ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/362155468

Bringing Artificial Intelligence to Business Management (Forthcoming at Nature Machine Intelligence)

Article *in* Nature Machine Intelligence · July 2022

CITATIONS
4

READS
8,572

4 authors, including:

Yash Raj Shrestha
University of Lausanne

67 PUBLICATIONS **1,253** CITATIONS

SEE PROFILE

All content following this page was uploaded by Yash Raj Shrestha on 21 July 2022.

The user has requested enhancement of the downloaded file.

Bringing Artificial Intelligence to Business Management

– Forthcoming at Nature Machine Intelligence –

Stefan Feuerriegel[1], Yash Raj Shrestha[2,3], Georg von Krogh[2], Ce Zhang[4]

[1] Institute of AI in Management, LMU Munich School of Management, LMU Munich, Munich, Germany

[2] Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

[3] Faculty of Business and Economics (HEC), University of Lausanne, Lausanne, Switzerland

[4] Institute for Computing Platforms, Department of Computer Science, ETH Zurich

Teaser: Artificial intelligence (AI) can support managers by effectively delegating management decisions to AI. Yet the road ahead is bumpy at best, with many organizational and technical hurdles that need to be mastered. We offer a first step on this journey by unpacking core factors that may hinder and foster effective decision delegation to AI.

Main

Despite widespread optimism and some early success, applications of AI in management remain limited to a relatively small subset of – mostly routine – decisions. The key to understanding the promise and challenges for AI in management is to focus on the problem of delegation of decisions by managers to AI [1,2,3,4]. Consider the problem of delegation in mergers & acquisitions by companies, where decisions may imply legal and ethical concerns that are best deliberated by managers [5,6]. Here, AI will provide input to human decision-making while the manager will remain in the “driver seat”. In effect, AI substitutes individual decisions or tasks (rather than entire positions), thereby complementing rather than substituting management decisions [2,4]. Moreover, both humans and AI have relative strengths, where humans excel in intuition, empathy, broad judgment, and complex reasoning. Hence, for companies to benefit from AI, the focus needs to shift towards promoting a full or partial delegation of decisions that augment managers.

Overcoming hurdles in delegating management decisions to AI

In the following, we discuss three salient organizational and technical hurdles facing AI in delegation of management decisions, and show potential solutions to overcome them. These hurdles cover (1) the managerial role, (2) the decision-making process, and its relationship with (3) the organization, thus encompassing the entire realm of genesis, process, and impact of management decisions (see **Box** for glossary).

Box. Glossary of AI terminology in management.

Term	**Definition**
Artificial intelligence	Artificial intelligence refers to machines performing cognitive functions that are usually associated with human minds, such as learning, interacting, and problem solving [2] (based on a definition by [27]). This gives rise to contemporary learning systems that can learn, adapt and act autonomously. This includes machine learning algorithms and its advanced instantiations of deep learning algorithms.
Human-in-the-loop	Human-in-the-loop is defined as an algorithm or model that requires human interaction. In the context of AI, the model may generate some output that the human can then override, or the AI may query the human for additional input [28].
Algorithmic bias	Algorithmic bias results from algorithmic outcomes that unfairly advantage some groups of people while disadvantaging others [3,29]
Algorithmic fairness	Algorithmic fairness subsumes mathematical approaches to uncover and prevent models' outcomes from exhibiting algorithmic bias, typically via alterations to standard models [29].
Algorithm aversion	Human tendency to prefer human forecasters over an algorithm despite the latter being more accurate in its predictions than the former [15].
Interpretability	Interpretability in machine learning refers to an 'explanation' that helps understanding how a model works [30]. Sometimes, authors further distinguish interpretable (when the decision logic itself is transparent) vs. explainable (when insights are based on a surrogate model).
Governance	In the context of AI, governance is a system of rules, practices, processes, and technological tools that are employed to ensure an organization's use of AI technologies aligns with the organization's strategies, objectives, and values; fulfills legal requirements; and meets principles of ethical AI followed by the organization [31].
Risk management frameworks	Risk management framework typically comprises of management tools, processes, and practice for (a) risk identification, (b) risk analysis, (c) risk-reducing measures, and (d) risk monitoring, aimed at protecting organizational assets from all external (e.g. natural disasters) as well as internal (e.g. technical failures, sabotage, unauthorized access) threats. The aim is that the costs of losses resulting from the realization of such threats are minimized [32,33].

Weak accountability of AI decisions

Classical literature in organizational design and decision rights provide theoretical conceptualizations of decision-making within organizations, their allocation, and delegation [7,8]. Foundational insights available in this literature inform how to design incentives and authority structures when decisions are delegated within companies. AI applications in companies and corresponding research on those applications provide a novel twist on this classic literature by addressing how existing insights and advice on organizational decision-making and delegation may change when organizations transform from all humans to human-AI hybrids [1, 2, 4].

