doi/10.1177/01492063211045023>
- Chowdhury, S., Dey, P., Joel-Edgar, S., Bhattacharya, S., Rodriguez-Espindola, O., Abadie, A., & Truong, L. (2023). Unlocking the value of artificial intelligence in human resource management through AI capability framework. *Human Resource Management Review*, 33(1), 100899. Retrieved November 3, 2023, from <https://linkinghub.elsevier.com/retrieve/pii/S1053482222000079>
- Cihon, P., Kleinaltenkamp, M. J., Schuett, J., & Baum, S. D. (2021). AI Certification: Advancing Ethical Practice by Reducing Information Asymmetries. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2105.10356>
- Cihon, P., Maas, M. M., & Kemp, L. (2020). Fragmentation and the Future: Investigating Architectures for International AI Governance. *Global Policy*, 11(5), 545–556. Retrieved November 3, 2023, from <https://onlinelibrary.wiley.com/doi/10.1111/1758-5899.12890>
- Cihon, P., Schuett, J., & Baum, S. D. (2021). Corporate Governance of Artificial Intelligence in the Public Interest. *Information*, 12(7), 275. Retrieved November 3, 2023, from <https://www.mdpi.com/2078-2489/12/7/275>
- Csaszar, F. A., & Steinberger, T. (2022). Organizations as Artificial Intelligences: The Use of Artificial Intelligence Analogies in Organization Theory. *Academy of Management Annals*, 16(1), 1–37. Retrieved November 3, 2023, from <http://journals.aom.org/doi/full/10.5465/annals.2020.0192>
- Das, A., & Rad, P. (2020). Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey. arXiv. Retrieved November 6, 2023, from <https://arxiv.org/abs/2006.11371>
- Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48(1), 24–42. Retrieved November 3, 2023, from <http://link.springer.com/10.1007/s11747-019-00696-0>
- De Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: A research agenda. *Journal of information technology*, 33(2), 124–135. SAGE Publications Sage UK: London, England.
- Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Kraymer, L., et al.

- (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *SSRN Electronic Journal*. Retrieved November 3, 2023, from <https://www.ssrn.com/abstract=4573321>
- Di Vaio, A., Palladino, R., Hassan, R., & Escobar, O. (2020). Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review. *Journal of Business Research*, 121, 283–314. Retrieved November 3, 2023, from <https://linkinghub.elsevier.com/retrieve/pii/S0148296320305191>
- Dillion, D., Tandon, N., Gu, Y., & Gray, K. (2023). Can AI language models replace human participants? *Trends in Cognitive Sciences*, 27(7), 597–600. Retrieved November 3, 2023, from <https://linkinghub.elsevier.com/retrieve/pii/S1364661323000980>
- Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2303.10130>
- Engler, A. (2023). The EU and US diverge on AI regulation: A transatlantic comparison and steps to alignment. Brookings Institution.
- Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2022). Artificial Intelligence and Business Value: A Literature Review. *Information Systems Frontiers*, 24(5), 1709–1734. Retrieved November 3, 2023, from <https://link.springer.com/10.1007/s10796-021-10186-w>
- Erdélyi, O. J., & Goldsmith, J. (2018). Regulating artificial intelligence: Proposal for a global solution (pp. 95–101). Presented at the Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society.
- Erdélyi, O. J., & Goldsmith, J. (2022). Regulating artificial intelligence: Proposal for a global solution. *Government Information Quarterly*, 39(4), 101748. Retrieved November 3, 2023, from <https://linkinghub.elsevier.com/retrieve/pii/S0740624X22000843>
- Ferrara, E. (2023). Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2304.03738>
- Ferraro, F., Etzion, D., & Gehman, J. (2015). Tackling grand challenges pragmatically: Robust action revisited. *Organization Studies*, 36(3), 363–390. Sage Publications Sage UK: London, England.
- Gal, M., & Lynskey, O. (2023). Synthetic Data: Legal Implications of the Data-Generation Revolution.
- Ganguli, D., Askell, A., Schiefer, N., Liao, T. I., Lukošiušė, K., Chen, A., Goldie, A., et al. (2023). The Capacity for Moral Self-Correction in Large Language Models. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2302.07459>
- George, G., Howard-Grenville, J., Joshi, A., & Tihanyi, L. (2016). Understanding and tackling societal grand challenges through management research. *Academy of Management Journal*, 59(6), 1880–1895. Academy of Management Briarcliff Manor, NY.
- Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment* (1st ed.). Cambridge University Press. Retrieved November 3, 2023, from <https://www.cambridge.org/core/product/identifier/9780511808098/type/book>
- Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., Nastase, S. A., et al. (2022). Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience*, 25(3), 369–380. Retrieved November 3, 2023, from <https://www.nature.com/articles/s41593-022-01026-4>
- Goodman, E. P., & Trehu, J. (2023). Algorithmic Auditing: Chasing AI Accountability. *Santa Clara High Technology Law Journal*, 39(3), 289.
- Gramitto Ricci, S. A. (2019). Artificial agents in corporate boardrooms. *Cornell L. Rev.*, 105, 869.

- Grozdanovski, L., & De Cooman, J. (2022). Of Hypothesis and Facts: The Curious Origins of the EU's Regulation of High-Risk AI. Presented at the Second Seminar of the EURA Center of Excellence.
- Guihot, M., Matthew, A. F., & Suzor, N. P. (2017). Nudging robots: Innovative solutions to regulate artificial intelligence. *Vand. J. Ent. & Tech. L.*, 20, 385. HeinOnline.
- Hacker, P., Engel, A., & Mauer, M. (2023). Regulating ChatGPT and other Large Generative AI Models. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2302.02337>
- Hagerty, A., & Rubinov, I. (2019). Global AI Ethics: A Review of the Social Impacts and Ethical Implications of Artificial Intelligence. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/1907.07892>
- Helberger, N., & Diakopoulos, N. (2023). ChatGPT and the AI Act. *Internet Policy Review*, 12(1). Retrieved November 3, 2023, from <https://policyreview.info/essay/chatgpt-and-ai-act>
- Hemphill, T. A. (2016). Regulating nanomaterials: A case for hybrid governance. *Bulletin of Science, Technology & Society*, 36(4), 219–228. SAGE Publications Sage CA: Los Angeles, CA.
- Hendrycks, D., Mazeika, M., & Woodside, T. (2023). An Overview of Catastrophic AI Risks. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2306.12001>
- Hickman, E., & Petrin, M. (2021). Trustworthy AI and Corporate Governance: The EU's Ethics Guidelines for Trustworthy Artificial Intelligence from a Company Law Perspective. *European Business Organization Law Review*, 22(4), 593–625. Retrieved November 3, 2023, from <https://link.springer.com/10.1007/s40804-021-00224-0>
- Hilb, M. (2020). Toward artificial governance? The role of artificial intelligence in shaping the future of corporate governance. *Journal of Management and Governance*, 24(4), 851–870. Retrieved November 3, 2023, from <https://link.springer.com/10.1007/s10997-020-09519-9>
- Ibarz, J., Tan, J., Finn, C., Kalakrishnan, M., Pastor, P., & Levine, S. (2021). How to Train Your Robot with Deep Reinforcement Learning; Lessons We've Learned. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2102.02915>
- Jacobides, M. G., Brusoni, S., & Candelon, F. (2021). The Evolutionary Dynamics of the Artificial Intelligence Ecosystem. *Strategy Science*, 6(4), 412–435. Retrieved November 6, 2023, from <https://pubsonline.informs.org/doi/10.1287/stsc.2021.0148>
- Jenkins, R., Cerny, D., & Hr̆bek, T. (2022). *Autonomous Vehicle Ethics: The Trolley Problem and Beyond*. Oxford University Press.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. Retrieved November 3, 2023, from <https://www.nature.com/articles/s42256-019-0088-2>
- Kelly, B. T., & Xiu, D. (2023). Financial Machine Learning. *SSRN Electronic Journal*. Retrieved November 3, 2023, from <https://www.ssrn.com/abstract=4501707>
- Kitsios, F., & Kamariotou, M. (2016). Decision Support Systems and Business Strategy: A Conceptual Framework for Strategic Information Systems Planning. *2016 6th International Conference on IT Convergence and Security (ICITCS)* (pp. 1–5). Retrieved November 3, 2023, from <http://ieeexplore.ieee.org/document/7740323/>
- Korinek, A., & Stiglitz, J. (2021). Artificial Intelligence, Globalization, and Strategies for Economic Development. *Institute for New Economic Thinking Working Paper Series*, 1–53. Retrieved November 3, 2023, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3812820
- Lai, V., Chen, C., Liao, Q. V., Smith-Renner, A., & Tan, C. (2021). Towards a Science of Human-AI Decision Making: A Survey of Empirical Studies. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2112.11471>

- Langer, M., & Landers, R. N. (2021). The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. *Computers in Human Behavior*, 123, 106878. Retrieved November 3, 2023, from <https://linkinghub.elsevier.com/retrieve/pii/S0747563221002016>
- Lawrence A. Cunningham, Arvin Maskin and James B. Carlson. (n.d.). Generative Artificial Intelligence and Corporate Boards: Cautions and Considerations. Retrieved from <https://corpgov.law.harvard.edu/2023/06/21/generative-artificial-intelligence-and-corporate-boards-cautions-and-considerations/>
- Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., et al. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. American Association for the Advancement of Science.
