4-1-2022

Teaching the Machine: How People Teach Algorithms to Replace People

Anna Zaitsev
The University of California, Berkeley, anna.zaitsev@berkeley.edu

TEACHING THE MACHINE: HOW PEOPLE TEACH ALGORITHMS TO REPLACE PEOPLE

Anna Zaitsev
University of California, Berkeley
School of Information
anna.zaitsev@berkeley.edu

ABSTRACT

Automation of work via artificial intelligence is becoming a significant issue for societies. This ethnographic study presents a case study that details the process that takes place to teach the algorithms that will eventually replace the need for human effort. The author was employed as a data quality analyst and a part of a process where algorithms were taught to generate better content and eventually replace human content creators. The study proposes a three-stage process of automation where the relationship between the humans and algorithms progresses from symbiotic to cannibalistic: the first phase is the commencement phase, where the human employees and algorithms live in symbiosis, reliant on each other. The symbiosis is followed by the expansion phase, where more work is delegated to the algorithms, and the final phase is the automaton phase, where human employees are no longer needed.

Keywords

Automation, artificial intelligence, ethnography, automatons

INTRODUCTION

Applications of artificial intelligence have swiftly permeated modern workplaces and homes, automating both mundane and complex tasks at an accelerating pace. A Pew Research Center report from 2014 states that experts who study automation and intelligent digital agents believe that these systems will significantly impact every aspect of daily life (Smith and Anderson, 2014). It has been suggested that by 2025, AI applications will perform many of the tasks that currently require human intelligence, including many currently existing blue-collar and, in increasing numbers, white-collar jobs (ibid.).

The benefits and disadvantages of automation are currently under vigorous debate, both in academia and among practitioners (Smith and Anderson 2014). Many researchers are concerned that the expansion in automation and the utilization of big data, intertwined with automation, create changes for which our societies are not yet prepared (Smith and Anderson, 2014; Markus and Topi, 2016, Kane et al., 2021).

However, organizations still must employ thousands of human labourers to monitor their content (Roberts, 2014). This workforce, largely hidden from the public view (Roberts, 2014), is knowingly or unknowingly participating in a less discussed and understood aspect of data analysis, the process of automation and machine learning. The human factor is significant but currently underresearched (Günther et al., 2017). Considering this gap in the literature, this study is asking:

How do organizations devise the processes that improve the algorithms to reduce the reliance on human employees?

Our response to this question is based on an ethnographic investigation into the world of data analysis work. This paper presents an observational account of how humans and algorithms interact during the machine learning process in a supervised machine-learning scenario. Where the extant research has focused mainly on applying the data and the tools for data analysis as the drivers for change (Günther et al., 2017), this paper has centred the study around the analysis process, which drives the data application and development of the tools.

THE CHALLENGES OF ARTIFICIAL INTELLIGENCE

There is a debate ongoing in Information Systems (IS), identified by Günther et al. (2017), on human intelligence and algorithmic intelligence. This debate highlights both the strengths and weaknesses in humans and algorithms alike, and augmented models are suggested, but, as Günther et al. note, it is unclear what insights are to be gained from hybrid models.

For example, Kane et al. (2021) are concerned that augmenting human decision-making with machine learning can lead to oppressive models and should be designed with emancipatory functions. Fügener et al. (2021) state that humans who are interacting with AI's discard their human individuality and lose their human knowledge. Newell and Marabelli (2015) highlight the some of the other social consequences which include dependence on algorithms, which can lead to a situation where humans forget how certain things are done without the use of algorithms. The trade-off is between comfort and independence from the devices. The aspect of tacit knowledge being replaced by algorithms is where more studies are required (Galliers et al., 2017). The interdependency can have adverse effects on decision making. Parasuraman and Manzey (2010) describe a study where the use of automation led to "automation complacency", overreliance on the automated systems, which led to errors and "automation bias", decisionmaking which favours the automatic advice over rigorous analysis. The bias in decision making is contextual, and the design of the systems which analyze data and support decision making is tied to the social and organizational environment in which the designers operate (Levenson, 2011). Markus (2017) argues that the context of the automated tasks, the level of risks to which the automated decision making is subjecting humans, should also be carefully considered. For example, Markus juxtaposes warehouse robots and self-driving cars. The latter can create much more significant damage if erroneously designed.

