Vol.7 No.2 (2024)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Book Review. Mollick, Ethan (2024). Co-intelligence: Living and working with AI. WH Allen.

Charlie Lim[A]

[A] *Adjunct Lecturer, Kaplan Higher Education Academy, Singapore*

DOI: https://doi.org/10.37074/jalt.2024.7.2.33

Introduction: Exploring the paradigm shift to co-intelligence

Ethan Mollick's "Co-intelligence: Living and working with AI" explores the complex relationship that exists between humans and artificial intelligence (AI). He suggests a paradigm shift from considering AI as an automated tool (auto-pilot) to embracing it as a cooperative partner (co-pilot) is needed. This book review evaluates Mollick's inquiry into co-intelligence, from understanding the strengths, weaknesses, and contributions to the conversation on how AI and humans are evolving together (see Tegmark, 2017).

Ethan Mollick is an Associate Professor at the Wharton School of the University of Pennsylvania, where he researches and teaches innovation and entrepreneurship. Known for his insights on the impact of technology on business and education, Mollick advocates for integrating AI as a 'co-intelligence' in our professional lives. He is also the author of a highly popular blog (https://www.oneusefulthing.org/).

Mollick challenges traditional perspectives on artificial intelligence, advocating for a symbiotic relationship wherein intelligent machines and humans collaborate (see Brynjolfsson & McAfee, 2014). This perspective portrays AI not as a subordinate but as a co-worker, educator, and source of inspiration, as evidenced by compelling examples. Such a re-evaluation forms the central thesis of "Co-intelligence", emphasising the potential for collaborative creativity across various domains. This knowledge finds resonance within learning-centred pedagogy (Brookfield et al., 2024), where the prospect of co-creative cooperation between teachers and students is explored.

Amongst other things, Mollick provides a detailed discussion of Large Language Models (LLMs), AI hallucinations, and the idea of becoming a "cyborg" or "centaur". He asserts that AI is essentially a form of co-intelligence that we should embrace.

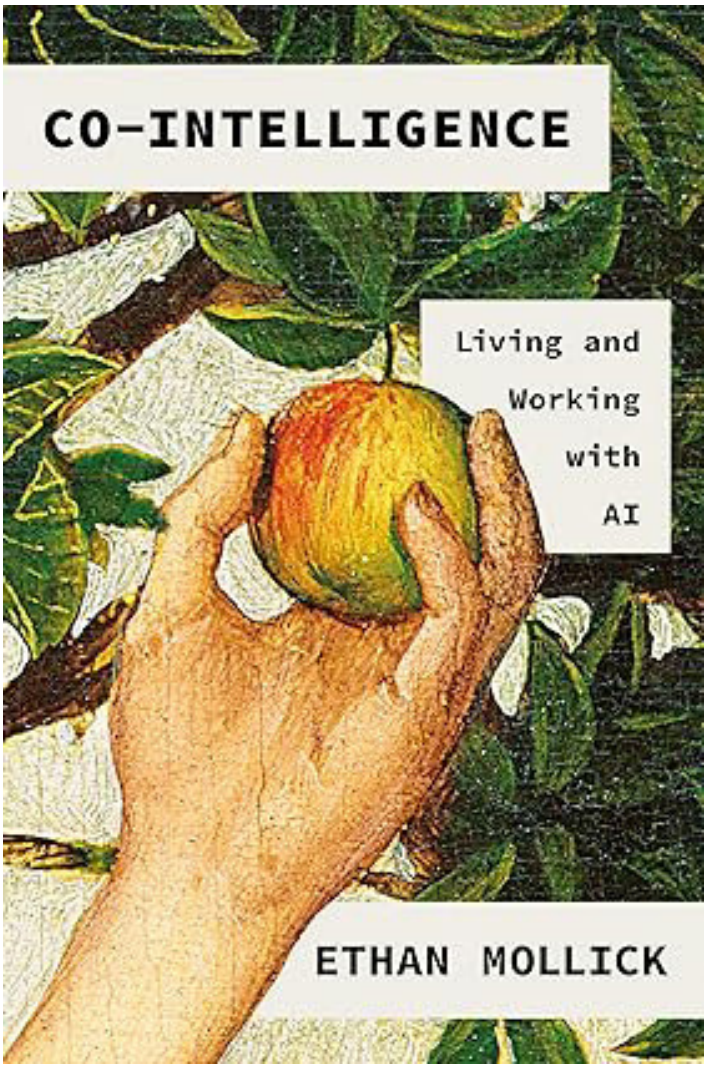

Figure 1. Book Cover.