- Li, Y., Zhu, X., & Darbandi, M. (2023). A comprehensive and bibliometric review on the blockchain-enabled IoT technology for designing a secure supply chain management system. *Journal of Management & Organization*, 29(4), 745–762. Retrieved November 3, 2023, from https://www.cambridge.org/core/product/identifier/S1833367222000748/type/journal_article
- Li, Z. (2023). Technology Governance Under Corporate Law. *U. of Pittsburgh Legal Studies Research Paper*, (2023–31).
- Liang, W., Tadesse, G. A., Ho, D., Fei-Fei, L., Zaharia, M., Zhang, C., & Zou, J. (2022). Advances, challenges and opportunities in creating data for trustworthy AI. *Nature Machine Intelligence*, 4(8), 669–677. Nature Publishing Group UK London.
- Liu, H., Wang, Y., Fan, W., Liu, X., Li, Y., Jain, S., Liu, Y., et al. (2023). Trustworthy AI: A Computational Perspective. *ACM Transactions on Intelligent Systems and Technology*, 14(1), 1–59. Retrieved November 3, 2023, from <https://dl.acm.org/doi/10.1145/3546872>
- Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., et al. (2023). Summary of ChatGPT-Related Research and Perspective Towards the Future of Large Language Models. *arXiv*. Retrieved November 3, 2023, from <https://arxiv.org/abs/2304.01852>
- Maas, M. M., & Villalobos Ruiz, J. J. (2023). International AI Institutions: A Literature Review of Models, Examples, and Proposals. *SSRN Electronic Journal*. Retrieved November 3, 2023, from <https://www.ssrn.com/abstract=4579773>
- Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research*, 120, 262–273. Retrieved November 3, 2023, from <https://linkinghub.elsevier.com/retrieve/pii/S0148296320305002>
- Mallinson, D. J., Azevedo, L., Best, E., Robles, P., & Wang, J. (2023). The Future of AI Is in the States: The Case of Autonomous Vehicle Policies. *Business and Politics*, 1–20.
- Manousakas, D., & Aydin, S. (2023). On the Usefulness of Synthetic Tabular Data Generation. *arXiv preprint arXiv:2306.15636*.
- Maple, C., Szpruch, L., Epiphaniou, G., Staykova, K., Singh, S., Penwarden, W., Wen, Y., et al. (2023). The AI Revolution: Opportunities and Challenges for the Finance Sector. *arXiv*. Retrieved November 3, 2023, from <https://arxiv.org/abs/2308.16538>
- Mariani, M. M., Machado, I., & Nambisan, S. (2023). Types of innovation and artificial intelligence: A systematic quantitative literature review and research agenda. *Journal of Business Research*, 155, 113364. Retrieved November 3, 2023, from <https://linkinghub.elsevier.com/retrieve/pii/S0148296322008293>
- Mariani, M. M., Machado, I., Magrelli, V., & Dwivedi, Y. K. (2023). Artificial intelligence in innovation

- research: A systematic review, conceptual framework, and future research directions. *Technovation*, 122, 102623. Retrieved November 3, 2023, from <https://linkinghub.elsevier.com/retrieve/pii/S0166497222001705>
- Mariani, M. M., Perez-Vega, R., & Wirtz, J. (2022). AI in marketing, consumer research and psychology: A systematic literature review and research agenda. *Psychology & Marketing*, 39(4), 755–776. Retrieved November 3, 2023, from <https://onlinelibrary.wiley.com/doi/10.1002/mar.21619>
- Marjeh, R., Sucholutsky, I., van Rijn, P., Jacoby, N., & Griffiths, T. L. (2023). Large language models predict human sensory judgments across six modalities. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2302.01308>
- Mathijssen, M., Overeem, M., & Jansen, S. (2020). Identification of Practices and Capabilities in API Management: A Systematic Literature Review. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2006.10481>
- Mazzucato, M. (2014). *The entrepreneurial state: Debunking public vs. private sector myths*. Anthem frontiers of global political economy (Revised edition.). London ; New York: Anthem Press.