However, not all research regarding AI's in information systems is all doom and gloom. Some studies take a different perspective and see the systems as valuable tools for businesses in an eclectic collection of domains ranging from prediction of movie ticket sales (Lash and Zhao, 2016) to automated interviewing (Nunamaker et al., 2011). Researchers have even suggested that other researchers could harness the power of artificial intelligence when studying big data and argue that machine learning would be a powerful tool for the researchers themselves (Müller et al., 2016).

METHOD

Our data collection method was based on a three-month-long ethnographic study during which the author worked as a part of the data analysis team. Ethnographies, originating from the field of anthropology, are first-hand observations of the researcher who observes, records and engages in the daily life in the culture or organization where they have immersed themselves and which they were unfamiliar with prior to the engagement (Schultze, 2000).

The Ethnographic Study Background

The goal of the study was to understand what was taking place in the organization and other similar organizations by taking part in the work and discussing it with the other employees, i.e., the author had to 'go native' and take part in all the aspects of the project, acting similarly to everyone else working on the job (Myers, 1999).

The author was fully embedded in the team during the study, working each day for 8 hours, located almost exclusively in the project premises. The ethnographic study began in late January 2018 and was concluded in May 2018. The decision to conclude the study was made based on the project plateauing. There were no new, significant developments predicted for the short term.

Data Gathering

From the beginning, the author kept a research diary of the events and impressions of the work. The author kept a visual diary with sketches of processes and organization, sketching the office layout and other visual cues that would help refresh the memory, later replicated in clearer format with online drawing tools. The author also occasionally dictated personal notes on a mobile phone, creating an audio log for some of the musings of the day and the written notes.

Table 1. summarises the data and the media used for the data collection.

Table 1. Ethnographic data	
Medium	**Data**
Daily diary notes	Daily diary of events, work process descriptions and thoughts on the process
Recordings	Supplement to diary, dictated notes
Notebook	Process doodles, office layout diagrams, organisation charts

Table 1. Ethnographic data

Data Analysis

The data analysis and theory building were conducted concurrently with the ethnographic study. A first version of the theory emerged after the first month of the study and was captured in the daily notes and the notebook. However, working on the project changed the theoretical model from data-centric to more process and human-centric.

The theoretical view became more focused and refined by comparing the new data with the older data. I followed the data analysis techniques described by Langley (1999), creating narratives and visual maps of the processes that occurred during the study period. Visual maps were drawn in an attempt to capture the high-level processes ongoing in the organization. These visualizations were drawn as part of the thought process, helping visually connect the steps of the process. A more polished version of the most relevant of these visual maps is presented in the following sections as Figure 2.

THE STUDY: TRAINING THE MACHINE

The author worked as a data quality assurance analyst for a web-based advertisement company. The data analysed by the author was localized advertisement material, short snippets aiming to entice a user to click on the advertisement. The author's first two weeks were spent in different training sessions, learning the guidelines and the data assurance process. In the beginning, the author had false expectations of the capabilities of the algorithms before being introduced to the workflows, but the first two weeks clarified these misconceptions. The algorithms were far less advanced than what the author had initially thought. However, the author soon recognized that the work currently being done was happening at a turning point; the machines would soon catch up with the author's original expectations.

After the training period, the first sets of actual data were given to the team members. A typical workday unfolded as follows: in the mornings, the analysts opened their computers, checked in with their teams in an online log tool and began working on the tasks assigned. The goal was to assure that the advertisement snippets were grammatically correct and according to the guidelines. The data was accessible on a platform, which was internal to the organization, and the analysis was conducted using the tools provided by the organization. There was also a summary that kept count of the results of the data analysis, automatically populated based on the data analysis done by the employees. The analysts would work through the data until they reached their daily targets and fill in three different reporting logs. The next day, the same routine would be repeated. The analysts were also asked to write up short memos or descriptions on the most common issues they saw in the data from time to time. According to the team leads, these summaries would help clarify some of the issues for the stakeholders.