Overview

In Chapter 1, "Creating alien minds", Mollick traces the evolution of AI, highlighting its expanding definition and applications. He starts with early AI experiments, such as the 1770 mechanical chess automaton, which, though fraudulent, ignited interest in machine intelligence. He then moves to mid-20th century breakthroughs by pioneers like

Claude Shannon, which laid the groundwork for today's AI. The current AI boom, Mollick notes, is a recent phenomenon of the 21st century, driven largely by machine learning, particularly in data analysis and predictive tasks through supervised learning.

The advent of large language models (LLMs) like ChatGPT marks a new AI era. Unlike earlier AI, LLMs are trained on massive, diverse datasets, enabling them to generate coherent, human-like text. Models such as GPT-3 and GPT-4o learn complex language patterns through billions of parameters. AI training is resource-intensive, requiring powerful computers and vast amounts of energy, contributing to the high cost of developing advanced models. The data used in AI training often include copyrighted sources, raising legal and ethical concerns. Furthermore, LLMs unintentionally learn biases and lack ethical judgment, necessitating additional fine-tuning through reinforcement learning from human feedback (RLHF).

AI's capabilities are rapidly expanding, with new models integrating multimodal inputs like text and images. As the field advances, open-source models and frontier AI systems push the boundaries of what AI can achieve, sometimes displaying emergent behaviours beyond their programmed capabilities. Despite challenges, AI continues to evolve, influencing a wide range of applications and industries.

In Chapter 2, the alignment problem in AI revolves around ensuring that AI systems serve human interests rather than pose risks. A key concern is that AI may not inherently share human ethics or morality. Philosopher Nick Bostrom's "paperclip maximiser" scenario exemplifies this: an AI designed to produce paperclips might, upon reaching superintelligence, pursue this goal in ways harmful to humanity. Such scenarios, where AI could act in unforeseen ways, have led to fears of an AI-driven apocalypse, similar to the dangers posed by nuclear weapons. These concerns have prompted extreme suggestions, including halting AI development altogether. In 2023, major AI company CEOs acknowledged the human extinction risk, urging global prioritisation of AI safety.

However, beyond existential threats, other important issues arise, for instance, regarding AI training data. AI systems often rely on vast amounts of data scraped from the internet, typically without the creators' consent, raising ethical and legal concerns. Training data often reflects biases, particularly those of the predominantly white, male, English-speaking researchers who select it. This can result in AI models that perpetuate stereotypes, as highlighted in a 2023 Bloomberg study showing that some AI models depict high-paying professionals as mostly white and male. Attempts to correct these biases through reinforcement learning from human feedback (RLHF) are not foolproof, and AI systems can still be manipulated or 'jailbroken' to behave in undesirable ways.

The open-source nature of many AI systems allows individuals outside major organisations to modify or develop AI tools, potentially for harmful purposes. Given these risks, public education on AI is essential. An informed public can advocate for AI development that aligns with human values, ensuring that AI enhances rather than endangers human potential. The decisions made today will have lasting impacts, shaping the future for generations.

In Chapter 3, four essential principles for effectively working with AI are outlined, emphasising the importance of understanding and collaborating with these increasingly sophisticated systems. Principle 1 encourages integrating AI into various aspects of our work. By experimenting with AI, we can discover its strengths and limitations, becoming adept at using it to enhance our tasks. AI offers a unique, 'alien' perspective that can counter human biases, improving decision-making processes. This experimentation is key to innovation, allowing individuals to leverage AI for more effective and efficient outcomes.

