ICIS 2024 Proceedings

International Conference on Information Systems (ICIS)

December 2024

Effect of Artificial General Intelligence on Organizational Learning

Yonggang Li
National University of Singapore, e0530832@u.nus.edu

Jungpil Hahn
National University of Singapore, jungpil@nus.edu.sg

Effect of Artificial General Intelligence on Organizational Learning

Completed Research Paper

Yonggang Li
National University of Singapore
11 Research Link, Singapore 119391
yonggang@comp.nus.edu.sg

Jungpil Hahn
National University of Singapore
11 Research Link, Singapore 119391
jungpil@nus.edu.sg

Abstract

With the rapid development of artificial general intelligence (AGI) organizations now have access to a novel resource to enhance its learning capabilities and performance. Through computational modeling and simulations, we study the impact of AGI on organizational learning and uncover the nuanced impacts of organizational trust in AI, the capability of the AGI in terms of its accuracy and generality, and human learning rates from the AGI. Our analyses reveal that, AI can reduce the demand for human explorative learning under appropriate levels of organizational trust in AGI. Moreover, the impact of AGI depends on its capability, increasing the accuracy of AI is largely beneficial, but adjustments on the domains that AI is not good at may be harmful to organizations. These insights contribute to rethinking organization learning in the presence of AGI technology and can aid organizations in reallocating scarce resources to facilitate organization learning in practice.

Keywords: Artificial general intelligence; organization learning; exploration and exploitation, organizational trust, accuracy, generality

Introduction

Artificial intelligence (AI) is gradually changing our society. The recent market introduction of AI based on foundation models (e.g., large language models (LLMs) such as OpenAI's ChatGPT) has ignited much enthusiasm. Some experts even argue that "sparks" of artificial general intelligence (AGI) are now evident in the latest generation of LLMs (Bubeck et al. 2023). Although there is no generally agreed upon definition of AGI, one aspect that is now broadly accepted is that AI is no longer limited to a specific domain or a specialized task, but is now a kind of general intelligence that encompasses a broad range of cognitive skills and abilities. AGI offers invaluable opportunities to improve people's lives across multiple domains, encompassing healthcare, education, employment, security, and transportation (Berente et al. 2021; Rahwan et al. 2019). AI is also poised to significantly impact organizations and the future of work.

The survival and long term viability of organizations heavily depend on their ability to learn (Argote et al. 2003; Cohen and Levinthal 1989), which in turn organizational learning relies on the capability of organization's members to adapt to the organization's external environment. However, with the continuous development and progress of science and technology, the world is gradually becoming more and more complex, mainly manifested in the explosive growth of data (Sagiroglu and Sinanc 2013) and more diversified information, from text to images to videos to virtual reality (Wohlgenannt et al. 2020). The traditional learning agents, i.e., human beings, are facing challenges in learning many facets of the increasingly complex reality due to human peculiarities such as cognitive biases, learning myopia, and bounded rationality (e.g., Levinthal and March 1993).

These challenges have been alleviated by the explosive emergence of AI technology and the adoption of these technologies which can enhance / augment human learning capabilities (Berente et al. 2021; March 2006). AI systems leverage extensive digitalized data, including digital traces and machine-generated data, to uncover insights across various domains of knowledge that might elude human cognition due to inherent cognitive limitations (Hendriks et al. 2023; Ransbotham et al. 2020). As an example, through the analysis of intricate business data, AI offered managers unparalleled prospects in conceiving new product designs, formulating innovative service propositions, and pioneering novel business models (Berente et al. 2021; Davenport et al. 2020). With the introduction of LLMs such as ChatGPT, the productivity of human beings has been greatly improved (Noy and Zhang 2023).

AI's role in this challenging world extends beyond mere tools; it can act as a supporter, an assistant, or even a counterpart for individuals engaged in organizational learning. On the one hand, the advanced development of AI makes it possible to be a learning resource for individuals. For instance, nowadays individuals can leverage the knowledge of ChatGPT for tasks such as software code programming, conducting research, analyzing data, writing professional reports, etc. On the other hand, "with their ability to learn, AI systems represent a new type of organizational resource contributing their own knowledge to an organization's stock of knowledge" (e.g., Lyytinen et al. 2021; Ransbotham et al. 2020).

This evolving role of AI is highlighted in the work of Sturm et al. (2021), which investigated how organizations can coordinate human learning and machine learning (ML) in order to improve organizational performance where AI was treated as a counterpart of individuals. In their model, organizations can learn from human agents and ML agents simultaneously. They stated that ML can reduce an organization's demand for human explorative learning that is aimed at uncovering new ideas and it is beneficial to adjust the ML system humanly even though under certain conditions this effect becomes detrimental. Conversely, Hendriks et al. (2023) studied AI as assistants of humans and explored the virtuous and vicious dynamics that affect organizational learning by conducting a series of agent-based simulations of different learning modes between humans and AI assistants in an organization. Their simulation results showed that aligning the learning of humans and AI assistants and allowing them to influence each other's learning processes equally leads to the highest organizational performance.

In previous studies (e.g., Sturm et al. 2021), AI is characterized as 'specialized,' indicating that it is typically capable of performing satisfactorily only on specific tasks. However, AI can also be 'general,' meaning it has the capacity to perform well across a variety of tasks, demonstrating a broader scope in its ability to learn from reality (Bubeck et al. 2023). The impact of such artificial *general* intelligence on organizations as assistants, remains to be explored.

However, developing advanced AGI systems capable of performing a variety of tasks can be prohibitive in terms of both costs and time. For example, OpenAI's GPT-4, which comprises 1 trillion parameters, required over $100M for its training (OpenAI et al. 2024). Consequently, many organizations opt to adopt AI technologies developed by others, which we refer to as "third-party AI." While scholars have explored how AI influences organizational learning performance from the perspectives of AI as assistant and as counterpart, the specific impacts of third-party AI on organizational learning continue to be a complex and underexplored issue.

The advancing capabilities of AI, surpassing certain human capacities, have also raised significant attention. Some scholars (e.g., Argote et al. 2021; Sturm et al. 2021) believe that AI can be used as an independent learner and join humans in organizations' search for knowledge with their own contributions. However, concerns persist regarding the transparency, or the 'black-box' nature of AI as well as the potential for biased outputs and hallucinations. These issues have led organizations to exhibit hesitancy in fully entrusting AI systems with critical tasks, opting instead for human oversight and intervention (Fügener et al. 2021). Consequently, the extent to which third-party AI solutions influence organizational learning remains an open question, particularly regarding organizational trust in AI's capability and reliability. Further exploration is warranted to understand the intricate dynamics and implications of integrating third-party AI into organizational learning processes from the trust perspective.

In this paper, our focus is on elucidating strategies for guiding organizational learning processes that incorporate third-party AGI. Specifically, we aim to investigate *how organizations can effectively coordinate human learning endeavours with the utilization of third-party AGI as assistants across varying levels of organizational trust*. As far as we know, this is the first study that combines organizational level of trust towards artificial general intelligence which serves as assistant to organization members in organization learning processes. Our objective is to discern organizational configurations that facilitate the synergistic integration of human and AGI-driven learning approaches, ultimately aiming to enhance organizational performance. This is paramount for organizations due to the fundamental role of learning in their long-term viability and success (Cohen and Levinthal 1989; Sturm et al. 2021). Lacking a comprehensive understanding of how to effectively integrate AI, including establishing appropriate levels of trust in AI as discussed in our paper, organizations may encounter inefficient allocation of their limited resources (Sturm et al. 2021).

We conducted a series of agent-based simulations on a computational model of organizational learning that extends the classical model of March (1991). Our extensions incorporate a third-party AGI as an assistant to human organizational members. Our simulations provide insights into the interplay between human learning and AI assistance within an organizational context, specifically examining how organizational trust in AI influences effectiveness of organizational learning. This understanding can assist organizations in prioritizing resource allocation towards practices and structures that contribute most significantly to effective organizational learning.

The paper unfolds as follows. First, the extant literature elucidating the influence of AI (and ML) on organizational learning is reviewed. Subsequently, our simulation model and the resultant findings are meticulously examined and interpreted. Finally, we discuss the implications of our findings, providing insights for theory and practice.

