Winter 12-2019

Man-Machine Collaboration in Organizational Decision-Making: An Experimental Study Using Loan Application Evaluations

Anh Luong

Nanda Kumar

Karl Lang

Man-Machine Collaboration in Organizational Decision-Making: An Experimental Study Using Loan Application Evaluations

Completed Research Paper

Anh Luong
Baruch College, CUNY
anh.luong@baruch.cuny.edu

Nanda Kumar
Baruch College, CUNY
nanda.kumar@baruch.cuny.edu

Karl R. Lang
Baruch College, CUNY
karl.lang@baruch.cuny.edu

Abstract

With artificial intelligence and machine learning (AI/ML) being increasingly employed in complex decision making, human experts are being required to collaborate with these technologies in order to improve their organizational outcomes. The use of AI/ML in organizational decision-making involves three components: the people, the technology, and the organizational practices. Using the loan application reviewing task context, we examine factors that can produce effective collaboration between AI/ML and human decision makers, leading to enhanced firm performance. Our experimental study shows that firm success when employing AI/ML for decision-making depends on both high-quality AI technologies and well-aligned organizational practices. Most importantly, the two factors are inseparable, and their co-presence is crucial if firms want to close the theoretical gap between their potential and their most fully realized capacity.

Keywords

Algorithmic decision-making, organizational decision-making, algorithm aversion, algorithm usage, experimental economics, organizational practices, incentive structure, incentive alignment, man-machine collaboration, intelligence augmentation

Introduction

In recent years, firms across industries have been rapidly obtaining artificial intelligence and machine learning technologies (AI/ML) to integrate into their operations (Webb 2018). AI/ML is being employed not only for monotonous tasks, but also for non-routine and complex operations previously reserved for human experts, such as diagnosing physical and psychological illnesses, detecting credit card frauds, conducting research and writing articles, even devising investment and corporate strategies. This has led to heightened concerns about full automation soon replacing humans in both routine and non-routine roles. Currently however, for many of such complex business tasks, AI is actually being deployed to aid (rather than replace) human experts, who ultimately still have the final say in organizational decisions. We believe this to be an interesting and important problem space, where humans are required to interact with and capitalize on AI insights for the betterment of their organizations' decision-making processes and outcomes. Early academic research on AI use reported extensively on the *algorithm aversion* phenomenon, where human experts showed serious reservations toward using AI (e.g. Rader & Gray 2015; Dietvorst et al. 2015; Rosenblat & Stark, 2016; Lee et al., 2015). Recent studies found that under certain circumstances, people actually do use AI/ML and sometimes even prefer AI/ML to human judgement (Dietvorst et al. 2016; Alexander et al. 2018; Logg et al. 2019; Yeomans et al. 2019; Packin 2019). Furthermore, recent industry studies focused on firms' adoption of AI have reported that the values from the majority of firms' AI implementations are yet to be capitalized, due to lack of foundational practices that can facilitate the full integration of AI/ML in business operations throughout the firms' various units (e.g. Webb 2018). Thus, additional understanding is needed on (1) how human decision makers can learn to work with and develop lasting effective professional relationship with AI; and (2) how certain organizational practices can best facilitate such relationship, for the enhancement of firm performance. Our research question is: *How do technology and organizational practices affect firms' performance when they deploy AI/ML for complex decision-making?*