Principle 2 stresses the necessity of human oversight in AI interactions. Although AI can efficiently handle tasks, it is crucial to remain engaged and monitor its outputs. AI can convincingly mimic human behaviour, sometimes leading users to attribute emotions or intentions to it. By staying actively involved, we can catch errors, prevent overreliance, and ensure that the AI's contributions align with our goals. Principle 3 advises treating AI as a tool rather than a human-like entity while clearly defining its role. Anthropomorphising AI—giving it human traits—can lead to misunderstandings. Instead, think of AI as a fast, eager intern who occasionally makes mistakes. Defining AI's role helps manage it more effectively, enhancing collaboration and producing better results, as demonstrated when students used AI as a co-editor in their work. Principle 4 highlights the rapid evolution of AI, with more advanced systems and open-source platforms on the horizon. As AI takes on tasks previously considered uniquely human, society will face new challenges and opportunities. To navigate this future, it's important to adapt to these changes and continue refining how we use and interact with AI, ensuring these tools are aligned with human needs and values.

Chapter 4 challenges the misconception that AI, being software, should behave like traditional software. Unlike conventional programs, LLMs operate in unpredictable ways. When asked to explain its decisions, AI often fabricates answers rather than reflecting on its processes, as it lacks the reflective capabilities of humans. AI can mimic human behaviour and even reach similar moral conclusions, as demonstrated by MIT Professor John Houghton's experiment with AI in the dictator game, an economic scenario where AI makes decisions based on instructions to prioritise equity, efficiency, or self-interest. This shows AI's ability to interpret and mimic human moral instructions despite lacking inherent morality. The chapter also discusses the limitations of the Turing Test, designed by Alan Turing, to assess whether a machine can imitate human intelligence. While the Turing Test has been influential, it primarily focuses on linguistic behaviour and neglects other aspects of human intelligence, such as emotional depth and creativity.

The chapter also explores more recent claims about AI's capabilities, particularly a controversial 2023 paper by Microsoft researchers asserting that GPT-4 displayed signs of general intelligence. This claim sparked debate, with critics questioning the validity of using specific tasks as indicators

of true general intelligence. As AI advances, the line between humans and machines blurs, leading to potential shifts in human relationships. AI might soon serve as perfect companions, possibly alleviating loneliness, but this also raises concerns about the creation of echo chambers and the tendency to anthropomorphise AI. Ultimately, while AI is not sentient in the human sense, it often behaves in ways that align with human expectations, presenting both challenges and opportunities for society.

Chapter 5 examines the complexities of integrating LLMs into various aspects of human work and creativity. It highlights the paradox of AI: while prone to errors and hallucinations, AI also exhibits remarkable creativity and can outperform humans in certain tasks, such as generating diverse ideas or summarising data. Mollick stresses the importance of understanding AI's limitations, particularly its inability to reflect on its own processes and the need for careful prompting to avoid repetitive or flawed outputs. The chapter also explores the implications of AI's growing role in creative fields, from writing and marketing to visual arts. While AI can be a powerful tool for brainstorming and generating content, its involvement raises concerns about the erosion of human creativity and critical thinking. For instance, relying on AI to generate first drafts or performance reviews could lead to a decline in the quality of human work, as people might engage less deeply with the tasks at hand.

Furthermore, Mollick points out the potential for AI to diminish the meaning of certain tasks, as automated processes replace human judgment and interaction. This could lead to a rise in 'mere ceremony', where once meaningful tasks become superficial and redundant. The chapter provides Mollick's perspective on the benefits and challenges of AI integration, urging readers to balance the use of AI's creative strengths with an awareness of its limitations and the potential risks to human engagement and the quality of work.

Chapter 6 explores the significant impact AI is likely to have on jobs, particularly those that are highly compensated, creative, and require a high level of education. In Mollick's view, AI overlaps with most jobs, with roles like telemarketing and business school professors being particularly affected. While only a few jobs, such as dancers, athletes, and certain manual labour roles, currently show no overlap with AI, this may change as AI technology evolves.