- Mazzucato, M., Kattel, R., & Ryan-Collins, J. (2020). Challenge-Driven Innovation Policy: Towards a New Policy Toolkit. *Journal of Industry, Competition and Trade*, 20(2), 421–437. Retrieved November 3, 2023, from <http://link.springer.com/10.1007/s10842-019-00329-w>
- Mazzucato, M., Schaake, M., Krier, S. & Entsminger, J. (2022). Governing artificial intelligence in the public interest. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2022-12). Retrieved November 3, 2023, from <https://www.ucl.ac.uk/bartlett/public-purpose/wp2022-12>
- McIntyre, D. P., & Srinivasan, A. (2017). Networks, platforms, and strategy: Emerging views and next steps. *Strategic Management Journal*, 38(1), 141–160. Wiley Online Library.
- Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, 58(3), 103434. Retrieved November 3, 2023, from <https://linkinghub.elsevier.com/retrieve/pii/S0378720621000082>
- Mosier, K. L., & Skitka, L. J. (2018). Human decision makers and automated decision aids: Made for each other? *Automation and human performance* (pp. 201–220). CRC Press.
- Murray, A., Kim, D., & Combs, J. (2023). The promise of a decentralized internet: What is Web3 and how can firms prepare? *Business Horizons*, 66(2), 191–202. Retrieved November 3, 2023, from <https://linkinghub.elsevier.com/retrieve/pii/S0007681322000714>
- Novelli, C., Taddeo, M., & Floridi, L. (2023). Accountability in artificial intelligence: What it is and how it works. *AI & SOCIETY*. Retrieved November 3, 2023, from <https://link.springer.com/10.1007/s00146-023-01635-y>
- OECD (2019). AI Principles. Retrieved from <https://oecd.ai/en/ai-principles>
- OECD. (2021). *OECD Regulatory Policy Outlook 2021*. OECD. Retrieved November 3, 2023, from https://www.oecd-ilibrary.org/governance/oecd-regulatory-policy-outlook-2021_38b0fdb1-en
- OpenAI. (2023). GPT-4 Technical Report. arXiv. Retrieved September 16, 2023, from <https://arxiv.org/abs/2303.08774>
- Passi, S., & Vorvoreanu, M. (2022). *Overreliance on AI: Literature Review* (No. MSR-TR-2022-12). Microsoft. Retrieved from <https://www.microsoft.com/en-us/research/publication/overreliance-on-ai-literature-review/>
- Patil, S. G., Zhang, T., Wang, X., & Gonzalez, J. E. (2023). Gorilla: Large Language Model Connected with Massive APIs. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2305.15334>

- Pegah Maham and Sabrina Küspert. (2023). *Governing General Purpose AI — A Comprehensive Map of Unreliability, Misuse and Systemic Risks*. Retrieved from [Www.stiftung-Nv.de. https://www.stiftung-nv.de/de/publikation/governing-general-purpose-ai-comprehensive-map-unreliability-misuse-and-systemic-risks](http://www.stiftung-nv.de/de/publikation/governing-general-purpose-ai-comprehensive-map-unreliability-misuse-and-systemic-risks)
- Petit, N., & De Cooman, J. (2020). Models of Law and Regulation for AI. *SSRN Electronic Journal*. Retrieved November 3, 2023, from <https://www.ssrn.com/abstract=3706771>
- Picciau, C. (2021). The (Un) Predictable Impact of Technology on Corporate Governance. *Hastings Bus. LJ*, 17, 67.
- Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, 48(1), 137–141. Retrieved November 6, 2023, from <http://link.springer.com/10.1007/s11747-019-00710-5>
- Rietveld, J., & Schilling, M. A. (2021). Platform competition: A systematic and interdisciplinary review of the literature. *Journal of Management*, 47(6), 1528–1563. SAGE Publications Sage CA: Los Angeles, CA
- Roberts, H., Hine, E., Taddeo, M., & Floridi, L. (2023). Global AI governance: Barriers and pathways forward. *SSRN Electronic Journal*. Retrieved November 3, 2023, from <https://www.ssrn.com/abstract=4588040>
- Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach*. Prentice Hall series in artificial intelligence (Third edition, Global edition.). Boston Columbus Indianapolis New York San Francisco Upper Saddle River Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montreal Toronto Delhi Mexico City Sao Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo: Pearson.