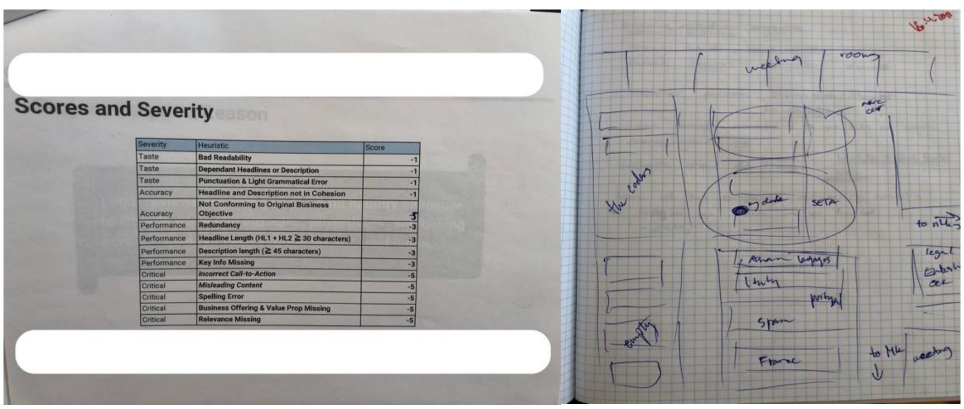

Figure 1 An (anonymized) example of training material and a seating chart doodle

Approximately every two weeks, a new batch of data was presented, and the analyst would switch over to this new dataset, even if the processing of the older dataset were not completed. Occasionally the new data heralded changes in the guidelines as well. The guidelines were always made more specific, adding new quality rules. These rules were presented in brief meetings, and they were regularly revised with the analysts in the so-called synchronization meetings. The guidelines had two facets: the data had to be accurate, and it also had to conform to the legal provisions, the requirements that mandated what kind of data was acceptable.

The primary role of the managers was to ensure that the stakeholder requirements were communicated to the analysts. However, to the author, the data providers were present as a black box. Data were coming in from an unknown source via unknown means. These business stakeholders, setting the quality rules for the data, would only communicate with the managers, and they were, according to the managers, also the consumers of the short data summaries, the automatically populated graphs that captured the data analysis of each data batch.

Furthermore, developers of the algorithm were unfamiliar to the analysts. These stakeholders communicated with the team only via new versions of the data. It was, however, clear from the changes in data that someone was taking the feedback on board and was tweaking the algorithm.

The three months that the author spent on the shop floor saw significant changes in the data. The first batches of data were very crude, and analysis revealed an abundance of errors. A translated excerpt from the notes illustrates the issues with the first batch:

13.3. After work notes: the content changed from curated ones to the algorithmic ones. The supplier has provided some guidelines that will auto-generate the content. There are two versions created by the algo. Some are identical, some are different. I've no idea why the duplication. Almost of the lines just repeats the same content in the header. Quality poor.

The second batch, developed a few weeks later, was already cleaner but still riddled with errors. The third batch of data was significantly better than the first two. There was a significant reduction in the number of errors, and this was considered to be business-critical. The fourth batch was similar to the third batch, with another slight reduction in the error rates. A diary entry describes the new data batch:

23.3. Lunch break notes: New set, potentially improved algo. No visible changes in the content text, thought. Quality is similar, lots of repetition. Too many capital letters. I think the original content was bad, but these are worse. Less inventive.

Every batch had introduced new rules for the analysis, and the fifth batch was no different. More complex analysis was needed to ensure that the data were conforming to the quality standards. This added complexity to the human efforts was translated into fewer errors for the algorithm. This was evident from work conducted by another similar team working on the data for a more extended period. That team was receiving a much more advanced version of the algorithmically generated data, with almost no errors but had also worked through a more significant number of iterations.