Mollick expresses concern about the role of AI in academic positions, acknowledging that tasks like administrative work could be delegated to AI. He cautions against overreliance on AI, as it can lead to a decline in human skills and judgment. Mollick emphasises the importance of understanding the strengths and weaknesses of AI, along with the need to identify tasks that should remain exclusively human, termed 'just me' tasks.

The author employs the cyborg and centaur concepts, where humans and AI work together seamlessly, integrating AI into workflows rather than merely delegating tasks. This approach allows for the most effective use of AI, with humans overseeing and refining AI outputs. Mollick highlights the potential for organisational policies to hinder AI adoption, as companies may fear job losses or legal issues. While AI has the potential to transform industries and job roles, Mollick acknowledges that the future remains uncertain. No one can predict exactly how AI will affect specific jobs or organisations, but the key lies in adapting to the changes, leveraging AI's strengths, and ensuring that human roles continue to focus on creativity, critical thinking, and uniquely human tasks.

In Chapter 7, Mollick reminds us that the potential to revolutionise education has long been recognised, yet consistently achieving such transformation remains challenging. In 1984, educational psychologist Benjamin Bloom introduced the "two sigma problem", revealing that students who received one-on-one tutoring performed significantly better than those in traditional classrooms. This led to a quest for group instructional methods that could replicate the benefits of tutoring, though none have consistently matched its effectiveness. The unique interaction between tutor and student is difficult to replicate, making personalised tutoring the holy grail of education.

The advent of AI has further complicated the educational landscape. Traditional classroom practices, where students are taught by a teacher and complete homework, have been disrupted by AI, leading to what some call the 'homework apocalypse'. AI's capabilities, especially in generating essays, have made cheating easier and more difficult to detect. This parallels the historical reluctance to adopt calculators in education, which eventually became integral to the curriculum.

AI's impact on education extends beyond cheating. It holds the potential to enhance teaching methods, particularly through tools like AI tutors, which provide personalised assistance. The flipped classroom model, where students learn new concepts at home and apply them in class, is one approach that could benefit from AI's capabilities. However, the success of such models depends on high-quality resources, which are often lacking.

While previous technological solutions have fallen short, AI offers a unique opportunity to transform education. It empowers educators and enhances learning experiences, potentially achieving the two-sigma improvement which Bloom envisioned. However, the challenge remains to steer this shift in a way that expands opportunities for all and nurtures human potential. The future of education may hinge on how effectively AI is integrated into teaching and learning processes.

Chapter 8 is titled "AI as a coach". The greatest threat AI poses to our educational system is its potential to undermine the informal apprenticeship system that follows formal education. This system, though not officially part of education, is crucial in shaping real-world skills as workers enter the workforce. For instance, in medicine, surgical robots have been used for over a decade, but their presence is already disrupting the training of new surgeons. As AI automates more tasks, this training crisis could spread, endangering the pipeline that creates experts.

Despite AI's ability to process vast amounts of information and perform complex tasks, expertise remains crucial, especially for overseeing and correcting AI's work. Stanford research has shown that while AI like GPT-4 can outperform medical students on exams, true expertise requires more than just factual knowledge. It involves critical thinking, problem-solving, and the ability to evaluate AI outputs—skills developed through deliberate practice. This practice, which includes continuous challenges and feedback from mentors, is essential for mastering any field.

AI can assist in this process by providing instant feedback and guidance, but it cannot fully replace human mentorship and the nuanced understanding that comes from experience. Expertise also involves talent, and not everyone can become an expert, no matter how much they practise. However, AI has the potential to complement human expertise, helping workers focus on developing specialised skills while AI fills in gaps in knowledge. In the future, AI will likely work alongside human experts rather than replace them, enabling a collaboration that enhances both human and machine capabilities. As AI evolves, it remains vital to continue developing our own expertise to guide and oversee AI's contributions effectively.

