



The process is the product: a review of *co-intelligence* by Ethan Mollick

John Bartucz¹

Received: 20 October 2024 / Accepted: 2 December 2024
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2024

Abstract

Co-Intelligence: How AI and Humans Can Work Together by Ethan Mollick presents an optimistic view of the future in which Generative AI tools complement and enhance human creativity, productivity, and decision-making. Mollick skillfully explains how healthcare, education, and businesses of all types will benefit from incorporating AI tools. However, the book falls short in addressing the critical ethical issues surrounding Large Language Models (LLMs). While Mollick touches on these topics, he fails to engage deeply with topics such as intellectual property, energy usage, and exacerbating inequalities. This review offers a critical analysis of Co-Intelligence, emphasizing the need for a deeper exploration of ethical frameworks to guide AI development.

Since ChatGPT 3.5 was made public on November 30, 2022, Ethan Mollick's publications and blog have established him as a preeminent authority on AI's practical applications and societal impact. Although not a computer scientist by training, his exploration and deep understanding of the methods used to create large language models (LLMs) informs his insight into the optimal ways to leverage their power. The premise of his book, *Co-Intelligence: How AI and Humans Can Work Together* [1], is that AI must necessarily become our partner in almost all aspects of life, from education to health care to business. The vision he offers is that instead of replacing humans, AI will augment our abilities, ushering in a new era of human–machine partnership. While the book offers valuable insights into the potential of AI, it falls short in addressing critical ethical concerns and the almost certain exacerbation of global societal inequalities.

Mollick demonstrates how AI could enhance human creativity, productivity, and decision-making in diverse fields such as education, healthcare, business, and scientific research. His ability to articulate the transformative potential of AI in these sectors is admirable, offering readers a glimpse into a future where human–AI collaboration leads to unprecedented advancements. In fact, one of the book's greatest strengths is in Mollick's ability to explain complex

technological concepts in layperson's terms without sacrificing nuance or depth. This approach allows readers from various backgrounds to grasp the implications of AI integration in their respective fields, fostering a broader understanding of the technology's potential impact.

However, despite its merits in exploring AI's potential, “Co-Intelligence” falls short in several critical areas, particularly in its failure to consider deeper ethical considerations and long-term societal consequences. While Mollick offers valuable insights into AI's potential, the book misses a pivotal opportunity for further exploration of the ethical frameworks that could guide AI development and deployment. With tens of thousands of copies selling in the first few days, *Co-Intelligence* immediately made the New York Times bestseller list. Mollick had the chance to raise our collective consciousness about the ethical process of building, delivering, and using AI, and chose not to engage that discussion. One extra chapter in this area would perfectly complement Mollick's visionary perspective on human–AI collaboration.

While he occasionally touches upon ethical concerns, these discussions fail to engage with the depth and complexity these issues demand. Beginning with the section titled “Artificial Ethics for Alien Minds,” Mollick brings up six points for consideration: first, AI can perpetuate and amplify societal biases. Second, AI can be used to manipulate people by quickly creating realistic-sounding messages from sources that the recipient trusts. Similarly, Mollick demonstrates how LLMs can be tricked into providing dangerous information, such as instructions on how to make

✉ John Bartucz
jbartucz@umn.edu

¹ Department of Computer Science and Engineering,
University of Minnesota, Minneapolis, USA

napalm. Third, he mentions that AI might lead to job displacement, although he only references one study that claims it will mostly affect “highly skilled and highly educated workers, as well as workers in creative and analytical fields.” Finally, he briefly mentions that over-reliance on AI can lead to decreased skill and judgment in humans, AI can make it easier for students to cheat, and AI could undermine apprenticeship-based learning.

All of that being said, the book's optimistic and inspiring tone tends to gloss over the more troubling aspects of widespread AI adoption. In this new era of wide-eyed admiration for one of the most powerful technologies ever invented, Mollick mostly hand-waves away the ethical issues and does not explicitly address potential solutions. For example, in the section on education, he does not cite research that AI tutors may actually harm students' learning [2], although he has more recently addressed this in public forums [3]. He also fails to mention that AI is already using as much electricity as a small country [4]. Finally, though he briefly mentions that “low-paid” Kenyan, Ugandan, and Indian workers filter out toxic, racist, and pornographic material, he does not mention the fact that they are paid less than \$2 per hour, or that they are currently filing lawsuits against the firm who hired them as subcontractors for OpenAI for the working conditions and mental trauma that they have suffered (Perrigo, 2023; Rowe, 2023). He does not state how this work could be done in the future or by whom, or even that we should be considering these issues before jumping on the AI bandwagon. If we accept these issues as “the cost of doing business” with AI as Mollick implies we should, then we are no longer bystanders. We become complicit in these activities by supporting the companies that use these processes to create and run their models.