A clear challenge in delegating decisions to AI rather than humans resides in accountability. When AI outputs decisions that are erroneous, accountability for those decisions still rests with managers. Such accountability raises significant governance issues for the company—in particular when decision errors have broad, long-term, and irreversible legal, ethical, financial, and strategic implications. For instance, Amazon recently applied an AI-based tool to make decisions on hiring new employees. The tool came under intense public scrutiny and was eventually abandoned, because it forged an unintentional bias and produced outputs that were unfair to women [9].

Managing accountability risks due to AI decision delegation becomes especially challenging in large organizations where manifold organizational roles, business units, departments, and stakeholders are simultaneously involved in the development and application of AI. Such difficulty is complicated by the fact that managers are often unaware of the ethical implications of AI decisions or fail to raise moral awareness at the time of the decision—a situation known as *ethical fading* [10]. The Amazon example illustrates this well: Who is to blame for the adverse outcomes? The AI engineer developing the tool, the manager overseeing the development, the HR department manager responsible for the hiring process, or the chief legal officer who didn’t spot the legal risks? Eventually, blame for weak accountability falls on the upper echelons of the company—the CEO, the CIO, or chairperson—although AI-based tools are used at much lower and operational levels.

While technical approaches are available to reduce the risk of AI-induced errors, these are not a magic bullet for eliminating weak accountability in delegating business decisions to AI. Instead, managers—in principle—need to accept accountability if something goes wrong. A way forward lies in studying where the AI intervention in the organization begins and where it ends.

Companies should track AI decision delegation, monitor where AI is deployed, and then proceed to assess the risks inherent in such arrangements. Additionally, enforcement of new and existing legal frameworks (e.g. Equal Credit Opportunity Act, EU AI Act) as well as collective efforts across various levels and departments in the organizations are required [11]. This is particularly important when AI makes predictions that impact people, for example in human resource management where AI may be deployed to monitor and appraise employees' performance. Hence, if the implications of AI for people cannot be controlled or the AI application is subject to high accountability risks, decision delegation to AI should not happen.

Finally, companies need to embed and regularly update governance structures (e.g., AI oversight boards, external AI audits) in order to overcome the hurdle of weak accountability. By now, several companies have implanted oversight boards for AI or enacted company-wide ethical guidelines principles on AI use. Effective governance rests on interdisciplinary collaborations where managers work closely with organization experts, AI experts, lawyers, and public relations specialists in establishing risk management frameworks. For instance, the public relations department may assist managers and AI teams up front in assessing reputational risks and formulating an action plan should the use of AI lead to adverse outcomes. Such assessment could benefit managers in striking the delicate balance between how much they need to understand about the workings of AI, while remaining accountable.

Incomplete frameworks for human-in-the-loop analytics

Due to the inherent challenges underlying delegation, not all AI applications in management can be automated end-to-end. Rather, managers often seek to augment their role with AI, where the AI essentially provides novel information for human decision-making ("human-in-the-loop") [12]. In human-in-the-loop frameworks, AI outputs can be cross-checked by domain experts, or human inputs can be used to overwrite and iteratively improve AI predictions based on their tacit domain knowledge. For instance, in order to mitigate biases within AI, one may complement AI outputs with human domain expertise [13]. Augmenting, rather than fully delegating management decisions to AI, is likely to lower costs for implementation of AI tools and support the risk mitigation discussed above, while contributing to decision-making efficiency and effectiveness in the organization. A case in point is the combination of AI predictions and complementary human judgment for generating accurate forecasts in stock analysis [14].

Despite the benefits of a human-in-the-loop approach, however, companies often find it difficult to combine managerial roles and AI applications across collaborative tasks. For example, when and how should managers engage? Should they sign off on all decisions, or only high-stakes decisions where the AI produces outcomes under large uncertainty? How might domain experts with tacit knowledge augment AI predictions? How is a potential loss in power, authority and responsibility perceived by managers when involving AI? Such conundrums result from behavioral operations (e.g., algorithm aversion [15], fairness [3,29]) when AI algorithms interact with a human supervisor. Despite extensive efforts [16,17,18], more research is needed to develop a principled approach to human-in-the-loop frameworks.

