Trust, Apprehension, and ChatBots: A Study On The Efficacy of AI in Pedagogical Settings

Introduction

This paper explores the intersection of machine learning and artificial intelligence (AI) through a comparative analysis of educational and vocational contexts. By examining studies such as Feng & Wang (2023) and Lei et al. (2023), it highlights how AI-assisted learning tools impact student engagement and motivation, particularly among tech-literate groups. The paper further distinguishes between machine learning as a tool for data analysis and AI as an evolving entity capable of adapting based on user interaction. It emphasizes the importance of trust and control in human-AI relationships, arguing that effective machine learning applications rely on a balanced approach between automation and manual oversight. The findings suggest that fostering a "community of inquiry" and interactive learning environments enhances AI literacy and efficacy. Future research should focus on the philosophical implications of AI's self-reflective nature and the role of interactive learning in improving machine learning outcomes.

1. Key concepts and historical overview of machine learning

Machine learning is a branch of AI and serves as a pathway to grant a given technology the eponymous *artificial intelligence*. While the term "AI" lends itself to a slew of connotations throughout culture and business, it is defined primarily by its capacity to learn from human behavior and thereby simulate the cognitive functions of its users. A predictive algorithm does not qualify as exhibiting artificial intelligence if it is explicitly programmed and never changes its thinking based on user feedback. For example, predictive text algorithms have existed for a long time but have not always been artificially intelligent. An actuating keyboard, particularly one for Chinese script, makes its word processing more efficient by suggesting new characters based on the one just used. This makes it so that the user does not have to look for one specific character out of a thousand but rather just out of a handful of likely ones. This is programmed based on natural language but does not account for style nor the typing habits of the user (McClure, 2012). An AI-powered predictive text algorithm has the capacity to modify what text it suggests based on the overall trends it records in how the user interacts with the word processor. By being able to change over time and adapt to user behavior, the technology learns and thus qualifies as AI.

The main connection this conception of AI has with machine learning is that they are both programmed with an intent to gather data and recognize patterns within it, more so than with an explicit, unshifting function. To this end, machine learning assists the programmer with the programming of itself. In American computer programmer Arthur Samuel's 1959 paper that originally coined the term, he predicted that programming computers to learn from experience would eventually eliminate the need for detailed programming effort. While this hypothesis was premised on his work programming computers to play (and win) checkers against humans, the basic premise of allowing machines to improve through trial and error remains revolutionary. Furthermore, the characteristic of being willing to fail in order to improve is what distinguishes AI as specifically human-centric. And as we will see, AI's functionality is dependent on how much its human users see themselves reflected in the algorithm itself (Finn et al., 2023).

For those working with the data hands-on, machine learning serves the purpose of identifying patterns within data sets and building predictive models. This tool is crucial for generalizing huge swaths of data. Furthermore, the data that the algorithm is trained on is provided and curated by the data scientists rather than being fed to the algorithm directly. The results are then evaluated and adjusted as needed, often by supplying more data if the model generated is overfitted on an evidently narrow-minded data set (Fawcett & Provost, 2013). This is to say that in data analytics the intent is always on refining the data being gathered, preparing it for analysis, and gleaning information from it––machine learning is simply the tool by which analysts reason with it (Bailie, et al. 2022). This is the difference between machine learning and AI: with the former the user has direct access to the data and remains interested in curating it towards desired ends. On the other hand, with AI, the program can be understood as an end-in-itself with the user's interest lying with the intent to understand it and thereby better understand themselves. This aspect of selfhood greatly affects how laymen interact with machine learning algorithms to varying degrees.

2. Findings and overview of predominant trends

The main feature of machine learning is its capacity to *learn*, so naturally we will begin this overview by looking at how machine learning algorithms have been employed in educational contexts. Branching out we will see how similar technology affects those in vocational contexts and then in contexts of daily life. This will provide a fluid shape to the findings that reflects how the use of machine learning as either passive or active determines its affectual response as well as how technological literacy varies between age groups. To begin, Feng & Wang (2023) offer a study that compares the use of AI-assisted learning with traditional, rote, pen-and-pencil learning within a Chinese primary school classroom focusing on Chinese-English bilingual reading. This study looks at left-behind and migrant students in particular with an intent to bridge the gap between their media-consumption tendencies and their learning habits. The study also emphasizes how this generation finds itself bored with book learning when not predisposed towards high achievement, which is expected of students marginalized in one way or another.