Two major changes occurred during the three months of the observation: the team tripled in size, and the data got significantly better. A brief mention in the notes describes the quality control process:

6.4. Lunch break notes: I checked the stats for errors in Arabic. Huge improvement! The algo really nailing it this round.

The project's undiscussed but implied end goal was to directly analyze all new data through the algorithms without the need for further human analysis and quality control. However, even though the author was attempting to estimate the project's duration, it gradually became apparent that there was no certainty on that matter. It was unrealistic for the author to observe the project until the conclusion, and the study was concluded after three months of extensive immersion into teaching the machines.

FROM SYMBIOTIC TO CANNIBALISTIC – THE HUMAN-ALGORITHM RELATIONSHIP

The observations made during this ethnographic study describe a situation where the relationship between humans and the algorithm begins as symbiotic and mutually beneficial but is destined to develop into a cannibalistic one, where one part of the relationship, the algorithm, devours the other.

Based on our understanding of the existing data validation process and the relevant literature discussing and predicting the future of algorithmic data analysis, we propose a theory that combines the insights from the study and the views of the experts. Many researchers, including machine learning and AI experts (see Smith and Anderson, 2014) and IS authors, such as Markus et al. (2008) and Newell and Marabelli (2015), have envisioned a future where human input is no longer needed in many areas where it persists. Building on this, this study proposes a theory of algorithm-human relationship, which describes the symbiotic-turned-cannibalistic relationship between the human teachers of the machine learning algorithm and the algorithm itself.

By combining the ethnographic experiences and first-hand accounts of the process and other cases discussed in the literature (e.g., Newell and Marabelli, 2015), the author has arrived at a three-phase process. First, the organization arrives at the stage at which the amount of data they need to analyze becomes overwhelming for the organization's staff. During this initial phase, the Commencement Phase, the organization focuses on the data sets that provide the most value, e.g., is responsible for the largest market share or are otherwise significant for the business. The organization decides to create an algorithm to analyze the data. The first version of the algorithm is crude and needs more data to improve. The organization has a small number of people analyzing the data and training the algorithm by providing labelled feedback to the algorithm developers.

During the second phase, the organization needs to expand their efforts beyond the initial dataset and focus on more complex and niche analysis. The Expansion Phase begins when the organization grows in terms of quantity of data and concurrently requires more people to handle the growing amount of data. The Expansion Phase is likely to correspond with the growing amount of data and internal and external pressures that demand better data analysis and management. At this phase the human decision-making is being augmented with the AI's (Fügener et al., 2021) but the AI's are not yet making independent decisions without human supervision. During the Expansion Phase, the organizations might recruit additional employees to analyze the data and improve the algorithm. The algorithm is constantly tuned according to the feedback, and the human counterparts subject these recalibrated versions of the algorithm to more analysis.

In the third phase, the Automaton Phase, the algorithm is deemed to function well enough without large training datasets, and it becomes a self-operation machine (from greek 'automatos'). The next steps for the algorithm might be unsupervised learning or learning from other algorithms. The results are good enough for the organization; it no longer needs to retain the cohort of analysts. The algorithm has made human analysis redundant. However, there is the question of the sustainability of this phase. What if there are changes to the data or the quality guidelines or concerns for fairness and bias ()? Are the algorithm developers sufficient resourced when it comes to maintaining the algorithm's accuracy, or does the process begin anew with the initiation of another Commencement Phase? This has yet to be seen.