Current efforts to design human-in-the-loop frameworks acknowledge that AI decision-making is not always easy to interpret. Hence, such work systematizes how managers can interact with AI systems through automation and augmentation [2,3,5,6]. However, technical challenges remain in such decision augmentation. Here, we argue that interpretability is a helpful lever that allows managers to generate knowledge on how the AI arrives at prediction. Through tools for interpretability, managers can probe the AI and ask “what-if” questions that inform subsequent decision-making and provide opportunities for human learning. For example, in quality management, such tools for interpretability can learn a ’digital twin’ of manufacturing sites (i.e., a real-time virtual representation of the corresponding physical decision environment) and facilitate identification of sources of low quality [16]. Notwithstanding, AI algorithms do not always lend themselves to human interpretation, which is especially challenging given the abundance of black-box models (e.g., deep learning) in companies.

To develop effective human-in-the-loop frameworks, we believe there are a number of promising paths forward. First, human-in-the-loop frameworks must be created from a human—rather than machine-centered perspective. Future research should combine behavioral studies with AI research to create frameworks that reduce AI aversion and instead promote trust, fairness, and acceptance. Second, research on AI systems should give ample attention to the development of “soft” skills [19], such as problem-solving, critical thinking, and adaptability. Third, we foresee advances in systems that continuously learn from a broad range of human input, and that may act “responsibly” by automatically withholding decisions and deferring them back to managers, or that actively query managers to deliver data where the AI is not sufficiently accurate.

Beyond the human-in-the-loop approach, more studies are needed on prescriptive algorithms as an end-to-end approach for business decision-making (e.g., [20]). These algorithms integrate a prediction model into another optimization model so that predictions (e.g., what is the likeliest course of events) are mapped onto decisions (e.g., what action should managers take to achieve a desired output). They learn treatment effects from observational data (and thus without the need for exploration and even without data on counterfactuals), and thereby recommend decisions that are optimal in generating long-term value for a company (e.g., via off-policy learning, inverse reinforcement learning). However, we also caution that the cost of developing such tailored AI solutions may exceed the potential gains to the company.

Organizational inertia

Organizational forces often hinder the adoption, diffusion, and subsequent potential business value that emerging technologies may deliver to a company [21]. Prior studies have noted that the process of adopting and diffusing technology within organizations, as well as actions taken in response to generated insights, necessitate fundamental organizational transformation. However, organizations resist or dampen such changes resulting in overall failure of the newly adopted technology, often labeled *organizational inertia* [21].

AI is likely to introduce fundamental changes within companies, including how decisions are formed and how authority is delegated [4]. Such changes may entail large-scale transformations of processes and structures. For instance, a retail chain that adopted an AI to monitor the performance of in-store sales staff underwent fundamental changes in the structures of organizational decision-making. Managers needed to accommodate AI in their decisions on personal appraisal and promotions, and employees had to get accustomed to AI-based supervision [22].

Organizational inertia may be particularly dominant for AI technologies (as opposed to traditional IT) due to the complexity of AI (e.g., it requires specialized data science skills), the perceived loss of control, or because the economic benefits of AI tools for the company are difficult to quantify. AI decision outcomes also directly interact with managerial roles and responsibilities, which may require rapid changes in organizing. Such challenges and imperatives for system-wide changes for AI adoption have been well documented in recent empirical findings [23,24,25,26].

Furthermore, delegation of decision authority to AI differs from that to humans [1]. Such changes in the authority structure within organizations introduce novel and important dynamics that need to be explored. This may, for instance, result in difficulty in accepting AI recommended choices despite AI's predictive power. A case in point is the accuracy-interpretability trade-off frequently observed in modern AI systems [30], which corroborates algorithm aversion among managers. In particular, deep learning models are considered highly accurate in their predictions while they pose major challenges to interpretation and, hence, may fail to receive support among organizational stakeholders.

Overcoming organizational inertia requires suitable frameworks that demonstrate the benefits and costs in adopting AI, as well as substantive interdisciplinary research along three directions. Future research should investigate the design of a transformation workforce that oversees the adoption of AI and acts to overcome organizational resistance. Such teams may, for instance, contain a mix of inside managers, AI experts, and operational staff, as well as outside AI experts and consultants. Research should also investigate the design of new incentives that encourage exploration of AI and foster stronger collaborations between various organizational units and AI teams (e.g., research funds or new organizational roles).

Finally, many AI systems are only trained on past data available to the company, which may not generalize well to future situations, especially in case of idiosyncratic events such as financial crises, supply chain disruptions, and pandemics. While we need robust AI systems, only companies with robust strategies will truly shine. Thus, future research should investigate the important intersections between AI evolution and deployment, and competitive advantage.

Conclusion

Bringing AI to management hinges on successfully delegating management decisions to AI. We have discussed three organizational and technical hurdles that need to be overcome through novel interdisciplinary research, as well as timely, prudent, and enthusiastic managerial action.

Author information

Authors are listed in alphabetical order.

Correspondence.
Georg von Krogh (gvkrogh@ethz.ch). ORCID: https://orcid.org/0000-0002-1203-3569

Competing interests.
The authors declare no competing interests.