Different from other e-reading software, the educational technology used in the Feng & Wang (2023) study employs predictive algorithms to "[enable] audio reading of bilingual picture books, mutual translation, real-time scoring and correction, as well as the ability to scan augmented reality (AR) image books" (p. 2). The robot not only enhanced the bilingual performance of the students, but also encouraged independent reading habits and overall motivation. Before discussing the control factors that may have contributed to the success of this study, the performance of the technology itself is worth remarking on. Similar to machine learning tools, students are able to supply the robot's AR camera with content from their homework or books. This allows the student to determine what data the algorithm will feed in and, by "shadowing-reading aloud," make sense of the text with assistance by the robot (p. 2). This simulates the tutor-pupil relationship while allowing the student to control by what means he or she wishes to learn as well as by what pace.

The multi-media approach to learning is salient because it allows for a greater interactivity for the student than what textbooks and worksheets would traditionally allow. In addition to the benefit of interactivity it encourages discovery, which in turn facilitates independence in personal learning activities. Regardless, the paper also measures anime and video game consumption in the group before and after receiving the treatment, with a 64% decrease as a result (p. 8). This implies that these students are already being trained on a heavily mediated diet of entertainment prior to the study, suggesting that the AI-assisted learning supplants an existing tendency rather than deterring it. This is supported by how the UI of the robot is designed to resemble cartoons and prime the children to approach the AI as a friendly entity. It also suggests that these students are exceptionally tech-literate, meaning that this method would work better with this population than with others.

The Feng & Wang (2023) study suggests that humans can cooperate with machine learning towards their best interest as inquisitive beings, but the data is still too narrow to confirm that the results would sustain long-term. A study by Lei et al. (2023) also looks at a group of primary school students learning English as a second language but controls for the factors previously mentioned. These being digital literacy and the priming of the students to respond to the AI agent comfortably. The study confirms that students feel less pressured working with an AI agent than with a teacher, though only when there is a greater inclination to interact with the algorithm. Students lacking the necessary trust or digital literacy interact with the learning material with a “surface approach” wherein they “spend the least amount of effort to meet the minimum learning requirements.” In other words, as if the program were just another authority figure. This is in direct contrast with students approaching the work on a “deep” or “organized” level and “join technology-enhanced learning activities with intrinsic motivation” (p. 4-5). The apprehension by the ‘surface approach’ students is understandable given that a lack of feedback causes the machine learning to work less effectively, leading to the same sense of struggle a traditional classroom would impart on such students. Ironically, those using the 'surface approach' ignore the AI-assistance and resort to rote memorization and re-reading.

Again, Lei et al. (2023) controls for environmental factors and recommends that a "community of inquiry" is crucial for human-AI learning to take place and to yield any positive results. This means that trust in the AI is instilled via a predisposition to do so, supported by a given environment or culture (much like the one of the Chinese students in the Feng & Wang study). This poses a nearly Pavlovian question of how to imbue a positive reaction to AI in a person where there is no such relationship, but this road of inquiry is both out of our disciplinary wheelhouse and contrary to what technological culture should try to manifest. Lei et al. (2023) concludes their study with a suggestion that students apprehensive about AI can have the human instructor intervene and "apply new knowledge acquired from learner-AI interactions to new settings" (p. 14). Meaning, data collected by the algorithm can be used to guide the student depending on their personal learning style, but the AI itself does not necessarily have to be doing the guiding as well. This affirms my point about machine learning as being more so a tool for data collection and refinement rather than as a kind of robot companion. Furthermore, it appears that control in alternating manual and automatic use of the AI is essential towards achieving this end.