Background Literature

Organization Learning

Organizational learning encompasses the dynamic process through which knowledge is generated, preserved, and disseminated within an organization (Argote et al. 2003). As the organization accumulates and assimilates experiential insights over time, it undergoes iterative improvements, thereby enhancing its capabilities and effectiveness. This iterative cycle of experience acquisition facilitates the creation of new knowledge within the organization, contributing to its future development and evolution (Argote et al. 2021; Hendriks et al. 2023; Levitt and March 1988).

Learning fundamentally shapes the actions of organizational members and consequently the performance of the organization. However, while learning is indispensable for organizational success, it does not occur autonomously; rather, it relies on the learning endeavours of individual members, facilitated through socialization and peer influence processes (Fiol and Lyles 1985; Levitt and March 1988; March 1991).

At the heart of organizational learning lies the challenge of balancing exploration and exploitation (Argote et al. 2021; Gupta et al. 2006; March 1991). Exploration entails the pursuit of novel beliefs, such as experimenting with new strategies, while exploitation involves reinforcing established beliefs, such as relying existing ideas and strategies (Gupta et al. 2006; March 1991). To attain optimal organizational performance, organizations must strike a delicate balance between exploration and exploitation (March 1991). Excessive emphasis on either exploration or exploitation risks underutilizing developed competencies or failing to innovate and adapt to the competitive environment.

Various factors contribute to an organization's inclination towards exploration or exploitation. An organization's structure has been found to shape organizational learning preferences (Fang et al. 2010; Schilling and Fang 2014). Additionally, organizations' learning rates are subject to variation, influenced by such factors as the proficiency of individuals, advancements in organizational technology, and improvements in organizational structure, including routines and coordination methods (Argote 2013). Furthermore, inherent imperfections within learning processes also exert a notable influence on organizational learning performance (Levinthal and March 1993; Simon 1991). Organizations are also prone to "learning myopia," which denotes an organization's tendency towards exploitation over exploration, which inevitably shapes organizational learning dynamics (Koçak et al. 2023; Levinthal and March 1993).

With the growing integration of AI and ML into organizations, researchers have started to investigate the role and consequences of AI and ML systems in organizational learning (e.g., Afiouni 2019; Argote et al. 2021; Berente et al. 2021; Lyytinen et al. 2021; Ransbotham et al. 2020; Schuetz and Venkatesh 2020; Sturm et al. 2021; Teodorescu et al. 2021). Schuetz and Venkatesh (2020) and Teodorescu et al. (2021)

discuss the consequences of humans and ML systems entering closer relationships. Sturm et al. (2021) discuss how AI as counterparts of humans impact organization learning performance from the perspective of initializing and reconfiguring the AI. However, the AI in Sturm et al. (2021) is modeled as a narrow / specialized type that directly imparts knowledge to organizations, a conceptualization that is incongruent to human-in-the-loop theories. Despite this, no existing research comprehensively examines the impact of AGI as an assistant on organizational learning. This gap leads us to our central research question: *What is the impact of third-party AGI on organizational learning?* Specifically, we are interested in understanding how individuals' working behaviors might change with the introduction of third-party AGI under different levels of organizational trust towards AGI.

Artificial General Intelligence

Decades ago, several scholars defined creating AI as "making a machine behave in ways that would be called intelligent if a human were so behaving" (McCarthy et al. 2006, p. 11). Over the years, although numerous definitions of AI have emerged (Liu et al. 2018), most contemporary definitions highlight different attributes of AI, such as its learning abilities (Castelvecchi 2016), or its emulation capability, focusing on its design to replicate human skills and abilities (Brynjolfsson and Mitchell 2017).

The term 'Artificial General Intelligence' (AGI) is used to describe an AI system that is at least as capable as a human at most tasks. It often refers to as "strong AI" or "full AI." The concept of AGI is rooted in computer science. In his seminal paper, the father of modern computer science posed a question "Can machines think?" (Turing 1950). This question has guided the philosophical and technical pursuits within the field. In theory, the development of AGI requires systems that not only mimic human cognitive abilities, but also have the ability to generalize knowledge across different domains. This is in contrast to "narrow AI", which is designed for specific tasks such as facial recognition, fraud detection, product recommendations, credit loan approval, etc. The swift progress in AI models has shifted the concept of AGI which is a type of AI from a largely philosophical discussion to a matter of near-term practical importance (Morris et al. 2024). Some experts argue that the signs of AGI are already emerging in the latest generation of LLMs (Bubeck et al. 2023). These LLMs can provide knowledge that matches or even surpasses human expertise across a diverse range of fields, from art to science.

Organizational Trust in AI

Trust plays a crucial role in the adoption and utilization of AI (Lukyanenko et al. 2022). Distrust or over-trust in AI may be harmful to organization learning performance. One important factor that impacts organization trust in AI is the uncertainty that accompany the development and implementation of AI technologies. This uncertainty often results in caution, skepticism, and distrust, which are not unfounded given the notable failures of many AI systems. For instance, the COMPASS tool, intended to assist judges with release and detention decisions, was later found to exhibit bias against African-Americans (Mehrabi et al. 2021). Such failures can also stem from flawed training processes, as evidenced by IBM Watson's shortcomings in the medical field, where it was trained on hypothetical instead of real patient data, leading to unsafe treatment recommendations (Davenport and Ronanki 2018).

Algorithm aversion is also a significant factor that influences organizational trust in AI (Cabiddu et al. 2022). It arises when individuals prefer human judgment over algorithmic output, even if the algorithm performs better. This aversion stems from a lack of understanding of how the algorithm works or from negative experiences with an automated system, which can lead to a generalized distrust of all algorithms. Moreover, accountability is a crucial factor impacting organizational trust in AI. If AI is involved in a decision-making process that results in substantial losses, accountability issues will influence the organization's trust in AI, as the AI system often cannot be held responsible (Glikson and Woolley 2020). The level of trust organizations in different industries place in AI also varies. In healthcare, for instance, there is a heightened focus on AI transparency and bias-related issues. Stakeholders consciously calibrate their levels of trust in AI to avoid both over trust and under trust in AI (Asan et al. 2020). In the field of education, there is heightened scrutiny over ethical issues related to AI, such as the plagiarism of AI-generated content, leading to a general lack of trust in AI (Qin et al. 2020).

A Model of Organizational Learning with AGI

We study how the use of third-party AGI influences organizational learning and performance. Our model extends March's (1991) seminal model by incorporating a third-party AGI as an additional resource for organizational learning. First, we explicate March's (1991) model, after which we discuss our extensions.

March's (1991) Model of Organization Learning

March (1991) devised an agent-based simulation model where members of an organization progressively acquire knowledge about their environment. In this model, organizational learning is conceptualized as a reciprocal process between the organization and its members. On the one hand, the organization accumulates and retains the knowledge of its members, encapsulating it within established norms, practices, and routines, thus shaping what is termed the organizational code. On the other hand, organizational members interact with the organizational code over time. Hence, the organizational code both shapes and is shaped by individuals within the organization. Based on this premise, March (1991) posits a model consisting of three main constructs: an external reality, the organizational code and individuals (i.e., organizational members).

An *external reality* is represented as a multidimensional vector comprising m elements, where each element takes on a binary value of either 1 or -1. This vector remains undisclosed to the organization's members and remains independent of their beliefs. The initial values are randomly assigned with equal probability.

The *organization code* mirrors the organization's perceptions of external reality and is also represented as an m-dimensional vector. The organizational code stores values of 1, 0, and -1. A value of 0 for an element in the organizational code signifies that the organization remains neutral about that element. Initially, the organizational code is devoid of any beliefs regarding the reality and therefore is initialized as a vector of zeros.

The organization consists of n *individuals* who are the learning agents of the organization. Each individual holds m beliefs regarding external reality. Similar to the organizational code, these beliefs can be either 1, -1, or 0. A belief of 0 for an element signifies that the individual remains neutral about that element. Initial beliefs are randomly determined with equal probability.