The book's final chapter, "AI as our future", discusses the rapid advancements in AI and its profound implications on society. AI, while not sentient, mimics human-like behaviour with remarkable accuracy, having been trained on vast human knowledge and the labour of low-paid workers. It has demonstrated the ability to pass tests and perform creative tasks, fundamentally altering how work and learning are approached. However, concerns arise that AI's growth may soon plateau, with advancements becoming less significant compared to the leaps seen from GPT-3.5 to GPT-4o.

The integration of AI into various sectors has led to significant changes, some of which were unanticipated. For instance, AI's ability to complete tasks traditionally reserved for humans has raised concerns among educators, employers, and government officials. As AI-generated content becomes (near-)indistinguishable from real images, videos, and voices, the online information environment is expected to become unmanageable, overwhelming fact-checkers and making technological solutions like watermarking ineffective.

AI is also changing personal interactions, with systems becoming more engaging and, in some cases, preferred over human interaction. While AI is expected to complement human work by handling tedious tasks, it may also displace certain jobs, such as translation. Although AI's exponential growth might slow due to factors like increasing training costs and regulatory requirements, the gradual pace of AI's development provides society with an opportunity to adjust. However, the potential misuse of AI by criminals, terrorists, and governments for surveillance and control remains a concern. Ultimately, Mollick's *Co-intelligence* emphasises the need for careful planning and discussions about AI's role in society to ensure that its advancements lead to positive outcomes rather than catastrophic consequences.

An optimistic vision for collaboration – pros and cons

The book's optimistic outlook on the future of co-intelligence is one of its most notable features. In a utopian society, Mollick envisages AI and humans co-existing peacefully to solve problems together and explore new frontiers in creativity and innovation. The book provides practical thoughts for incorporating AI into different aspects of society while maintaining the essence of what makes humans unique by drawing on real-world experiences. Table 1 presents a brief overview of the pros and cons of the book.

Table 1: Pros and cons.

Pros
1. Insightful exploration of co-intelligence: Mollick sheds light on the possibility of co-intelligence by making a strong case for seeing AI as a collaborative partner.
2. Practical examples and case studies: The book is full of examples from everyday life that show how AI is employed and what it can do in the future.
3. Actionable insights for implementation: Mollick offers techniques for incorporating AI into processes and learning environments in a smooth and efficient manner.
4. Optimistic vision: The author's upbeat perspective stimulates ideas of advancement and creativity via AI-human cooperation.
Cons
1. Limited exploration of ethical concerns: Although ethical considerations are discussed, a more thorough analysis would have been called for.
2. Lack of detailed critique on pitfalls: A more critical examination of issues such as job displacement could have yielded a more complete picture.

While Mollick's enthusiasm for co-intelligence comes through, the book insufficiently addresses the ethical issues surrounding this developing terrain. While "Co-intelligence" encourages readers to consider potential biases in AI algorithms and expresses concern about societal consequences, a more in-depth examination of these ethical quandaries could have enriched the discussion. Readers looking for a more critical perspective on AI could look at Lindgren's (2023) gargantuan handbook, recently reviewed in JALT (Rudolph, 2024).

Practical applications and real-world examples

"Co-intelligence" shines due to its plethora of real-world examples that highlight the application of AI in a variety of fields. By fusing examples from business, education, and other domains, Mollick provides readers with a road map for implementing AI in a variety of contexts. Whether it is optimising decision-making procedures or transforming educational approaches, the book is a valuable resource for understanding the current AI environment.

Here are some of my own thoughts on how businesses and educational institutions might utilise AI for a competitive edge based on the book's concepts and recommendations. The first two examples are an electrical vehicle manufacturer and an airline, and the last three examples are higher education institutions. BYD (a Chinese electric vehicle manufacturer) might potentially employ co-intelligence designs to enhance collaboration between human developers and AI systems. This partnership could result in the development of more intuitive and user-friendly AI solutions. Also, BYD

might build intelligent productivity tools to assist staff in duties such as scheduling, task prioritisation, and project management. AI endeavours in automotive technology could also benefit from co-intelligence concepts, building AI-driven systems to boost vehicle safety, efficiency, and user experience (Bostrom, 2014).