In higher education, the use of chatbots powered by large language models (LLMs), like ChatGPT and Bard, have become ubiquitous. Because there is a lack of publicly-available studies on this fast-growing topic, I ventured to analyze a dataset on my own. This survey conducted by the Industrial Engineering department of a research university in Jordan constitutes the data (Abbad et al., 2023). Students at German Jordanian University were polled on how the use of LLM-based chatbots have impacted their learning experience. On average, AI tools had a favorable view by students and were thought to improve their quality of education. A stacked linear regression proved that the chatbots were primarily being used to help students catch up with their coursework and improve poor grades, which led to this positive view on AI (fig. 1). While the positive effect AI has had for struggling students is undeniable, it cannot be said that they are using the technology in an iterative fashion conducive to learning. Information recall, although reported to be improved on the whole, did not correlate with any other trend.

Recalling the Feng & Wang study, the students at German Jordanian University appear to be employing the "surface approach" to learning with chatbots. This is evident by the lack of actual comprehension of the material exhibited in students that use AI tools the most. Ironically, these same students had the most favorable perception of the technology and saw the most improvement overall, though they did not hold the highest grades (fig. 2). Because chatbots are neither deliberately implemented nor explicitly permitted at this university, it cannot be said that this particular setting was emblematic of a "community of inquiry" as professed in Lei et al. With LLM-powered chatbots, the repository of data at its disposal is so vast that the user can hardly affect the iterative capacity of the machine learning algorithm at play. This prevents a "community of inquiry" from taking place and limits the technology to the function of that of a cheatsheet. By contrast, successful studies saw the AI being perceived as a buddy or something of a mirror to the user's own curiosity.

Figure 1

Students using the technology to catch up also cut down on time spent studying

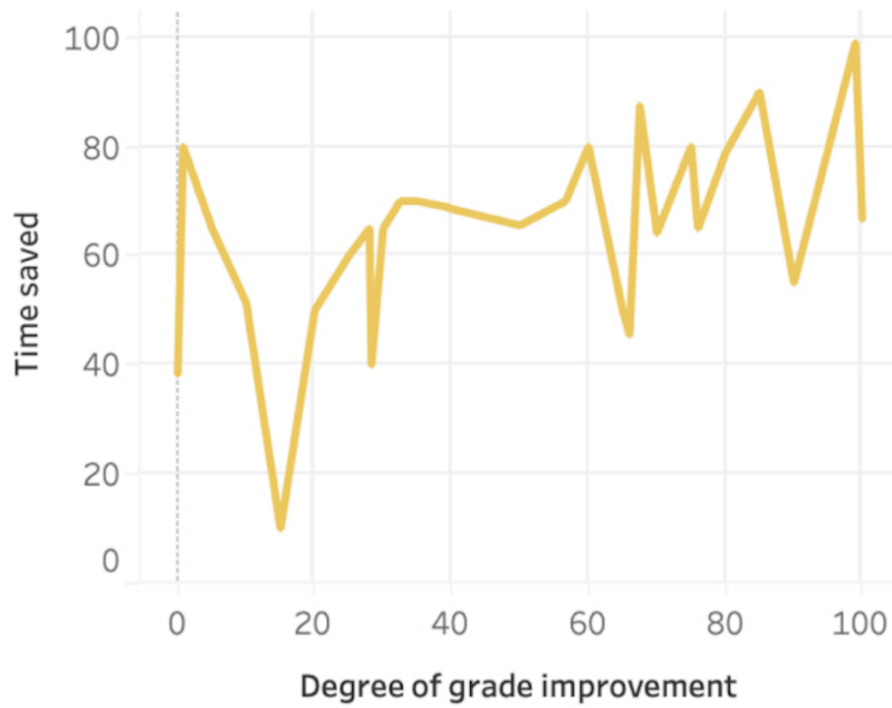
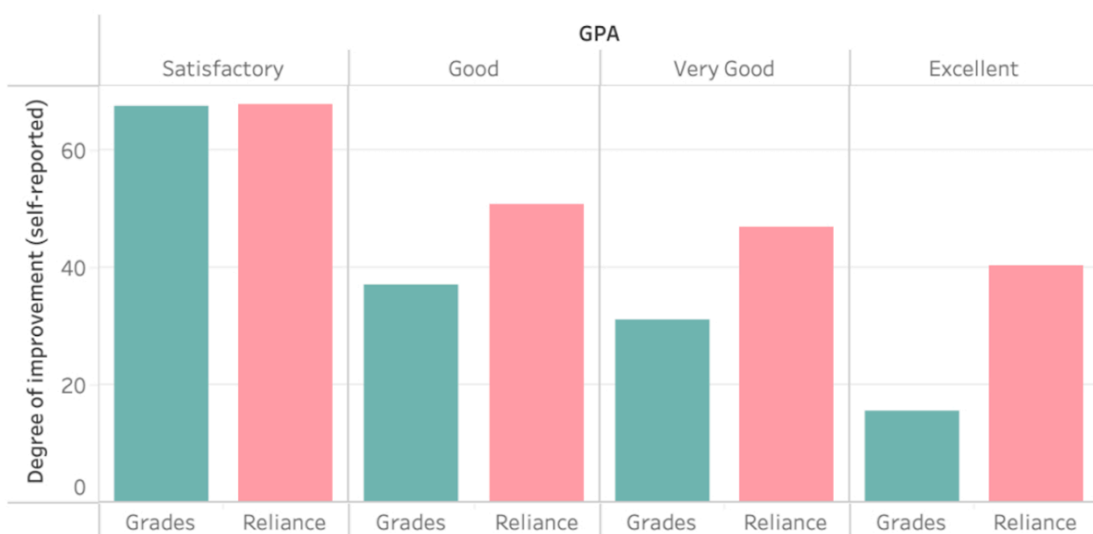