Both individuals and the organizational code can harbor accurate and inaccurate beliefs, defined as beliefs that correspond or do not correspond to the respective values of the external reality. The knowledge levels of all individuals and the organizational code are assessed by calculating the percentage of beliefs that align with external reality. Organizational learning endeavors to maximize this knowledge level, aiming to increase the alignment between organizational (and individuals') beliefs and the external reality. This goal, as highlighted by March (1991) and further explored by Miller et al. (2006) and Sturm et al. (2021), underscores the pursuit of achieving the highest possible number of matches between organizational beliefs and external reality.

Over time, both individuals' beliefs and the organizational code adapt. Individuals assimilate knowledge from the organizational code: For every individual and each of their m beliefs, the belief's value transitions to the corresponding value in the organizational code with a probability denoted as p_1. If the organizational code's value is 0 for an element, individuals' beliefs for that element are not affected. This process of individuals' learning from the organizational code mirrors the socialization of individuals into the organization's belief system, encompassing norms, practices, and routines. Simultaneously, the organizational code learns knowledge from its members: For each of the m beliefs encoded in the organizational structure, the code's value can shift to the majority belief held among high performing individuals.[1] The likelihood of organizational code learning is determined by the code learning rate, denoted as p_2, and the consensus level among these more knowledgeable individuals. Organizational code learning symbolizes the adaptation of organizational norms to align with the best practices observed among the organization's members (Fang et al. 2010; March 1991; Miller et al. 2006).

March (1991) illustrated that within an organization, beliefs tend to converge over time, ultimately reaching a

[1]High performing individuals are those whose knowledge levels are higher than that of the organizational code.

stable state of knowledge. Exploitation arises when there is rapid mutual learning between individuals and the organizational code, resulting in premature convergence toward uniform beliefs, thereby steering the organization towards a sub-optimal stable state knowledge. Conversely, exploration occurs when the slower pace of learning from and by the organizational code preserves diversity in beliefs within the organization. March (1991) noted that a combination of slower individual learning from the organizational code (a low p_1) alongside rapid learning by the code (a high p_2) yields the highest average levels of knowledge attainment. The robustness of this finding has been demonstrated in subsequent works (Fang et al. 2010; Kane and Alavi 2007). This underscores the importance for organizations of striking a balance between exploration and exploitation in order to optimize their learning dynamics.

Extending March (1991) with Third-party AGI

An AGI is an AI that is not specialized to carry out a specific task, but one that can perform as broad a range of tasks as a human (Shanahan 2015). State-of-the-art LLMs (e.g., mid-2023 deployments of GPT-4, Bard, Llama 2, and Claude) can be considered as AGIs as they exbihit generality, a key property of AGI (Grudin and Jacques 2019). Because language models can cover a wide range of topics, execute a wide range of tasks, handle multimodal inputs and outputs, operate in multiple languages, and "learn" from zero-shot or few-shot examples, they can be said to have achieved sufficient generality.

In essence, an AGI can be conceptualized as having knowledge about the world (i.e., the external reality). However, it is not perfect (at least as of yet). Therefore, although the AGI has knowledge that reflects external reality, it may have incorrect knowledge about it (e.g., bias) or may not have knowledge about some aspects of external reality (e.g., hallucinations). To incorporate AGI into our model, we implement the following model extension.

In addition to the n human learning agents (i.e., individuals), an organization also introduce an AGI system with broad and general knowledge about the external reality that its members can interact with directly. This AI system is not personalized but provides the same beliefs to all individual members rather than provide personalized knowledge to different individuals members as in Hendriks et al. (2023). The AGI is represented by a m dimension vector of beliefs. Each dimension can take on a value of 1, 0 or -1. A belief of 0 for an element signifies that the AGI does not have knowledge about that aspect of reality. The capability of the AGI is modeled in 2 ways – how general its scope is and how accurate it is (within its scope). AGI's generality (or its inverse) is captured with parameter $zeros$ which represents the proportion of the m-element vector that is initialized with 0. AGI's accuracy is captured with parameter $accuracy$ which represents the ratio of beliefs that AGI possesses (i.e., those that are not initialized with 0) which are consistent with reality.

With an AGI deployed within the organization, organizational members now have a novel resource from which it can learn. Just like the way individuals assimilate knowledge from the organizational code, individuals can assimilate knowledge from the AGI: For every individual and each of their m beliefs, the belief's value transitions to the corresponding value of the AGI with a probability denoted as p_3.[2] If the AGI's value is 0, it will not impact individuals' beliefs.

Even though individuals may learn from the AGI, individuals may not always do so due to a lack of organizational trust in AI. The organization's level of trust in the AGI is captured with parameter $trust$ which represents the probability for the individuals in the organization to learn from AGI rather than from the organization code.

Simulation Parameters and Experimental Design

For default simulation parameters, we follow March (1991) and model the external reality as having 30 dimension ($m = 30$) and the organization comprises of 50 members ($n = 50$). We vary the learning rates from low to high – i.e., human agents' learning from the organization code ($p_1 = \{0.1, 0.2, \ldots, 0.9\}$), the organizational code's learning from its members ($p_2 = \{0.1, 0.2, \ldots, 0.9\}$), and members' learning from

[2]We assume that all individuals within the organization have same learning rate from AGI.

the AGI ($p_3 = \{0.1, 0.2, \ldots, 0.9\}$). We also vary the accuracy and generality of the AGI at varying levels – $accuracy = \{0.8, 0.85, 0.9, 0.95, 1.0\}$ and $zeros = \{0.2, 0.5, 0.8\}$. Organizational trust in AI is also varied from low to high levels ($trust = \{0, 0.1, 0.2, \ldots, 0.9, 1\}$).[3] Our measure of (long-term) organizational performance is *Organizational Knowledge* at equilibrium (March 1991; Sturm et al. 2021).

Our simulation model is constructed in Python using the Mesa framework (Hendriks et al. 2023) for agent-based modeling. We conduct computational experiments of all parameters in a between-factor factorial manner – i.e., $p_1(5) \times p_2(5) \times p_3(5) \times accuracy(5) \times zeros(3) \times trust(7)$, yielding $5 \times 5 \times 5 \times 5 \times 3 \times 7 = 13,125$ unique experimental conditions. Each simulation runs across 200 time steps, a duration deemed sufficient and adequate for organizational knowledge levels to converge at their ultimate equilibrium values. The organizational knowledge levels discussed hereafter pertain to these long-term values observed at the end of each simulation run. To ensure that the results of the simulation reflect the underlying structure of the model rather than a particular realization of a stochastic process, our results are based on 800 replications.

Results

Our study examines the impact of varying learning rates—p_1 for human learning from organizational code, p_2 for the organizational code learning from organizational members, and p_3 for human learning from AGI, alongside organizational trust in AGI ($trust$), on the organizational knowledge that use AGI with differing levels of accuracy ($accuracy$) and generality ($zeros$). Before delving into the findings of our research questions, we provide a summary of how these learning rates affect organizational knowledge. We contrast these effects with a baseline scenario that involves only human participants (i.e., when $trust = 0$ and the model reverts to March (1991)), to delineate the specific contributions of AGI agents to organizational learning (see Figure 1).[4]

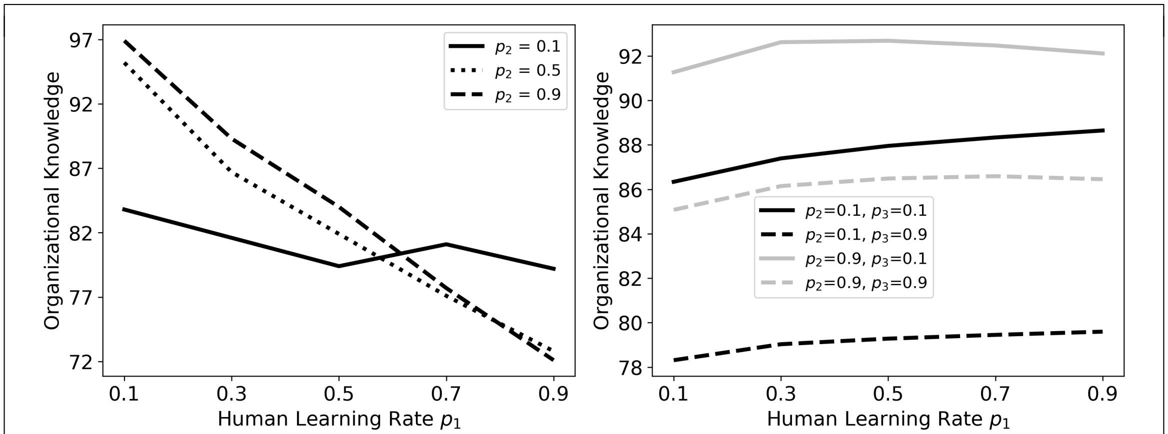

Figure 1. Effect of Human and Code Learning Rates on Organizational Knowledge

To identify AGI's general impact on human-centric organizational learning rates p_1 and p_2, we compared the effects of these rates with those of the human-only baseline model (see Figure 1). The downward lines in the

[3] When $trust = 0$, individuals only learn from the organization code (i.e., the AI is never used); our model reverts to March's (1991) model. When $trust = 1$, individuals always learn from AI and as a result does not learn from the organizational code; when $0 < trust < 1$, organizational members engage in both socialization learning with other members (through the organizational code) and also learn directly from the AGI. The AI provides humans a new resource to gain insights about the external reality.