Theoretical Foundation and Research Model

Our hypotheses are developed based on the Absorptive Capacity (ACAP) theoretical framework by Cohen and Levinthal (1990), which states that organizations that are more able to recognize new valuable external knowledge, attain it, and integrate it in their operations, should see increased performance and generate greater competitive advantage (Lichtenthaler, 2009; Lane et al. 2006). This capability of organizations is called their absorptive capacity. We examine our research question in the specific context of a well-known decision problem in banking—evaluating consumer loan applications. Thus, the two principal factors that determine organizational ACAP (organizational knowledge structures, organizational learning processes) are represented in our study by the **AI/ML system** that predicts loan risks (new IT capability) and the **incentive alignment mechanism** (compensation policy that fits the organizational lending strategy), drawing on the *ACAP*, *compensation*, and *incentive alignment* literatures. Firstly, the AI/ML, with its associated database of historic loan performance records, enhances the organizational knowledge base and the firm. Secondly, the organization's risk-reward structure, through its incentive alignment mechanism, contributes to better decision-making. These combined enable the organization to capture more of its potential capacity, meaning that adopting an AI/ML technology which is embedded within a learning process and adopting a policy that aligns the right incentives with the firm strategy can move a firm's performance closer to its theoretical capacity. It also means, however, only adopting high quality AI/ML may not generate much organizational benefit if it is not coupled with an incentive structure that connects decision logic with firm strategy. On the other hand, enforcing a suitable incentive structure is not enough if the technology assisting the decision-making process is sub-par. We posit that a high-quality AI/ML system can mitigate the negative impacts that a misaligned incentive structure exerts on firm performance. Thus, our study examines the impacts of AI/ML quality (in terms of prediction accuracy) and organizational incentive structure (in terms of alignment with a conservative lending strategy) and most importantly, their interdependent effect on firms' absorptive capacity. Our hypotheses are: ***H1.*** *Prediction accuracy of the AI system in the loan review process is positively associated with AI/ML usage level (in terms of (a) number of Revised Decisions, (b) degree of Change of Mind)* | ***H2***. *Prediction accuracy of the AI system in the loan review process is positively associated with firm performance (in terms of (a) lower Type I error rates, (b) lower Type II error rates)* | ***H3.*** *The incentive structure in an organization impacts its performance in terms of achieving its objectives. More specifically, for organizations with a conservative, risk-avoiding lending strategy, a well-aligned (misaligned) risk-reward incentive structure is (a) positively (negatively)* *associated with a Type I error rate and is (b) negatively (positively) associated with a Type II error rate* | ***H4.*** *The effects of incentive alignment on (a) Type I error rate and (b) Type II error rate are weaker (stronger) in the presence of high (low) AI/ML prediction accuracy.*

We examine AI/ML usage level and firm performance as these constructs represent the components of absorptive capacity: firms' ability to (1) recognize the value of and (2) assimilate (internalize) the new valuable knowledge to apply it to commercial outcomes (Cohen & Levinthal 1990). The first two components are proxied by AI/ML usage level, and the last component is proxied by firm decision-making performance in our study, following studies in the ACAP literature. Revised Decisions are the decisions that are modified per AI/ML predictions, and thus represent willingness to listen to the AI/ML. Change of Mind (Marks et al., 2018) is a more granular measure of AI/ML usage, as it considers the changes in both decisions and confidence levels and thus more closely reflects the belief in the joint work between the human experts and the AI/ML. Type I and Type II errors are the two most common measures used by lending firms to assess their loan review decision outcomes, as these two metrics have direct and significant effects on their bottom lines (Goh 2018).

Methodology & Results

We follow the major guidelines of experimental economics methodology and design a controlled lab experiment where financially incentivized subjects represent rational agents that are traditional retail banks (conservative lending strategy). They are asked to review a set of 100 loan applications, spread over 10 decision rounds. All subjects earn a completion reward ($5) plus a performance-based reward (total ranging $5-$20). The experimental platform is run using oTree (Chen et al. 2016). We conducted experiments from March 27 to April 18, 2019. The subjects were 152 undergraduate students in the Zicklin School of Business, Baruch College. Our experiment has a 3x2 factorial design, the 3 factors being AI Prediction Accuracy (Low vs. High), Incentive Alignment (Aligned vs. Misaligned) and Lending Strategy (Conservative vs. Aggressive). We fix Lending Strategy at Conservative for this study. Our four treatment conditions are High Accuracy-Aligned (n=26), High Accuracy-Misaligned (n=37), Low Accuracy-Aligned (n=23), and Low Accuracy-Misaligned (n=21). We also ran a Baseline condition (n=45) that does not involve AI predictions in the loan review process. AI Prediction Accuracy is manipulated through the accuracy rates of the AI's predictions—60% for Low and 80% for High Accuracy. Incentive Alignment is manipulated through the pay-off schedule: All subjects earn 20 cents for each correct approve/reject, and nothing for each false reject. For each false approve, the penalty is $1 in the Aligned condition and $0 in the Misaligned condition. In each of the 10 rounds, subjects review 10 loan applications, in 3 stages: (1) make initial decisions (approve/reject) and rate confidence (0-100) in the decisions; (2) view the AI predictions (whose accuracy rates are informed to subjects) and finalize their decisions and confidence levels (which can be revised/updated from the initial stage, or kept unchanged); (3) view the results of how each loan actually performs, and how much they earn. Subjects also answer two questionnaires—one after and one before the experiment—which assess subjects' demographics information and various personality measures.

We employed multiple linear regression analysis to test our hypotheses. To control for the effect of repeated measurement, especially learning effects throughout the rounds, we included 9 dummy variables for the 10 decision rounds in our regression models. Additionally, we controlled for *subjects' quantitative and financial competencies, financial risk tolerance, ethnicity, gender, along with the number of poor performing loans (per round) and the decision performance of the AI/ML in each round*. Our results support our argument (all hypotheses supported) that firm success when employing AI for decision making depends on both the quality of the IT resource (AI systems with more reliable predictions) and the alignment of the organizational practices (incentive structure being aligned with firm's strategy). Most importantly, our results show that in order for firms to reap benefits from algorithmic decision making, it is crucial that both proper IT resources and organizational processes be concurrently implemented. Only when the two exist together can firms narrow the theoretical gap between their highest possible potential capacity and their realized capacity.