Figure 2

Students with lower GPAs saw the most improvement with AI tools, but also relied on them the most



In cases for primary school students, designing the AI agent in the form of a cartoon character helps facilitate trust and encourage constant interaction. But in most cases, the AI is considered a “black box” that users are forced to construct a “mental model” for in order to approach it a certain way. A study by Finn et al. (2023) showed that priming the user to imagine the AI as either caring, manipulative, or neutral correlated with the efficacy of the program itself (p. 1076-1078). Thinking poorly of the program in advance caused users to form their own conclusions about the results, thereby retarding mutual positive interaction. Naturally, those expecting the program to perform well were more open to results and facilitated better performance for the AI (p. 1081). While this confirms the necessity of trust in the human-AI relationship, those provided with a neutral mental model had only their preconceived notions to rely on and no opportunity to learn about the technology itself. A study by Barda et al. (2020) observed how patients in a hospital setting confide trust in predictions made by a machine learning model only to the degree that the supervising doctor or nurse successfully explains how the model works. This means that a layman’s success using machine learning models in their daily life is dependent on either cultural connotation or arbitrary supervisors standing-by.

In situations where adults use machine learning models for their jobs, the technology itself is reserved to automate repetitive and time-sensitive tasks. This is not so different from data scientists that use machine learning to generalize impossibly vast sets of data. An old study by Lee & Moray (1992) finds that operators in a pasteurization plant work best with automated controllers when they are given a choice between automatic and manual control. While this study was conducted in the 90s, the efficacy of the machinery was still contingent on trust with the technology itself, much like how AI operates today. Keeping this in mind, there was also an issue of too much trust placed in the automated controllers, wherein operators became “complacent, and [failed] to intervene when the technology failed” (p. 1243-1244). This affirms the need to modulate constantly between automatic and manual control, much like how an AI-assisted classroom would need human instructor intervention to clarify the salience of the data gathered. Furthermore, this reliance on automation subscribes to a passive framework also found in those students in Lei et al. (2023) that interact with the AI on a surface level. Those working in the niche field of interlingual respeaking that utilize automatic speech recognition (ASR) also face this challenge, wherein they must continuously revise the suggestions made by the AI in real time. In these fast-paced conditions, AI serves solely to automate repetitive tasks without any space for users to interact with the data itself (Evans et al., 2023, p. 1-2). This signals another instance of passive-yet-involved interaction with AI, wherein "sustained attention was not associated with [high accuracy]" in a "time-pressured process with high levels of task demand" (p. 9). Whatever cooperation facilitated between man and machine here would not necessarily sustain long-term. This harkens back to the 'surface approach' to human-AI interaction that uses the technology solely to meet the minimum requirements of the task at hand.