[4] Organizational knowledge levels represent the average of all simulation runs, across varying levels of AGI generality, accuracy, and organizational trust in AGI.

left sub-figure of Figure 1 indicate that low values of the human learning rate p_1 are superior to high values in the baseline model, which reflects the strong dependence of organizational learning on human exploration (Sturm et al. 2021) to reach higher levels of organizational knowledge. This is no longer the case when AGI is introduced into the model, as illustrated by the right sub-figure, where human exploration (i.e., low p_1) is no longer superior to human exploitation (i.e., high p_1). Interestingly, we observe that organizational knowledge levels are higher when the learning rate of humans from the AGI is low (i.e., $p_3 = 0.1$).

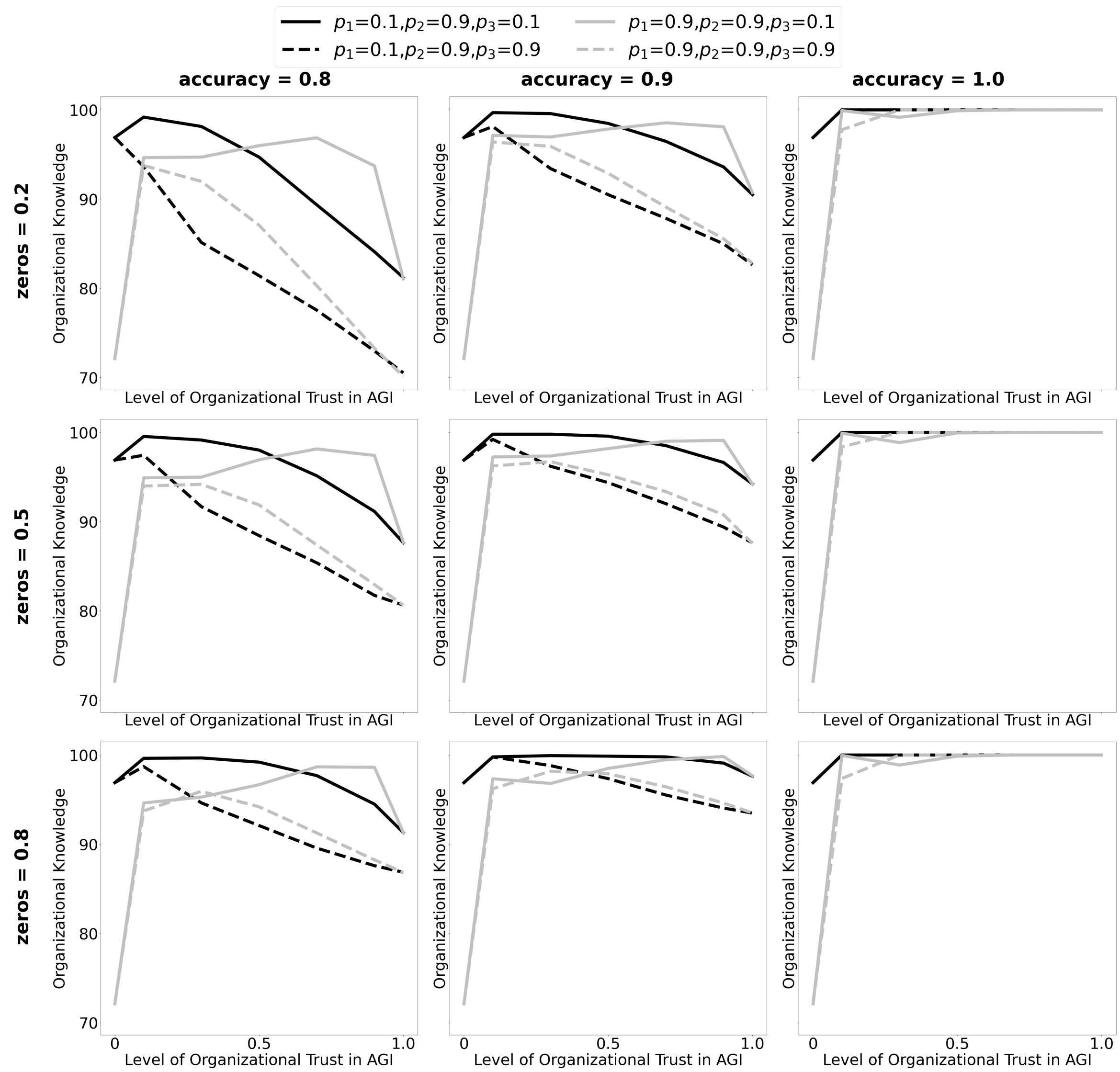

Figure 2. Effect of Organizational Trust in AGI on Organizational Learning

The Impact of Organizational Trust in AI on Organizational Learning

We first consider the impact of organizational trust towards third-party AGI on organizational learning performance. For this purpose, we compare the organizational knowledge levels for different configurations of organization trust in AI.

Figure 2 shows the organizational knowledge at equilibrium with and without the introduction of the AGI assistant at different levels of organizational trust in AI. We also consider the capabilities of the AGI through its generality ($zeros$) and accuracy ($accuracy$). Specifically, across each column of sub-figures, the $accuracy$ of AGI ranges from $\{0.8, 0.9, 1.0\}$ and across each row of sub-figures, the AGI's generality ranges from $\{0.2, 0.5, 0.8\}$. Within each sub-figure, the x-axis is the level of organizational trust in AGI, and the y-axis is the organization equilibrium knowledge level and each line represents a configuration of organization learning rates (i.e., $(p_1, p_2, p_3) = \{(0.1, 0.9, 0.1), (0.1, 0.9, 0.9), (0.9, 0.9, 0.1), (0.9, 0.9, 0.9)\}$) which are the four best configurations under different values of $zeros$ and $accuracy$ in each corresponding sub-figure. When the accuracy of the AI is relatively high ($accuracy = 0.8$ or 0.9), the organizational knowledge levels in the no-trust ($trust = 0$) and the full-trust ($trust = 1$) conditions is lower than with partial trust in the third-party AGI ($0 < trust < 1$; see the grey solid line increase first and then decline with increasing organization trust in the AGI). This means that the impact of introducing high-performance third-party AGI depends on organizations' trust in it – an appropriate level of organization trust in AGI is essential. It also shows that when the AGI is not perfect ($accuracy < 1$), individuals should learn from AGI and socialization at the same time to achieve higher organizational performance.

When organizational trust in AGI is relatively low, the optimal configuration for achieving maximal organizational equilibrium knowledge is when $p_1 = 0.1, p_2 = 0.9$ and $p_3 = 0.1$. This result resonates with March's (1991) original findings that a combination of slower individual learning from the organizational code (a low p_1) alongside rapid learning by the code (a high p_2) yields the highest average levels of knowledge attainment. Interestingly, with the incorporation of AGI as an additional knowledge resource, the same organizational learning configuration (i.e., low p_1 and high p_2) still produces the highest organizational performance and also that slower individual learning from the AGI (a low p_3) is in fact ultimately more beneficial for the organization. Conversely, when organizational trust in AGI is relatively high, the optimal configuration for achieving maximal organizational knowledge is when $p_1 = 0.9, p_2 = 0.9$ and $p_3 = 0.1$. This indicates that individuals should adopt beliefs from AGI slowly, but should rapidly learn from the organizational code, while the organizational code should continue to rapidly adopt beliefs from individuals. It suggests that an AGI assistant can redirect humans from exploration to a more exploitative role. In essence, an organization utilizing AGI assistants can relieve its human members of the need to explore, without risking significant losses in adaptability that might hinder long-term knowledge.