- Schoeffler, J., De-Arteaga, M., & Kuehl, N. (2022). On Explanations, Fairness, and Appropriate Reliance in Human-AI Decision-Making. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2209.11812>
- Schrepel, T. (2023). Decoding the AI Act: A Critical Guide for Competition Experts. *Amsterdam Law & Technology Institute Working Paper*, 3–2023.
- Schuett, J. (2023). Risk Management in the Artificial Intelligence Act. *European Journal of Risk Regulation*, 1–19. Retrieved November 3, 2023, from https://www.cambridge.org/core/product/identifier/S1867299X23000016/type/journal_article
- Shelby, R., Rismeni, S., Henne, K., Moon, Aj., Rostamzadeh, N., Nicholas, P., Yilla, N., et al. (2022). Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2210.05791>
- Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1), 99. Retrieved November 3, 2023, from <https://academic.oup.com/qje/article-lookup/doi/10.2307/1884852>
- Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2022). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society*, 9(2), 205395172211151. Retrieved November 3, 2023, from <http://journals.sagepub.com/doi/10.1177/20539517221115189>
- Sternberg, R. J. (2020). Rethinking what we mean by intelligence. *Phi Delta Kappan*, 102(3), 36–41. Retrieved November 3, 2023, from <http://journals.sagepub.com/doi/10.1177/0031721720970700>
- Sujata, J., Aniket, D., & Mahasingh, M. (2019). Artificial intelligence tools for enhancing customer experience. *International Journal of Recent Technology and Engineering*, 8(2), 700–706.
- Taeiagh, A. (2021). Governance of artificial intelligence. *Policy and Society*, 40(2), 137–157. Retrieved November 3, 2023, from <https://academic.oup.com/policyandsociety/article/40/2/137/6509315>
- Taeiagh, A. (2021). Governance of artificial intelligence. *Policy and Society*, 40(2), 137–157. Retrieved November 3, 2023, from <https://academic.oup.com/policyandsociety/article/40/2/137/6509315>
- Tartaro, A. (2023). Towards European Standards Supporting the AI Act: Alignment Challenges on the Path to Trustworthy AI (pp. 98–106). Presented at the Proceedings of the AISB Convention.
- The state of implementation of the OECD AI Principles four years on* (OECD Artificial Intelligence Papers No.

- 3). (2023). OECD Artificial Intelligence Papers (Vol. 3). Retrieved November 3, 2023, from https://www.oecd-ilibrary.org/science-and-technology/the-state-of-implementation-of-the-oecd-ai-principles-four-years-on_835641c9-en
- Thompson, S., Ekman, P., Selby, D., & Whitaker, J. (2014). A model to support IT infrastructure planning and the allocation of IT governance authority. *Decision Support Systems*, 59, 108–118. Retrieved November 3, 2023, from <https://linkinghub.elsevier.com/retrieve/pii/S0167923613002571>
- Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W.-G. (2021). Framing governance for a contested emerging technology: insights from AI policy. *Policy and Society*, 40(2), 158–177. Retrieved November 3, 2023, from <https://academic.oup.com/policyandsociety/article/40/2/158/6509326>
- UNESCO, C. (2021). Recommendation on the ethics of artificial intelligence. UNESCO France.
- Uuk, R., Gutierrez, C. I., & Tamkin, A. (2023). Operationalising the Definition of General Purpose AI Systems: Assessing Four Approaches. *SSRN Electronic Journal*. Retrieved November 3, 2023, from <https://www.ssrn.com/abstract=4471151>
- van Breugel, B., Kyono, T., Berrevoets, J., & van der Schaar, M. (2021). DECAF: Generating Fair Synthetic Data Using Causally-Aware Generative Networks. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2110.12884>
- Van Der Heijden, J. (2019). Risk Governance and Risk-Based Regulation: A Review of the International Academic Literature. *SSRN Electronic Journal*. Retrieved November 3, 2023, from <https://www.ssrn.com/abstract=3406998>
- Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4), 97–112. Retrieved November 3, 2023, from <https://www.degruyter.com/document/doi/10.9785/cri-2021-220402/html>
- Veale, M., Matus, K., & Gorwa, R. (2023). AI and Global Governance: Modalities, Rationales, Tensions. *Annual Review of Law and Social Science*, 19. Annual Reviews Inc.