3. Evaluation of findings

Similar to my description of the data scientists at work, the failure of such automated technology reveals what needs to be adjusted in the data. This goes to show that trust in machine learning is contingent on its fallibility: the experience of trial-and-error that humans may relate to and subsequently reason with towards mutual benefaction and increased intellect (Zuboff, 1989, p. 69). The key to this kind of result appears to be a modal approach to human-AI interaction that allows the user to alternate between automatic and manual control of the technology in order to facilitate a relationship not unlike trust between humans based on "performance… individual differences… predictability, dependability, and faith" (Lee & Moray, 1992, p. 1286). To this personalized degree, the automatic/manual dichotomy finds a parallel in the alternating rift between faith and control. In other words, trust would be conditioned by experience with the technology as well as the proper space to interrogate its functions at will––an aspect blocked off in situations like the one described by the Evans et al. (2023) study. To this point, user-centered design should “convey performance information” in “creative, understandable ways… dynamic and tailored to the context at various levels” (Benda et al., 2021, p. 210). Otherwise, trust in the AI “guides reliance when complexity… make[s] a complete understanding of the automation impractical” (Lee & See, 2004, p. 50). In such cases the cultural connotation, the personification of the AI-agent, and the aptitude of supervisors to explain the purpose of the AI defines the success of the technology for use by laymen. Suffice it to say this framework is unreliable and depends on its users’ passivity for success.

Recalling the students in the Lei et al. (2023) study who approached AI apprehensively as if it were an authority figure, success must be premised on a sense of independence and control in order for feedback to be provided willingly. Just like the pasteurization plant workers in Lee & Moray (1992), primary school students approach technology (intelligent or not) as a means to an end. This surface approach treats the data as authoritative and unchanging, which prevents the introduction of new or altered data when needed. The deep approach, by contrast, involves data in a reflexive relationship with the algorithm. By iterating between these two poles, the “deep approach” user is able to experience the subject from countless different angles. When the data is presented as authoritative, as it would be as the study material for an exam, rote memorization appears to be a less meandering route to understanding it. The introduction of AI tools reveals the overly orthodox methods of learning that fail to keep up with this advancing technology and the iterative approach to learning that it offers.

Differing from the Evans et al. study, the goal in a “deep approach” must be self-improvement with simultaneous refinement of the AI’s pattern-recognition abilities, rather than fulfilling the minimum requirements of a given task. This is contingent on a given environment being a “community of inquiry” that encourages experimental and multi-faceted interaction with the AI in order to garner as much knowledge as possible from it. While this concern belongs more to the domain of sociology, it is still an essential factor in promoting machine learning-literacy. Those with predisposed apprehension towards AI will otherwise be left behind, resorting to passive reliance on algorithms or total rejection of the emerging technology. In such cases, test subjects "tend to reject recommendations provided by a [machine learning] model even when the model outperforms human predictions, leading to the phenomenon of 'algorithm aversion'" (Chiang & Yin, 2022, p. 149). To bridge the gap between the averse and the willing, machine learning "literacy interventions are perceived to be more understandable and slightly more useful when they are interactive" (Chiang & Yin, 2022, p. 159). This "sandbox" approach allows users to control what data they feed into the algorithm and test its functionality accordingly. To this point, the machine learning-literacy being promoted is more effective when the user has a range of experiences to draw upon.

Conclusion

Generative human-AI interaction relies on trust in the system provided by a degree of literacy in the machine learning algorithm itself. The "community of inquiry" recommended by Lei et al. (2023) promotes this literacy better than plain explanation as it permits the user to experiment with the algorithm, relate the machine learning models to their own life and environment, and ultimately control and refine the data that the models implement. This goes to show that machine learning's main potential is not simply as a tool for completing a task, but a method to inquire into the meaning of a provided dataset, which in itself is a reflection of the user's own reality. In turn, machine learning may provide insight into the user's learning habits and work ethic and help improve these factors all the while improving itself. To reiterate the earliest definition of machine learning, it allows for the user to program the computer by interaction alone rather than by explicit instruction.