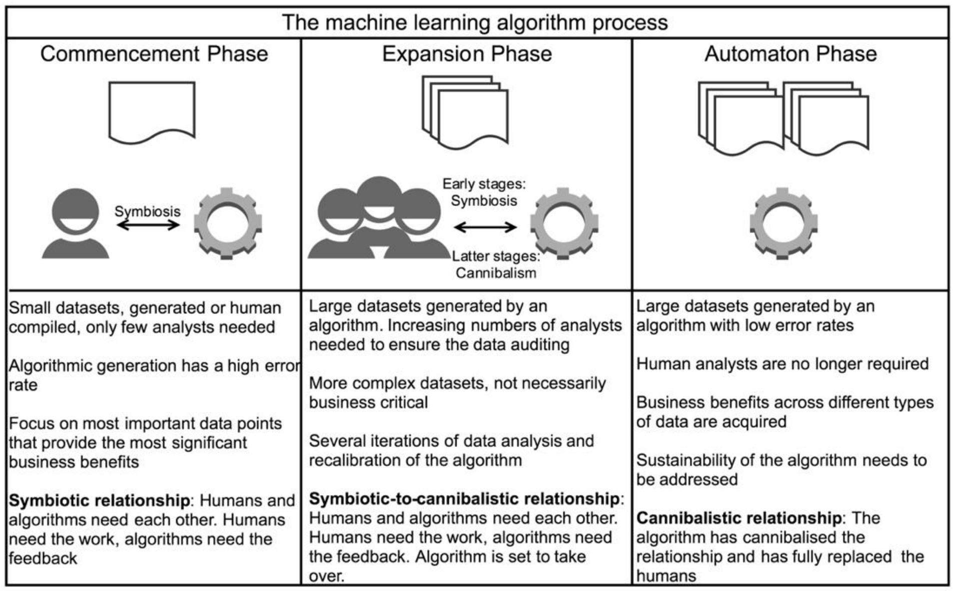

Figure 2 From Symbiotic to Cannibalistic Relationship

Figure 2 illustrates the process from Commencement to Automaton Phase. Two of these three phases of the theory of algorithm-human relationship in the modern workplace are observable in both practice and literature, whereas the third phase, the Automaton Phase, might require a few more years to properly manifest. During this study, the author experienced the first two phases of this model. Experts predict that the near future will usher in the Automaton Phase and permanently change modern societies (Smith and Anderson, 2014).

CONCLUSION

Artificial intelligence is a significant part of daily work in many large organizations, especially those dealing with big data analysis (Günther et al., 2017). However, there is a need for more discussion in information systems about algorithms and the short- and long-term effects they have, as stated by Newell, Markus and others. Reversing the algorithms' effect on society does not seem to be an option discussed; the only way to address the social issues is to look to the future, not the past.

This study has presented an ethnographic account of a process where human labour is applied to teach a machine to become independent of most humans. A process that consists of three parts, Initiation Phase, Expansion Phase and Machine Phase, illustrates how this independence is gained and maintained. It is unclear when the Machine Phase will be commonplace in large part of data analysis; the very complex applications of machine learning are yet in their nascent phase (LeCun et al., 2015). However, we should remain cognizant of the high probability of this development, which is likely to take place gradually in the coming years.

There are multiple benefits that these techniques provide. LeCun et al. (2015) describe several applications where complex machine learning algorithms are making significant headway, beneficial for society: analysis of drug molecules, particle accelerator data, brain circuits and effects of mutation in the DNA (pp. 436). When it comes to analyzing pictorial and video data, the machine learning algorithms that parse Internet content can have surprising benefits: they can safeguard the human psyche. Anecdotal evidence from people who have been employed as data reviewers for YouTube and similar video content outlets have described accounts of deep mental scarring from seeing horrific human depravities during their work as content reviewers (e.g. Roberts, 2014). Fully automated screening would reduce the need for human employees in such lines of work.

However, there are threats that the Automaton Phase will bring. Are societies ready for the loss of both blue and white-collar jobs to automation? The topic needs to be further addressed. In addition, the algorithms can become less objective and more dependent on the subjective views of their creators, against the intention of their creators. Bias added to the system during many development process steps can be hard to detect, especially when creating very complex data analysis tools that require joint efforts of both humans and algorithms (Teodorescu et al., 2021). Addressing the inherited bias in the systems is an interesting topic which warrants more attention.

Limitations

Ethnographic studies are limited in the breadth of their scope, limiting the study to a particular context (Myers, 1999). However, the study presented in this paper is not attempting to claim that every organization is organizing their process to develop machine learning in the same way but that this relationship will follow similar patterns to the Symbiotic-to-Cannibalistic human-algorithm relationship, developed based on the data and the literature. This study generalizes from description to theory, as is commonplace in the qualitative studies conducted in the field of IS (Lee and Baskerville, 2003).