Instead of hand-wringing over individual issues, we should approach AI through the lens of an ethical framework to help guide us. Let us consider, for example, the utilitarian approach to ethics, which focuses on maximizing overall well-being or happiness for the greatest number of people. A more rigorous analysis would explore how AI systems, designed to optimize specific outcomes, might align with or diverge from utilitarian principles. Questions such as “How do we ensure AI systems consider the well-being of all stakeholders, not just the most visible or quantifiable ones?” or “How do we balance short-term efficiency gains with long-term societal impacts?” are largely absent from Mollick's analysis. As he admits, virtually all of the training data in ChatGPT was scraped from the history of Western, English-speaking countries, which espouse Western ideals, ideology, and exceptionalism. Unless the user explicitly decides to give another context, any suggestions provided by such an advisor will tend to support Western, white, capitalist supremacy.

Perhaps it would be more appropriate to use a deontological framework since there is no “end-point”; AI will never be done, it will always be in development. If we only look through the practical lens of “how can we immediately use AI to our benefit,” we ignore the fundamental principle of deontology: the ends do not justify the means. Deontology requires us to ask the question—would we want the current practices (as outlined above) to become universal laws? The lack of transparency in AI systems and the absence of any kind of consent violate the duty of truthfulness. Questions of moral agency and responsibility surrounding both the acquisition of training data and the potential harm caused by AI remain unanswered. To align AI with deontological principles, we must prioritize human dignity, develop universalizable practices, increase transparency, clarify moral responsibility, consider long-term impacts, and ensure AI enhances rather than undermines democratic societies.

Mollick's vision of human-AI collaboration, while inspiring, assumes a level playing field that simply doesn't exist. The author doesn't explore how differences in access to AI technologies, digital literacy, and educational opportunities could accelerate these inequities. It does not seem a stretch to say that those with the easiest access to AI or with the most creative and supportive teachers have the potential to gain the most. It is clear that those with greater access to technology outpace the rest of the world in terms of income, trade, and virtually every other economic dimension [5]. John Rawls' “veil of ignorance” thought experiment could have provided a platform for considering how AI should be developed: if we were randomly assigned to any position in society, anywhere in the world, how would we want AI to be built, and how would we want it to work? What specific measures can be taken to ensure that AI doesn't simply amplify existing power structures and inequalities? Most importantly, how do we convince those already in power that these goals are worthwhile?

As we cross the threshold into a “Co-Intelligent” future, it is crucial that we engage in deeper, more critical discussions about the ethical frameworks that should guide AI development and use. Readers should approach this book as an excellent primer on the potential of human-AI collaboration, but only as a starting point. We must supplement Mollick's optimistic vision with a commitment to addressing the complex societal implications of this transformative technology. History is littered with examples of morally reprehensible processes being used to “advance society.” Envisioning a brighter future is always the first step, but we must take heed; the process is also the product.

Author contributions J.B. wrote the review in its entirety.

Funding The author received no funding from public, commercial, or not-for-profit entities.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

1. Mollick, E.: Co-intelligence: Living and working with AI. Portfolio/Penguin (2024)
2. Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakçı, Ö., Mariman, R.: Generative AI can harm. *Learning* (2024). <https://doi.org/10.2139/ssrn.4895486>

3. Elon webinar focuses on challenges and opportunities of AI in higher ed. (2024, March 12). Today at Elon. <https://www.elon.edu/u/news/2024/03/12/elon-webinar-focuses-on-challenges-and-opportunities-of-ai-in-higher-ed/>
4. de Vries, A.: The growing energy footprint of artificial intelligence. *Joule* 7(10), 2191–2194 (2023). <https://doi.org/10.1016/j.joule.2023.09.004>
5. Sampson, T.: Technology gaps, trade, and income. *Am. Econ. Rev.* 113(2), 472–513 (2023). <https://doi.org/10.1257/aer.20201940>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.