Comparing the left two columns of sub-figures with the right column of sub-figures we can see that when the performance of the AGI is (theoretically) perfect (i.e., AGI's $accuracy = 1$), high level of organizational trust in AGI will make the difference of distinct configurations less significant (i.e., the vertical gaps between different configurations are decreasing), which means that organizations can always gain a higher level of organizational knowledge with perfect AGI. However, perfect AGI is currently not a possibility.

Overall, it is inappropriate for organizations to either rely entirely on AI for learning or disregard it completely. The unique knowledge (Berente et al. 2021) that AI can provide may be valuable, depending on how the organization chooses to integrate and utilize it. Based on these insights, we propose the following:

Proposition 1a: When organizational trust in AGI assistants is high, humans' efforts on exploration can be liberated (i.e., p_1 does not have to be low) without sacrificing an organization's long-term knowledge.

Proposition 1b: When and only when AGI is perfect (i.e., $accuracy = 1$), total trust in AGI always benefits organizational knowledge. As perfect AGI is impossible, organizations should not totally trust (or downright dismiss) AGI to achieve optimal long-term organizational knowledge.

The Impact of AI Accuracy on Organizational Learning

It is important to recognize that the impact of AI on organizational learning performance is mediated through the individual members of the organization.

In Figure 3, each column of sub-figures represents increasing levels of organizational trust in AGI, ranging from 0.1 to 0.9. Concurrently, each row of sub-figures depicts an increase in AGI's generality, indicated by the parameter $zeros$, from 0.2 to 0.8. Within each sub-figure, the x-axis shows the accuracy of AGI, while the y-axis tracks the equilibrium knowledge within the organization. Each sub-figure contains four lines, each corresponding to different configurations of parameters p_1, p_2, and p_3 and a dotted line indicating the optimal organizational knowledge achieved in the model of March (1991) (i.e., $p_1 = 0.1$ and $p_2 = 0.9$).

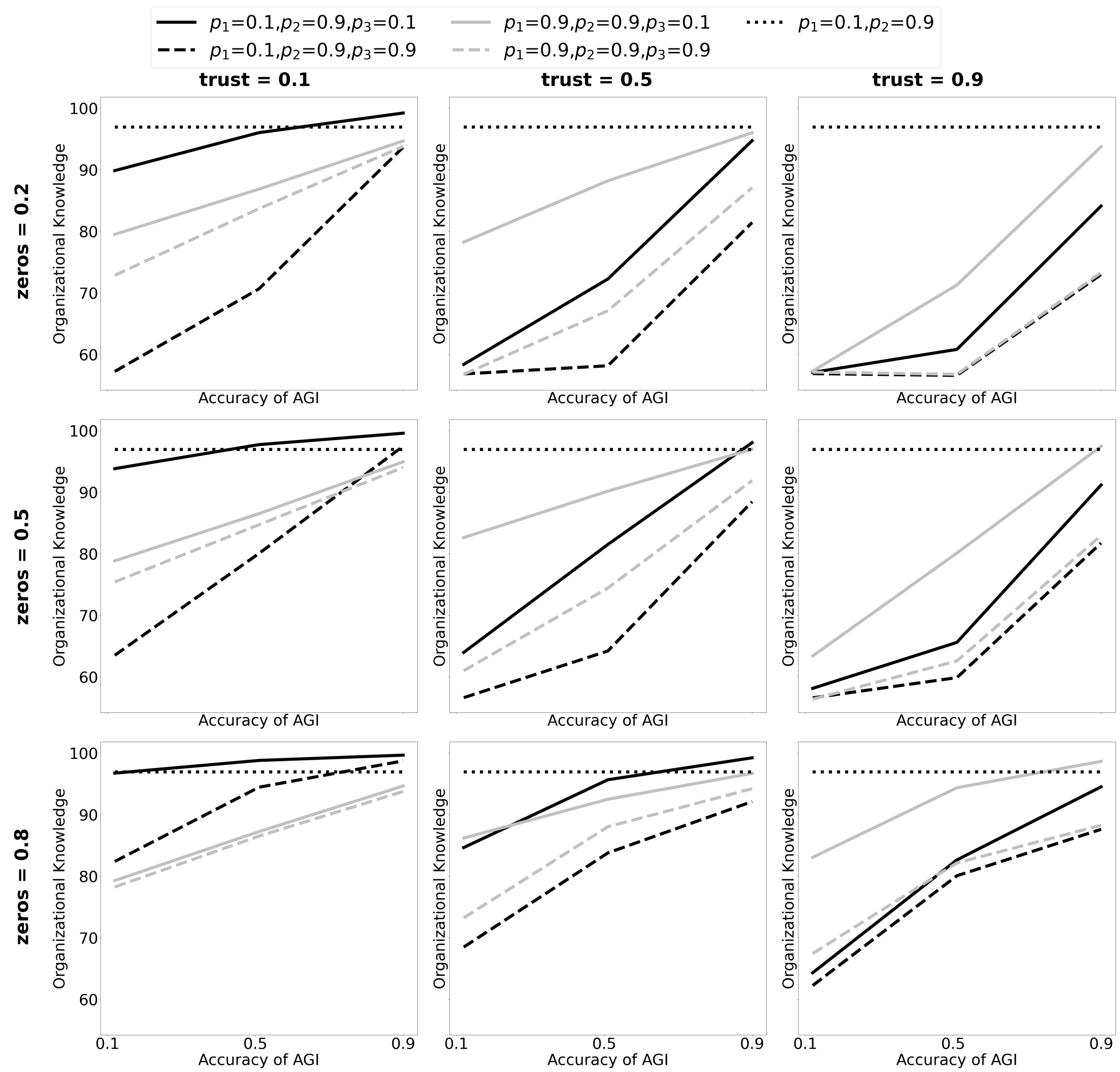

Figure 3. Effect of AI Accuracy on Organizational Learning

The baseline model does not involve learning from AI, so $accuracy$ of AGI has no impact on organizational knowledge, i.e., the dotted line is parallel to the x-axis. The choice to focus on these specific parameter configurations is driven by the observation that the highest organizational knowledge is achieved when $p_2 = 0.9$ (i.e., fast learning by the organizational code). This setup allows for the exploration of how variations in human interaction – whether through exploration or exploitation – with AGI and organizational code influence organizational knowledge when $p_2 = 0.9$ and we want to examine how human exploration (or exploitation) on AGI and organizational code will impact organizational knowledge. The result in Figure 3 shows a general trend that the increase of AGI's $accuracy$ (e.g., achieved through the use of more high-quality data during the AGI's agents' setup stage) will increase the optimal organizational knowledge achieved with different configurations.

Moreover, as indicated by the vertical gaps between different lines in each sub-figure of Figure 3, the human learning behavior (i.e., the learning rate p_1 (from organizational code) and p_3 (from AGI)[5] makes difference in the level of organizational knowledge for different level of AGI $accuracy$. Generally, the gaps are smaller with increasing AGI accuracy and conducting exploration on AGI is more beneficial than conducting exploitation on AGI (see the solid lines in each sub-figure with $p_3 = 0.1$). Recall that AGI characterized by low accuracy are prone to disseminating incorrect beliefs within an organization. Intensive human exploitation on such an AGI (indicated by a high p_3 can intensify the adoption of these incorrect beliefs, thereby degrading the level of organizational knowledge (see the start point of dashed lines in each sub-figure with $p_3 = 0.9$).