REFERENCES

1. Fügener, A., Grahl, J., Gupta, A. and Ketter, W., 2021. "Will Humans-in-the-Loop Become Borgs? Merits and Pitfalls of Working with A,". *Management Information Systems Quarterly (MISQ)-Vol, 45*.
2. Galliers, R., Newell, S., Shanks, G., and Topi, H. 2017. "Datification and Its Human, Organizational and Societal Effects: The Strategic Opportunities and Challenges of Algorithmic Decision-Making." Elsevier.
3. Günther, W.A., Mehrizi, M.H.R., Huysman, M., and Feldberg, F. 2017. "Debating Big Data: A Literature Review on Realizing Value from Big Data," *The Journal of Strategic Information Systems*).
4. Kane, G.C., Young, A.G., Majchrzak, A. and Ransbotham, S., 2021. "Avoiding an oppressive future of machine learning: A design theory for emancipatory assistants," *MIS Quarterly*, (45:1), pp.371-396.
5. Langley, A. 1999. "Strategies for Theorizing from Process Data," *Academy of Management Review* (24:4), pp. 691–710.
6. Lash, M.T., and Zhao, K. 2016. "Early Predictions of Movie Success: The Who, What, and When of Profitability," *Journal of Management Information Systems* (33:3), pp. 874–903.
7. LeCun, Y., Bengio, Y., and Hinton, G. 2015. "Deep Learning," *Nature* (521:7553), p. 436.
8. Lee, A.S., and Baskerville, R.L. 2003. "Generalizing Generalizability in Information Systems Research," *Information Systems Research* (14:3), pp. 221–243.
9. Leveson, N. 2011. Engineering a Safer World: Systems Thinking Applied to Safety. MIT press.
10. Markus, M.L. 2017. "Datification, Organizational Strategy, and Is Research: What's the Score?," *The Journal of Strategic Information Systems* (26:3), pp. 233–241.
11. Markus, M.L., and Topi, H. 2015. "Big Data, Big Decisions for Science, Society, and Business," *Report on a Research Agenda Setting Workshop. Bentley University*.
12. Müller, O., Junglas, I., vom Brocke, J., and Debortoli, S. 2016. "Utilizing Big Data Analytics for Information Systems Research: Challenges, Promises and Guidelines," *European Journal of Information Systems* (25:4), pp. 289–302.
13. Myers, M. 1999. "Investigating Information Systems with Ethnographic Research," *Communications of the AIS* (2:4)
14. Newell, S., and Marabelli, M. 2015. "Strategic Opportunities (and Challenges) of Algorithmic Decision-Making: A Call for Action on the Long-Term Societal Effects of 'Datification'," *The Journal of Strategic Information Systems* (24:1), pp. 3–14.
15. Nunamaker, J.F., Derrick, D.C., Elkins, A.C., Burgoon, J.K., and Patton, M.W. 2011. "Embodied Conversational Agent-Based Kiosk for Automated Interviewing," *Journal of Management Information Systems* (28:1), pp. 17–48.
16. Parasuraman, R., and Manzey, D.H. 2010. "Complacency and Bias in Human Use of Automation: An Attentional Integration," *Human factors* (52:3), pp. 381-410.
17. Roberts, S.T. 2014. Behind the Screen: The Hidden Digital Labor of Commercial Content Moderation. University of Illinois at Urbana-Champaign.
18. Schultze, U. 2000. "A Confessional Account of an Ethnography about Knowledge Work," *MIS Quarterly*, (24:1), pp. 3-41.
19. Smith, A., and Anderson, J. 2014. "Ai, Robotics, and the Future of Jobs," *Pew Research Center* (6).
20. Teodorescu, M.H., Morse, L., Awwad, Y. and Kane, G.C., 2021. "Failures of Fairness in Automation Require a Deeper Understanding of Human-ML Augmentation," *MIS Quarterly*, (45:3).