- Vered, M., Livni, T., Howe, P. D. L., Miller, T., & Sonenberg, L. (2023). The effects of explanations on automation bias. *Artificial Intelligence*, 322, 103952. Retrieved November 3, 2023, from <https://linkinghub.elsevier.com/retrieve/pii/S000437022300098X>
- Vlačić, B., Corbo, L., Costa E Silva, S., & Dabić, M. (2021). The evolving role of artificial intelligence in marketing: A review and research agenda. *Journal of Business Research*, 128, 187–203. Retrieved November 3, 2023, from <https://linkinghub.elsevier.com/retrieve/pii/S0148296321000643>
- Von Eschenbach, W. J. (2021). Transparency and the Black Box Problem: Why We Do Not Trust AI. *Philosophy & Technology*, 34(4), 1607–1622. Retrieved November 6, 2023, from <https://link.springer.com/10.1007/s13347-021-00477-0>
- Voss, P., & Jovanovic, M. (2023, September 4). Concepts is All You Need: A More Direct Path to AGI. arXiv. Retrieved November 3, 2023, from <http://arxiv.org/abs/2309.01622>
- Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., & Trichina, E. (2022). Artificial intelligence, robotics, advanced technologies and human resource management: A systematic review. *The International Journal of Human Resource Management*, 33(6), 1237–1266. Retrieved November 3, 2023, from <https://www.tandfonline.com/doi/full/10.1080/09585192.2020.1871398>
- Wachter, S. (2022). The theory of artificial immutability: Protecting algorithmic groups under anti-discrimination law. *Tul. L. Rev.*, 97, 149.
- Wang, W., Lumineau, F., & Schilke, O. (2022). *Blockchains: Strategic Implications for Contracting, Trust, and Organizational Design* (1st ed.). Cambridge University Press. Retrieved November 3, 2023, from

<https://www.cambridge.org/core/product/identifier/9781009057707/type/element>

- Wang, Y., Su, Z., Zhang, N., Xing, R., Liu, D., Luan, T. H., & Shen, X. (2023). A Survey on Metaverse: Fundamentals, Security, and Privacy. *IEEE Communications Surveys & Tutorials*, 25(1), 319–352. Retrieved November 3, 2023, from <https://ieeexplore.ieee.org/document/9880528/>
- Wang, Y., Zhong, W., Li, L., Mi, F., Zeng, X., Huang, W., Shang, L., et al. (2023). Aligning Large Language Models with Human: A Survey. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2307.12966>
- Webb, M. (2019). The Impact of Artificial Intelligence on the Labor Market. *SSRN Electronic Journal*. Retrieved November 3, 2023, from <https://www.ssrn.com/abstract=3482150>
- Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., et al. (2021). Ethical and social risks of harm from Language Models. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2112.04359>
- Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.-S., Mellor, J., Glaese, A., et al. (2022). Taxonomy of Risks posed by Language Models. *2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 214–229). Retrieved November 3, 2023, from <https://dl.acm.org/doi/10.1145/3531146.3533088>
- Weking, J., Desouza, K. C., Felt, E., & Kowalkiewicz, M. (2023). Metaverse-enabled entrepreneurship. *Journal of Business Venturing Insights*, 19, e00375. Retrieved November 3, 2023, from <https://linkinghub.elsevier.com/retrieve/pii/S2352673423000045>
- Wu, J., Gan, W., Chen, Z., Wan, S., & Lin, H. (2023, March 25). AI-Generated Content (AIGC): A Survey. arXiv. Retrieved September 16, 2023, from <http://arxiv.org/abs/2304.06632>
- Wu, J., Gan, W., Chen, Z., Wan, S., & Lin, H. (2023). AI-Generated Content (AIGC): A Survey. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2304.06632>
- Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., Yin, B., et al. (2023). Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2304.13712>
- Yang, K., & Stoyanovich, J. (2016). Measuring Fairness in Ranked Outputs. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/1610.08559>
- Ye, H., Liu, T., Zhang, A., Hua, W., & Jia, W. (2023). Cognitive Mirage: A Review of Hallucinations in Large Language Models. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2309.06794>
- Yeo, W. J., van der Heever, W., Mao, R., Cambria, E., Satapathy, R., & Mengaldo, G. (2023). A Comprehensive Review on Financial Explainable AI. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2309.11960>
- Zhao, J. (2022). Artificial intelligence and corporate decisions: Fantasy, reality or destiny. *Cath. UL Rev.*, 71, 663.
- Zhou, C., Liu, P., Xu, P., Iyer, S., Sun, J., Mao, Y., Ma, X., et al. (2023). LIMA: Less Is More for Alignment. arXiv. Retrieved November 3, 2023, from <https://arxiv.org/abs/2305.11206>