References

[1] Athey, S. C., Bryan, K. A., & Gans, J. S. (2020). The allocation of decision authority to human and artificial intelligence. In *AEA Papers and Proceedings* **110**, 80−84.

[2] Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review* **46**, 192−210.

[3] Teodorescu, M. H., Morse, L., Awwad, Y., & Kane, G. C. (2021). Failures of fairness in automation require a deeper understanding of human-AI augmentation. *MIS Quarterly* **45**, 1483−1500.

[4] Von Krogh, G. (2018). Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. *Academy of Management Discoveries* **4**, 404−409.

[5] Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics* **160**, 835−850.

[6] Morse, L., Teodorescu, M. H. M., Awwad, Y., & Kane, G. C. (2021). Do the ends justify the means? Variation in the distributive and procedural fairness of machine learning algorithms. *Journal of Business Ethics*, forthcoming.

[7] Athey, S., & Roberts, J. (2001). Organizational design: Decision rights and incentive contracts. American Economic Review **91**, 200−205.

[8] Jensen, M. C., & Heckling, W. H. (1995). Specific and general knowledge, and organizational structure. *Journal of Applied Corporate Finance* **8**, 4−18.

[9] Reuters. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Edited by Jeffrey Dastin. URL: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G, last accessed 01-11-2021.

[10] Tenbrunsel, A. E., Diekmann, K. A., Wade-Benzoni, K. A., & Bazerman, M. H. (2010). The ethical mirage: A temporal explanation as to why we are not as ethical as we think we are. *Research in Organizational Behavior*, **30**, 153−173.

[11] Ajunwa, I. (2020). The “black box” at work. *Big Data & Society* forthcoming.

[12] Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human–AI symbiosis in organizational decision making. *Business Horizons* **61**, 577–586.

[13] Choudhury, P., Starr, E., & Agarwal, R. (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal* **41**, 1381–1411.

[14] Cao, S., Jiang, W., Wang, J. L., & Yang, B. (2021). From man vs. machine to man+machine: The art and AI of stock analyses. *National Bureau of Economic Research* w28800.

[15] Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* **144**, 114–126.

[16] Senoner, J., Netland, T., & Feuerriegel, S. (2022). Using explainable artificial intelligence to improve process quality: Evidence from semiconductor manufacturing. *Management Science*, forthcoming.

[17] Madras, D., Pitassi, T., & Zemel, R. S. (2018). Predict responsibly: Improving fairness and accuracy by learning to defer. In *Conference on Neural Information Processing Systems (NeurIPS)*.

[18] Sun, J., Zhang, D. J., Hu, H., & Van Mieghem, J. A. (2021). Predicting human discretion to adjust algorithmic prescription: A large-scale field experiment in warehouse operations. *Management Science*, forthcoming.

[19] Deming, D. J. (2021). The growing importance of decision-making on the job. *National Bureau of Economic Research* w28733.

[20] Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science* **66**, 1025−1044.

[21] Kelly, D., & Amburgey, T. L. (1991). Organizational inertia and momentum: A dynamic model of strategic change. *Academy of Management Journal* **34**, 591−612.

[22] Tong, S., Jia, N., Luo, X., & Fang, Z. (2021). The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal*, forthcoming.

[23] Agrawal, A. K., Gans, J. S., & Goldfarb, A. (2021). AI adoption and system-wide change. *National Bureau of Economic Research* w28811.

[24] Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation, and work. In *The Economics of Artificial Intelligence: An Agenda* (pp. 197−236). University of Chicago Press.

[25] Brynjolfsson, E., & McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review* **106**, 133−39.

[26] Goldfarb, A., Taska, B., & Teodoridis, F. (2022). Could machine learning be a general purpose technology? A comparison of emerging technologies using data from online job postings. *National Bureau of Economic Research* w29767.

[27] Nilsson, N. J. 1971. Problem-solving methods in artificial intelligence. New York: McGraw-Hill.

[28] Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *Journal of Strategic Information Systems* **29**, 101614.

[29] De-Arteaga, M., Feuerriegel, S, & Saar-Tsechansky, M. (2022). Algorithmic fairness in business analytics: Directions for research and practice.

[30] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* **1**, 206−215.

[31] Mäntymäki, M., Minkkinen, M., Birkstedt, and Vilijanen, M. (2022). Defining organizational AI governance. *AI Ethics*.

[32] Gottfried, I.S. (1989). When disaster strike. *Journal of Information Systems Management* **6**, 86−89.

[33] Bandyopadhyay, K., Mykytyn, P. P., & Mykytyn, K. (1999). A framework for integrated risk management in information technology. *Management Decision* **37**, 437−445.