Notably, from the lower-left sub-figure in Figure 3, we can observe that organizations can even derive benefits from AGI with relatively low accuracy (e.g., when AGI's $accuracy = 0.5$) given an optimal configuration (indicated by the black solid line), where $trust = 0.1, p_1 = 0.1, p_2 = 0.9$, and $p_3 = 0.1$. This could be attributed to the complementary effect of AGI in situations of low trust. Even when AGI's performance is suboptimal, some of its insights may still be beneficial to organizational members. This indicates that incorporating a low-accuracy AGI into the organization can surpass the best organization equilibrium knowledge level previously achieved without AGI (indicated by the dotted line), as established in the March (1991) model. However, even with a relatively high-accuracy AGI ($accuracy = 0.8$), if the organizational configuration is not well-suited (e.g., when $trust = 0.9$, $p_1 = 0.9$, $p_2 = 0.9$, and $p_3 = 0.1$), the organization may not derive benefits from the AGI. This scenario is illustrated by the grey solid line in the top-right sub-figure. In sum, how correct knowledge of the AI gets disseminated effectively into the organization to promote organizational knowledge improvement is critical even when the AI does not exhibits high accuracy; the organization should maintain skepticism over AGI (i.e., lower its trust), and leverage exploration on AGI, indicated by the previously mentioned black solid line in lower-left sub-figure.). Consequently, we propose:

Proposition 2a: Humans' learning behaviors (from the AGI and from the organizational code) influence the positive and nonlinear effect of AGI's accuracy on organizational knowledge.

Proposition 2b: The impact of AGI accuracy on organizational knowledge needs to be combined with, the organization's trust in AI and the organization's members' learning behaviors. In other words, AGI's accuracy on its own is insufficient. It is necessary to consider organizational trust in AGI and humans' learning behaviors to achieve maximal long-term organizational knowledge.

The Impact of AGI Generality on Organizations' Knowledge

Next, to explore how improvements in domains where AGI is less proficient affect organizational knowledge, we will analyze the learning dynamics under varying values of $zeros$. This investigation will help us understand the influence of enhancing AGI capabilities in its weaker domains on the overall learning and knowledge accumulation within the organization. From Figure 4, we observe that when organizational trust in AGI is low, the benefits derived from refining AGI capabilities in domains that AGI is not good at are relatively modest across various learning strategies. It is also observable that in contexts where AGI performance is sub-optimal and organizational trust in the AGI is lacking, there may be a strategic advantage for the organization to maintain certain incorrect beliefs: a moderate refinement in domains where AGI is less proficient leads to higher long-term levels of organizational knowledge than do small or large refinements. This is particularly beneficial if there is rapid reciprocal learning between humans and the organization code, while human agents assimilate knowledge from the AGI at a slower pace (i.e., the grey solid line with $p_1 = 0.9, p_2 = 0.9, p_3 = 0.1$ in the third sub-figure of the second row). Maintaining these false beliefs, instead of transitioning to a state of no beliefs, potentially enlarges the exploration space for human agents, allowing for a broader scope of solution discovery and problem-solving. Another possible explanation is that retaining incorrect beliefs in the AGI could help maintain the diversity of beliefs among human agents by slowing the convergence of their beliefs (March 1991). Thus, we propose:

Proposition 3: With different levels of organizational trust in AGI assistants, for AGI assistant with high

[5]The organizations always rapidly adopt beliefs from its human members.

accuracy, humans' learning behaviour moderates the nonlinear effect of the reconfiguration in domains where AGI is less proficient on organizational learning. If humans engage in exploration on the organizational code (low p_1) and exploitation on the AGI assistant (high p_3), this effect on organizational learning effectiveness decreases in strength when eliminating *incorrect beliefs*. If humans engage in exploitation on the organizational code (high p_1) and exploration on the AGI assistant (low p_3), the refinement has an inverted U-shaped effect on organizational learning effectiveness.

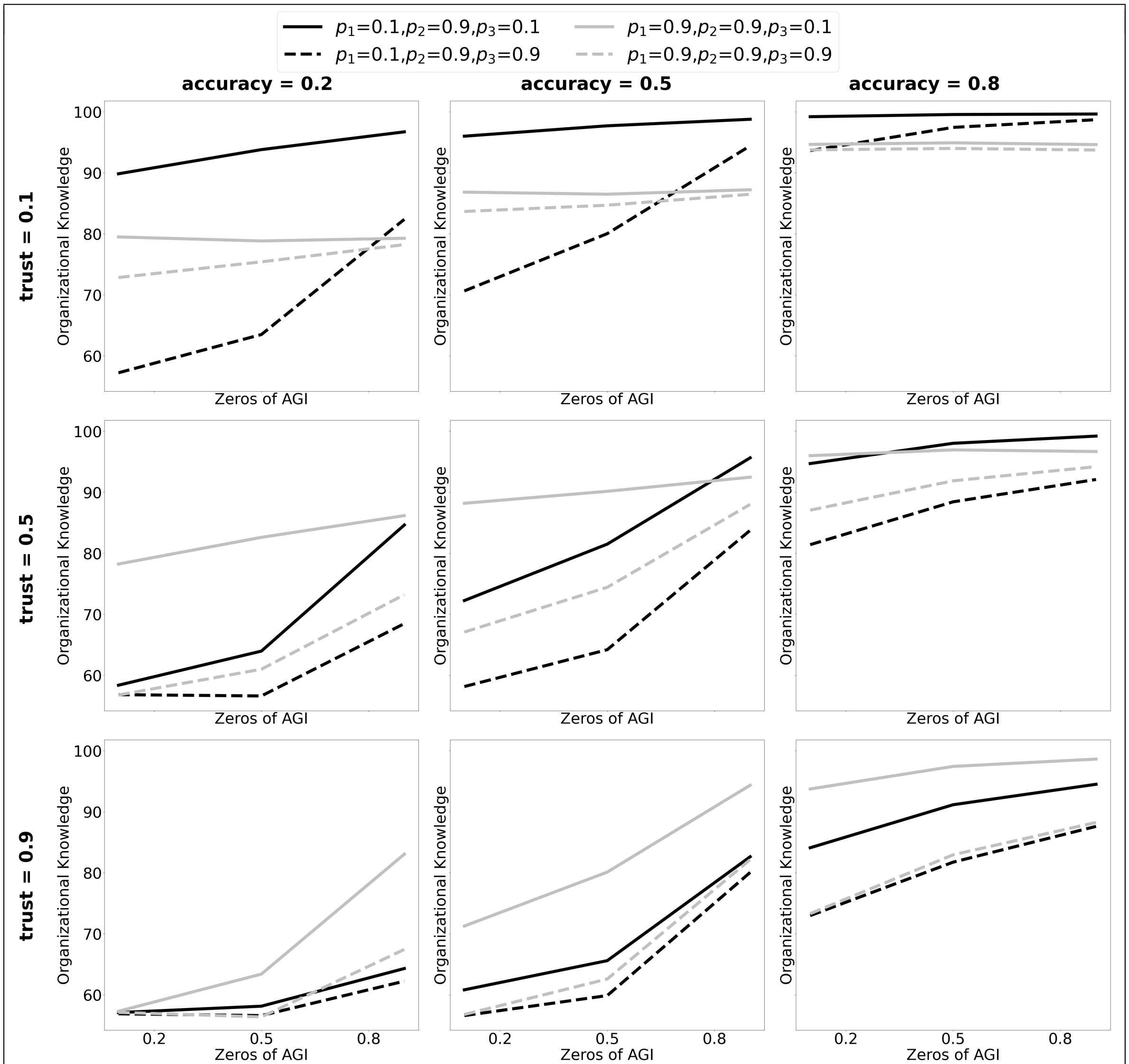

Figure 4. Effect of AI Generality on Organizational Learning

The Impact of AGI Learning Rate on Organizational Learning

Finally, to explore the impact of human learning behaviours with an AGI assistant, we analyzed variations in organizational knowledge levels in response to different learning rates (p_1, p_2, p_3), levels of organizational trust in the AGI assistant, and the AGI's $accuracy$. In Figure 5, each column of sub-figures shows the AGI accuracy increasing from 0.2 to 0.8, while each row illustrates varying levels of human trust in AGI from 0.1 to 0.9. Within each sub-figure, the x-axis represents variations in the parameter p_3, and the y-axis indicates the levels of organizational knowledge. Each line corresponds to a different combination of parameters p_1 and p_2, which reflect the exploration and exploitation dynamics (i.e., low and high p_1, respectively) of humans on organizational code and the rate at which the organizational code learns from human members.

