

# Review of Co-Intelligence: Living and Working with AI

**Joel Blechinger**

*Mount Royal University Library*

ORCID: 0000-0002-4907-3601

ABSTRACT: Review of Mollick, E. (2024). *Co-intelligence: Living and working with AI*. Penguin Portfolio.

Keywords: generative artificial intelligence, labour



This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Journal of Radical Librarianship*, Vol. 10 (2024) pp.63-65. Published 20 May 2024.

Since 2022, Ethan Mollick, a Professor of Management at the University of Pennsylvania's Wharton School, has emerged as an influential popular commentator on generative artificial intelligence (GenAI) and its impact on work and education. His Substack, *One Useful Thing*, has over 147,000 subscribers and he has close to 212,000 followers on X. Mollick's new book, *Co-Intelligence: Living and Working with AI*, arrives to capitalize on this expanded readership and to shore up his reputation as a leading GenAI futurist.

*Co-Intelligence's* nine chapters are split across two parts. Part I is short, providing a surface-level overview of artificial intelligence's history as a field, discussing the "alignment problem," and ending with Mollick's "four rules for co-intelligence": (1) always invite AI to the table; (2) be the human in the loop; (3) treat AI like a person (but tell it what kind of person it is); and (4) assume this is the worst AI you will ever use. Part II comprises six chapters, all with similar titles ("AI as a Person," "AI as a Creative," "AI as a Coworker," etc.) where Mollick fully indulges his futurist tendencies, holding forth on how GenAI is changing—and, more crucially, will change—creative production, labour, education, mentorship, and, eventually, our future.

I found it instructive to read *Co-Intelligence* not for its content—which is, in large part, little more than scantily researched trend forecasting more suited to informal online publishing genres like X threads and Substack posts—but for its style of argumentation. Repeatedly throughout the text, Mollick rhetorically performs awareness of ethical issues raised by GenAI whether related to copyright concerns, the exploitative labor practices involved in GenAI training, or the environmental impacts of data centers. Performing this awareness, however, merely grants him license to further the main aims of *Co-Intelligence*: building the case for "prompt engineering" and, more generally, advocating for a fulsome embrace of GenAI across all the economy's sectors with a particular focus placed on the teaching and learning context (see the book's "AI as a Tutor" and "AI as a Coach" chapters).

Reading *Co-Intelligence* is particularly useful for librarians because Mollick's argumentation style is one that we presently see downstream in library discourse. Every presentation on GenAI and libraries that one takes in contains the obligatory "ethical issues" slide (or slide set) before the speaker inevitably moves on—more or less disjointedly—to a discussion of how to teach with or about GenAI, use it to triage virtual reference, integrate it into collections work, etc.

As an example of Mollick's performative awareness, consider the following brief aside in his "Aligning the Alien" chapter during a discussion of Reinforcement Learning from Human Feedback (RLHF), a GenAI training method employed by companies like OpenAI to "detoxify" their models before release to a broader market where severely underpaid contract labor is exposed to harmful outputs that describe phenomena such as murder, incest, and child sexual abuse. Mollick (2024) begins, "It's important to note that the process is not without human cost"—a sentence written in a distinct, nonspecific but vaguely moralizing style strikingly similar to ChatGPT. He continues:

Low-paid workers around the world are recruited to read and rate AI replies, but, in doing so, are exposed to exactly the sort of content that AI companies don't want the world to see. Working under difficult deadlines, some workers have discussed how they were traumatized by a steady stream of graphic and violent outputs that they had to read and

rate. In trying to get AIs to act ethically, these companies pushed the ethical boundaries with their own contract workers. (p. 38)

We have to examine the book's notes to find that Mollick is obliquely referencing Perrigo's (2023) vital reporting for *Time* early last year, which broke the story of the "hidden labor" behind GenAI to a wider audience. It takes Mollick a mere eight pages, however, to move from this box-checking description of labor exploitation (to which he provides no satisfying resolution) to his first "rule for co-intelligence": always invite AI to the table. This is just one of many moments in *Co-Intelligence* where the text seems to lack a moral lodestar.

This style of argumentation reaches an apotheosis when Mollick explains his third rule for co-intelligence: Treat AI like a person (but tell it what kind of person it is). He begins this section with a solid two pages cautioning against anthropomorphizing GenAI. Says Mollick (2024), "Even the developers and researchers who design these systems can fall into the trap of using humanlike terms to describe their creations. We say that these complex algorithms and computations 'understand', 'learn', and even 'feel', creating a sense of familiarity and relatability but also, potentially, confusion and misunderstanding" (p. 55). Having adequately demonstrated his awareness of issues related to anthropomorphization—his conscience now clear—Mollick can then turn to the gist of his principle, which is, essentially, constant practical anthropomorphization: "[A]s imperfect as the analogy is, working with AI is easiest if you think of it like an alien person rather than a human-built machine. So, let's start sinning. Imagine your AI collaborator as an infinitely fast intern, eager to please but prone to bending the truth" (p. 57). This type of argumentative incoherence creates a whiplash effect in the reader where, suddenly, anthropomorphization is not only morally sound, but we can use it to maximize our productivity gains from GenAI, even if this means extending that anthropomorphization not just to a person but to "an infinitely fast intern." Notably, Mollick feels no need to explore the political import of analogizing between a GenAI chatbot and an intern (Perlin, 2011), and it is hard to determine whether this comes from willful ignorance or a more insidious impulse.

Mollick's "AI as a Creative" chapter is problematic in other ways, the most obvious example being when he reductively compares all human creativity to idea generation and sees no meaningful difference between the history of art making and "com[ing] up with 20 ideas for marketing slogans for a new mail-order cheese shop" (p. 106). This position, in turn, allows him to ominously imply that human creative practice is at risk of being "out invented" by GenAI. Confusingly, however, Mollick can also be quite a clear communicator when articulating GenAI's limitations. At the beginning of the chapter, when discussing the GenAI hallucination problem, Mollick conveys why large language models (LLMs) fabricate textual output with precision: "hallucinations are a deep part of how LLMs work. They don't store text directly; rather, they store patterns about which tokens are more likely to follow others. That means the AI doesn't actually 'know' anything. It makes up its answers on the fly" (p. 94). Lucid moments like this in *Co-Intelligence*—which Mollick frequently follows with underdeveloped futurist prognostications—lend the text a wildly uneven quality in addition to the abovementioned moral incongruity.

Unfortunately, *Co-Intelligence* reads more like a book that has been rushed to market to meet demand and less like one that needed to exist because its author had interesting or vital ideas to

share about artificial intelligence. It is telling that, multiple times in the text, Mollick integrates textual output from ChatGPT: he bizarrely uses it to admonish him to write *Co-Intelligence* in the first place, he quotes its responses to prompts throughout the book, and he even admits to having employed it to summarize a piece of research that he glosses towards the end. In a way wholly unintended by Mollick, then, *Co-Intelligence* is symptomatic of publishing in the AI age: on the surface, it has many of the formal features of a useful or engaging book but, if one looks deeper or scrutinizes Mollick's ideas at greater length, that first impression quickly fades and one is left mystified and unsatisfied. In the brave new GenAI future that he is keen to inaugurate, this will likely become an increasingly common experience for readers of texts.

## References

- Mollick, E. (2024). *Co-intelligence: Living and working with AI*. Penguin Portfolio.
- Perlin, R. (2011). *Intern nation: How to earn nothing and learn little in the brave new economy*. Verso.
- Perrigo, B. (2023, January 18). OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic. *Time*. <https://time.com/6247678/openai-chatgpt-kenya-workers/>