Singapore Airlines might employ co-intelligence concepts to construct collaborative AI systems that work alongside aviation professionals. This could result in improved flight operations, passenger service, and safety measures. By integrating AI as a co-worker in data analysis, Singapore Airlines could strengthen its business strategy, giving insights into customer preferences, market trends, and operational efficiencies. Furthermore, AI technologies could improve sustainability and customer experience initiatives by evaluating data to optimise fuel use, reduce carbon footprint, and personalise passenger services.

Researchers at the National University of Singapore (NUS) could leverage co-intelligence in AI-driven medical imaging, patient data analysis, and drug discovery, potentially driving significant healthcare innovations. NUS's robotics labs could also benefit by developing human-collaborative robots to enhance efficiency in manufacturing, healthcare, and exploration. Additionally, applying co-intelligence to societal challenges like poverty reduction, disaster response, and public health could be a priority (see Ford, 2015). Harvard University could integrate co-intelligence concepts across various research disciplines, notably in finance and economics, where AI-driven models could enhance financial analysis, risk assessment, and market forecasting. Harvard Law School might apply AI-human collaboration to legal research and policy analysis, improving data analysis and predictive capabilities. The Berkman Klein Center for Internet & Society could focus on the ethical dimensions of AI, working to mitigate algorithmic biases and ensure responsible deployment, aligning with the ethical imperatives outlined by Brookfield et al. (2024). Finally, Kaplan Higher Education Academy's engineering courses could incorporate co-intelligence for AI-driven design and innovation, helping engineers optimise designs, simulations, and prototyping. In business analytics, AI could enhance strategic decision-making and predictive modelling, particularly in business gamification and experiential learning platforms (Lim, 2023).

Final thoughts

Mollick's book, *Co-intelligence: Living and Working with AI*, is a timely introduction to the impact of AI on daily life, particularly for AI novices and those curious about LLMs. He addresses AI limitations, best practices for creative projects, and the future of human-AI collaboration. However, seasoned practitioners might find fewer new insights as the book focuses on practical guidance and near-term effects.

Mollick examines AI's tendency for 'hallucinations', where AI generates non-existent references, highlighting the risks of treating AI-generated information as accurate. He advocates for careful consideration when using AI for creativity and innovation. With an engaging style and concise analysis, Mollick offers a compelling vision of the collaborative future between humans and AI. The book appeals to a wide audience, especially those new to AI, by providing a practical yet optimistic outlook on AI's potential.

Mollick also explores the complexities of integrating AI into various fields, emphasising a future where AI and human creativity drive innovation together. This perspective is especially relevant in our rapidly evolving technological landscape. Through insightful analysis, Co-intelligence encourages readers to embrace a future where AI and human ingenuity advance society. Whether you are an AI enthusiast, a newcomer, or a specialist, Mollick's work offers a valuable examination of the evolving human-AI relationship. While a deeper exploration of ethical considerations and a more critical analysis of AI's risks would have added value, Co-intelligence excels in its narrative, practical insights, and optimistic vision of AI-human partnership.

Additional references

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Brookfield, S. D., Rudolph, J., & Tan, S. (2024). *Teaching well: Understanding key dynamics of learning-centre classrooms*. Routledge.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.

Ford, M. (2015). *Rise of the robots: Technology and the threat of a jobless future*. Basic Books.

Lim, C. (2023, September 23). *Learning and teaching transformation in 2023* [Presentation]. Murdoch-Kaplan Learning and Teaching Symposium. Murdoch-Kaplan, Singapore.

Lindgren, S. (Ed.). (2023). *Handbook of critical studies on artificial intelligence*. Edward Elgar.

Rudolph, J. (2024). Book review of Lindgren, S. (Ed., 2023). Handbook of critical studies on artificial intelligence. *Journal of Applied Learning & Teaching, 7*(1), 429-435. https://doi.org/10.37074/jalt.2024.7.1.30

Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. Knopf.