Besides the main analysis for our hypothesis testing, we also examined whether embedding algorithmic decision-making in organizational business practices actually leads to better firm outcomes and thus to absorbing additional firm capacity at all. We ran separately two ANOVA tests to compare between: (1) the Baseline and the Low Accuracy sessions, and (2) the Baseline and the High Accuracy sessions. The results show that while firm outcomes did not significantly differ between the Baseline and the Low Accuracy sessions, they indeed did differ between the Baseline and the High Accuracy conditions. Additionally, even though the p-value is slightly higher than 0.1, the Low Accuracy condition even produced a slightly higher type II error rate than the Baseline. These findings further emphasize our position that simply investing in new technology does not necessarily increase firm absorptive capacity, as technologies lacking in reliability cannot help human decision makers translate organizational knowledge into effective decisions and sustainable competitive outcomes for the firm, and thus cannot result in increased organizational absorptive capacity. Moreover, since our analysis controls for financial and quantitative competencies, we cannot disregard the importance of expertise in human decision makers. They are still critical for organizational decision making in that with expertise and skills they can differentiate between high quality and low quality algorithmic advice, and in the case of having to use a low quality algorithm, they should be more likely than non-expert decision makers to be able to override the bad algorithmic advice.

Our study contributes to the current algorithm usage literature by focusing on the characteristics of both the IT resources (AI decision support) and the organizational practices (the incentive alignment mechanism) and examining their interdependent effects on firm performance. In the current state of rapidly growing data, mistakes that start from one decision or one decision maker can easily lead to a massive number of additional mistakes that can wreak havoc on firm performance. Without proper practices securely implemented to align decision logic with firm strategy, neither the AI nor the human decision makers can be properly managed and relied on to make the best possible decisions for all the stakeholders involved. Another contribution of our study is adding a dynamic perspective. Unlike most present behavioral studies on AI use, our experiment features repeated round design (as opposed to one-shot), which allows us to perform a longitudinal analysis that shows the more nuanced dynamics of how decision makers interact with AI over time. Our experiments show that when firms deploy a mature, reliable AI system and give human decision makers adequate time to interact with it, the decision makers do actually use it and develop a productive working relationship with it, leading to noticeably improved firm performances.

References

Alexander, V., Blinder, C., & Zak, P. J. (2018). Why trust an algorithm? Performance, cognition, and neurophysiology. Computers in Human Behavior, 89, 279-288.

Chen, D.L., Schonger, M., Wickens, C., 2016. oTree - An open-source platform for laboratory, online and field experiments. Journal of Behavioral and Experimental Finance, Vol 9: 88-97

Cohen, W. M., and Levinthal, D. A. 1990. "Absorptive Capacity: A New Perspective on Learning and Innovation," Administrative Science Quarterly (35:1), March, pp. 128-152.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2016) Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. Management Science.

Goh, G. (2018). Automated Machine Learning is a Game Changer for Wellen Capital. Datarobot, Inc.

Lane, P. J., Koka, B. R., and Pathak, S. 2006. "The Reifícation of Absorptive Capacity: A Critical Review and Rejuvenation of the Construct," Academy of Management Review (31:4), pp. 833-863.

Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015, April). Working with machines: The impact of algorithmic and data-driven management on human workers. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 1603-1612). ACM.

Lichtenthaler, U. 2009. "Absorptive Capacity, Environmental Turbulence, and the Complementarity of Organizational Learning Processes," Academy of Management Journal (52:4), pp.822-846.

Logg, J. M.; Minson, J. A.; Moore D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151 (2019) 90–103.

Marks, Joseph and Copland, Eloise and Loh, Eleanor and Sunstein, Cass R. and Sharot, Tali, Epistemic Spillovers: Learning Others' Political Views Reduces the Ability to Assess and Use Their Expertise in Nonpolitical Domains (April 13, 2018). Harvard Public Law Working Paper No. 18-22.

Packin, Nizan Geslevich, Algorithmic Decision-Making: The Death of Second Opinions? (March 28, 2019).

Rosenblat, A., & Stark, L. (2016). Algorithmic labor and information asymmetries: A case study of Uber's drivers. International Journal of Communication, 10(27).

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. Journal of Behavioral Decision Making.

Webb, N. (2018). AI adoption advances, but foundational barriers remain. McKinsey & Company.