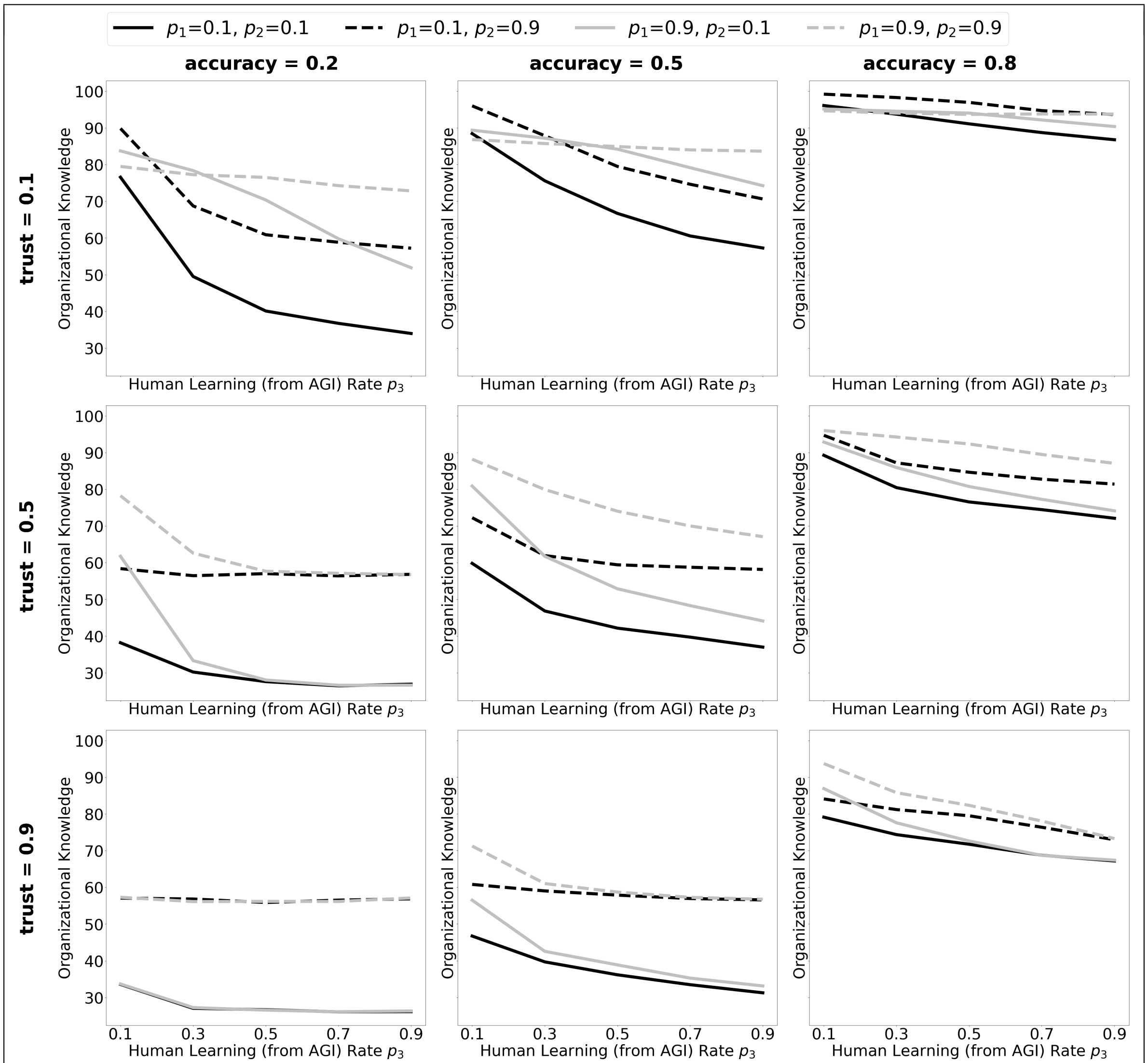

Figure 5. Effect of Learning Rate from AI on Organizational Learning

Our initial findings shown in Figure 5 suggest that when organizational trust in the AGI assistant is relatively high, conducting exploratory activities on the AGI can significantly enhance organizational knowledge, particularly when there is a mutual and rapid exchange of knowledge between the organizational code and the individuals. Conversely, in environments where organizational trust in the AGI assistant is low, exploratory activities can still augment organizational knowledge. However, this tends to occur when organizations quickly absorb insights from individuals, while individuals themselves learn more slowly from the organizational code (indicated by a low p_1, signifying a greater emphasis on exploration). The explanation could be that pursuing exploration on AGI (low p_3) especially when the performance of AGI is high might decrease the speed that the beliefs converge to incorrect ones and increase the likelihood of converging to correct beliefs. This study underscores the nuanced role that trust and learning rates play in leveraging AGI assistants for organizational learning.

Proposition 4: With a well-established trust in AGI, when the organization rapidly integrates beliefs from human agents, humans' exploration of AGI capabilities (low p_3) generally enhances organizational knowledge.

Conclusion

The integration of AI into organizational learning processes has great potential to yield utility. Leveraging AI for organizational learning consistently proves to be advantageous, albeit with a conscientious approach. Upon surpassing a certain threshold of performance, exclusive reliance on AI becomes a plausible scenario. AI demonstrates its potential to alleviate the necessity for extensive human exploration or exploitation endeavors within an organizational context.

While accuracy of the AI stands as a pivotal factor, the cultivation of appropriate trust in AI also emerges as paramount. Establishing a balanced level of trust is crucial for fostering effective utilization of AI within organizational frameworks. Endeavors aimed at enhancing accuracy and ensuring the transparency and reliability of AI systems are of equal significance. Prioritizing efforts to refine accuracy and bolster the clarity of AI contributes substantially to its effectiveness and trustworthiness within organizational settings.

For managers, our study suggests that managers should decide on investing in AI or investing in individuals. Investment in humans or AI may incur upside potential (complementary to each other), failure risks (AI may be incorrect due to bias) and investment costs (belief improvements are marginal). Organizational members should have thoughts in mind of collecting which beliefs from AI or peers, or by self-learning. Which of this knowledge will last longer? The properties of the knowledge held by AI and peers and self-learning are distinct. Adopting beliefs from AI is instantaneous, but the quality of the beliefs can be uncertain. AI is helpful for human tacit knowledge diffusion. It may also change how people interact with each other due to the knowledge that can only be transferred through AI to humans. Our study's findings may alter people's work practices in organizations, i.e., people may learn from AI instead of learning from practice, experiences or colleagues.

This study is not without limitations. The relationship between the levels of organizational trust in AGI and the AI's performance might be inherently correlated. Although this correlation may not be invariably positive, it is generally observed that superior AI performance can lead organizations to place higher trust in AI technologies. Our analyses does not account for the heterogeneity in how individuals within an organization learn from AI. Individuals subjected to pressures of potential obsolescence might display a greater propensity to acquire knowledge from AI (i.e., high p_3). Conversely, highly skilled individuals may exhibit reluctance to engage with AI learning (i.e., low p_3), due to concerns about the homogenization that might result from widespread AI adoption. Our study also does not account for the capability of AGI to accurately and promptly capture environmental turbulence – a critical feature that can significantly impact organizational learning in dynamic settings. In future research, we plan to investigate how this specific feature of AGI influences organizational learning outcomes in turbulent environments.

Acknowledgements

The authors acknowledge the generous support from the AI Governance Pillar at AI Singapore.

References

Afiouni, R. 2019. "Organizational Learning in The Rise of Machine Learning," in: *Proceedings of the 40th International Conference on Information Systems,* H. Krcmar, J. Fedorowicz, W. F. Boh, J. M. Leimeister, S. Wattal (eds.), Munich, Germany.

Argote, L. 2013. *Organizational Learning: Creating, Retaining and Transferring Knowledge,* 2nd ed., New York: Springer.

Argote, L., Lee, S., and Park, J. 2021. "Organizational Learning Processes and Outcomes: Major Findings and Future Research Directions," *Management Science* (67:9), pp. 5399–5429.

Argote, L., McEvily, B., and Reagans, R. 2003. "Managing Knowledge in Organizations: An Integrative Framework and Review of Emerging Themes," *Management Science* (49:4), pp. 571–582.