The development of machine learning reflects humanity's own learning process that is iterative, creative and receptive to new experiences. AI, similarly, constantly seeks out new data with which to experiment with and generate new possibilities, learning all the while. As in the analysis of students at German Jordanian University, the popular application of AI betrays its inherent instincts by limiting the scenarios it may iterate with and thus generating something of a negative feedback loop (Abbad et al., 2023). Our conception of AI is limited by thinking of it as exponentially growing in cleverness until it may outpace us. Rather, it is more like humanity than we realize, in that overspecialization leads it to become burnt out and uninspired, much like ourselves. As with our example with Arthur Samuel's checkers program, specialized uses of machine learning eventually uncover a set amount of possible games to play and can no longer learn anything new. These specialized algorithms have their place, but may lead users astray when made to assume that the AI is a reflection of themselves. Inversely, embracing the underlying function of machine learning may put users better in touch with the data that powers AI. A technologically advanced society and a "community of inquiry" must support users as though they were all lay-data scientists, capable of refining and altering the data that comes to shape their world.

Sources

Abbad, W., Ammouri, M., Elashi, M., Qaadan, S. (2023). Data-Driven Insights into the Effectiveness of Artificial Intelligence Tools in Education. Available at SSRN: https://ssrn.com/abstract=4609385 or http://dx.doi.org/10.2139/ssrn.4609385

Baillie, M., Le Cessie, S., Schmidt, C. O., Lusa, L., Huebner, M., & for the Topic Group "Initial Data Analysis" of the STRATOS Initiative. (2022). Ten simple rules for initial data analysis. *PLOS Computational Biology*, *18*(2), e1009819. https://doi.org/10.1371/journal.pcbi.1009819

Barda, A. J., Horvat, C. M., & Hochheiser, H. (2020). A qualitative research framework for the design of user-centered displays of explanations for machine learning model predictions in healthcare. BMC Medical Informatics and Decision Making, 20(1), 257. https://doi.org/10.1186/s12911-020-01276-x

Benda, N. C., Novak, L. L., Reale, C., & Ancker, J. S. (2021). Trust in AI: Why we should be designing for APPROPRIATE reliance. *Journal of the American Medical Informatics Association*, *29*(1), 207–212. https://doi.org/10.1093/jamia/ocab238

Chiang, C.-W., & Yin, M. (2022). Exploring the effects of machine learning literacy interventions on laypeople's reliance on machine learning models. *27th International Conference on Intelligent User Interfaces*, 148–161. https://doi.org/10.1145/3490099.3511121

Evans, S. L., Davitti, E., & Wallinheimo, A.-S. (2023). Training in new forms of human-AI interaction improves complex working memory and switching skills of language professionals. Frontiers in Artificial Intelligence, 6, 1253940. https://doi.org/10.3389/frai.2023.1253940

Fawcett, T., & Provost, F., (2013). Data science for business: What you need to know about data mining and data-analytic thinking. *O'Reilly Media, Inc.*

Feng, Y., & Wang, X. (2023). A comparative study on the development of Chinese and English abilities of Chinese primary school students through two bilingual reading modes: Human-AI robot interaction and paper books. Frontiers in Psychology, 14, 1200675. https://doi.org/10.3389/fpsyg.2023.1200675

Finn, E., Liu, R., Maes, P., & Pataranutaporn, P. (2023). *Mitmedialab/nmi-ai-2023: V1. 0* (public) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.8136979

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), 1243–1270. https://doi.org/10.1080/00140139208967392

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors: The Journal of the Human Factors and Ergonomics Society, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50.30392

Lei, J., Li, L., Liu, Q., Pang, H., Tan, S. C., Wallace, M. P., & Wang, X. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education*, *194*, 104703. https://doi.org/10.1016/j.compedu.2022.104703

Mcclure, M., & University, S. (n.d.). *Chinese typewriter anticipated predictive text, finds historian*. Retrieved November 27, 2023, from https://phys.org/news/2012-11-chinese-typewriter-text-historian.html

Zuboff, S. (1989). *In the age of the smart machine: The future of work and power*. Basic Books.