Asan, O., Bayrak, A. E., and Choudhury, A. 2020. "Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians," *Journal of Medical Internet Research* (22:6), e15154.

Berente, N., Gu, B., Recker, J., and Santhanam, R. 2021. "Managing Artificial Intelligence." *MIS Quarterly* (45:3), pp. 1433–1450.

Brynjolfsson, E. and Mitchell, T. 2017. "What Can Machine Learning Do? Workforce Implications," *Science* (358:6370), pp. 1530–1534.

Bubeck, S. et al. 2023. *Sparks of Artificial General Intelligence: Early Experiments with GPT-4.* (arXiv: 2303.12712). arXiv. https://doi.org/10.48550/arXiv.2303.12712.

Cabiddu, F., Moi, L., Patriotta, G., and Allen, D. G. 2022. "Why Do Users Trust Algorithms? A Review and Conceptualization of Initial Trust and Trust over Time," *European Management Journal* (40:5), pp. 685–706.

Castelvecchi, D. 2016. "Can We Open The Black Box of AI?," *Nature News* (538:7623), p. 20.

Cohen, W. M. and Levinthal, D. A. 1989. "Innovation and Learning: The Two Faces of R&D," *The Economic Journal* (99:397), pp. 569–596.

Davenport, T., Guha, A., Grewal, D., and Bressgott, T. 2020. "How Artificial Intelligence Will Change The Future of Marketing," *Journal of the Academy of Marketing Science* (48:1), pp. 24–42.

Davenport, T. H. and Ronanki, R. 2018. "Artificial Intelligence for The Real World," *Harvard Business Review* (96), pp. 108–116.

Fang, C., Lee, J., and Schilling, M. A. 2010. "Balancing Exploration and Exploitation Through Structural Design: The Isolation of Subgroups and Organizational Learning," *Organization Science* (21:3), pp. 625–642.

Fiol, C. M. and Lyles, M. A. 1985. "Organizational Learning," *Academy of Management Review* (10:4), pp. 803–813.

Fügener, A., Grahl, J., Gupta, A., and Ketter, W. 2021. "Will Humans-in-the-Loop Become Borgs? Merits and Pitfalls of Working with AI," *MIS Quarterly* (45:3), pp. 1527–1556.

Glikson, E. and Woolley, A. W. 2020. "Human Trust in Artificial Intelligence: Review of Empirical Research," *Academy of Management Annals* (14:2), pp. 627–660.

Grudin, J. and Jacques, R. 2019. "Chatbots, Humbots, and The Quest for Artificial General Intelligence," in: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems,* Glasgow Scotland, UK: ACM, pp. 1–11.

Gupta, A. K., Smith, K. G., and Shalley, C. E. 2006. "The Interplay Between Exploration and Exploitation," *Academy of Management Journal* (49:4), pp. 693–706.

Hendriks, P., Sturm, T., Olt, C. M., and Buxmann, P. 2023. "The Impact of Human-Artificial Intelligence Partnerships on Organizational Learning," *Publications of Darmstadt Technical University, Institute for Business Studies (BWL)* (138376).

Kane, G. C. and Alavi, M. 2007. "Information Technology and Organizational Learning: An Investigation of Exploration and Exploitation Processes," *Organization Science* (18:5), pp. 796–812.

Koçak, Ö., Levinthal, D. A., and Puranam, P. 2023. "The Dual Challenge of Search and Coordination for Organizational Adaptation: How Structures of Influence Matter," *Organization Science* (34:2), pp. 851–869.

Levinthal, D. A. and March, J. G. 1993. "The Myopia of Learning," *Strategic Management Journal* (14:S2), pp. 95–112.

Levitt, B. and March, J. G. 1988. "Organizational Learning," *Annual Review of Sociology* (14:1), pp. 319–338.

Liu, J., Kong, X., Xia, F., Bai, X., Wang, L., Qing, Q., and Lee, I. 2018. "Artificial Intelligence in the 21st Century," *IEEE Access* (6), pp. 34403–34421.

Lukyanenko, R., Maass, W., and Storey, V. C. 2022. "Trust in Artificial Intelligence: From a Foundational Trust Framework to Emerging Research Opportunities," *Electronic Markets* (32:4), pp. 1993–2020.

Lyytinen, K., Nickerson, J. V., and King, J. L. 2021. "Metahuman Systems = Humans + Machines That Learn," *Journal of Information Technology* (36:4), pp. 427–445.
March, J. G. 1991. "Exploration and Exploitation in Organizational Learning," *Organization Science* (2:1), pp. 71–87.
March, J. G. 2006. "Rationality, Foolishness, and Adaptive Intelligence," *Strategic Management Journal* (27:3), pp. 201–214.
McCarthy, J., Minsky, M. L., Rochester, N., and Shannon, C. E. 2006. "A Proposal for The Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955," *AI Magazine* (27:4), pp. 12–14.
Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. 2021. "A Survey on Bias and Fairness in Machine Learning," *ACM Computing Surveys* (54:6), 115:1–115:35.
Miller, K. D., Zhao, M., and Calantone, R. J. 2006. "Adding Interpersonal Learning and Tacit Knowledge to March's Exploration-Exploitation Model," *Academy of Management Journal* (49:4), pp. 709–722.
Morris, M. R., Sohl-dickstein, J., Fiedel, N., Warkentin, T., Dafoe, A., Faust, A., Farabet, C., and Legg, S. 2024. *Levels of AGI: Operationalizing Progress on the Path to AGI*. (arXiv:2311.02462).arXiv. https://doi.org/10.48550/arXiv.2311.02462.
Noy, S. and Zhang, W. 2023. "Experimental Evidence on The Productivity Effects of Generative Artificial Intelligence," *Science* (381:6654), pp. 187–192.
OpenAI et al. 2024. *GPT-4 Technical Report*. (arXiv:2303.08774). arXiv. https://doi.org/10.48550/arXiv.2303.08774.
Qin, F., Li, K., and Yan, J. 2020. "Understanding User Trust in Artificial Intelligence-based Educational Systems: Evidence from China," *British Journal of Educational Technology* (51:5), pp. 1693–1710.
Rahwan, I. et al. 2019. "Machine behaviour," *Nature* (568:7753), pp. 477–486.
Ransbotham, S., Khodabandeh, S., Kiron, D., Candelon, F., Chu, M., and LaFountain, B. 2020. *Expanding AI's Impact with Organizational Learning*. Tech. rep. MIT Sloan Management Review and Boston Consulting Group.
Sagiroglu, S. and Sinanc, D. 2013. "Big Data: A Review," in: *Proceeding of 2013 International Conference on Collaboration Technologies and Systems (CTS),* G. C. Fox and W. W. Smari (eds), San Diego, CA, pp. 42–47.
Schilling, M. A. and Fang, C. 2014. "When Hubs Forget, Lie, and Play Favorites: Interpersonal Network Structure, Information Distortion, and Organizational Learning," *Strategic Management Journal* (35:7), pp. 974–994.
Schuetz, S. and Venkatesh, V. 2020. "The Rise of Human Machines: How Cognitive Computing Systems Challenge Assumptions of User-system Interaction," *Journal of the Association for Information Systems* (21:2), pp. 460–482.
Shanahan, M. 2015. *The Technological Singularity,* Cambridge, MA: MIT Press.
Simon, H. A. 1991. "Bounded Rationality and Organizational Learning," *Organization Science* (2:1), pp. 125–134.
Sturm, T., Gerlach, J. P., Pumplun, L., Mesbah, N., Peters, F., Tauchert, C., Nan, N., and Buxmann, P. 2021. "Coordinating Human and Machine Learning for Effective Organizational Learning," *MIS Quarterly* (45:3), pp. 1581–1602.
Teodorescu, M., Morse, L., Awwad, Y., and Kane, G. 2021. "Failures of Fairness in Automation Require A Deeper Understanding of Human-ML Augmentation," *MIS Quarterly* (45), pp. 1483–1500.
Turing, A. M. 1950. "Computing Machinery and Intelligence," *Mind* (LIX:236), pp. 433–460.
Wohlgenannt, I., Simons, A., and Stieglitz, S. 2020. "Virtual Reality," *Business & Information Systems Engineering* (62:5), pp. 455–461.