

Of Machines and Men: Judgment and Thinking in the Age of Artificial Intelligence

Kar-yen Leong

Associate Professor, PhD Program in Cognitive Sciences,
National Chung Cheng University Taiwan

Abstract

In this short reflective piece, I reflect on my experiences co-teaching a class on artificial intelligence and the issues that were raised. Through several assigned readings, both the students and the instructors were able to reflect on the broader challenges related to AI. What came to the fore was a deeper discussion on the ethics of AI and the notion of responsibility.

Keywords

AI, unmanned aerial vehicles, violence, war, human rights, ethics, judgment

Introduction

As they shuffled into class, there was a discernible cloud in their eyes, a lingering effect of a heavy afternoon lunch. *It will be difficult getting through to them*, I thought. This was a class composed of students from various faculties, all at different stages of their academic careers. The class had begun as an experiment, prompted by my colleague's research interest in artificial intelligence within the realm of public administration. On my part, AI was a relatively new topic, and teaching it through a language I was not entirely familiar with seemed, to me, to invite disaster. Nevertheless, as the semester progressed, I sensed that our weekly three-hour meetings had taken on a natural momentum, leading myself, my colleague, and our students toward

novel ways of thinking about the underlying issues surrounding AI.

The decision to co-teach this course was also a way to learn from a senior colleague, both about the practical and intellectual investments required by this form of technology. I have grown skeptical of the frenzied rush many universities exhibited in embracing all things AI. Institutes of higher education appear eager to attach the letters “AI” to nearly every one of their programs, in a conspicuous effort to attract more students. What is painfully lacking, however, is a measured approach to questioning AI’s impact on society, as well as its implications for my own area of research: human rights.

Thus, amid an alphabet soup of LLMs, GANs, algorithms, and neural networks, my colleague and I, in a generically titled course called “Selected Readings on AI”, chose an interdisciplinary approach, drawing on multiple perspectives. After all, we had a cohort of students pursuing degrees in engineering, communications, labor studies, law, as well as philosophy. The primary text was Ben Buchanan and Andrew Imbrie’s *The New Fire: War, Peace, and Democracy in the Age of AI*, accompanied each week by a selection of related articles.¹ The students were expected to respond to the chapters by formulating questions, and as a group, we would work collaboratively to answer those questions. There were instances when language barriers almost derailed my attempts to engage with students, but fortunately, there was almost nothing Google Translate could not handle. We, as instructors, were also always ready to provide deeper explanations or guidance concerning unfamiliar terms and concepts. Yet, in some instances, the students proved savvier than we were.

The course was intentionally designed to mimic the dynamics of a “book club”, placing less emphasis on hierarchy and order, thereby giving students ample space to act, react, and provide their individual interpretations of the text. For me, it was an opportunity to embark on a foray into unknown territory, making me as much of a neophyte as the other participants in the class. While written for generalists, the text offered some technical insights into the workings of artificial intelligence. As we progressed through the book, it became clear that the authors aimed to issue a warning about the paths this ever-evolving technology might take.

1 See Ben Buchanan and Andrew Imbrie. 2022. *The New Fire: War, Peace, and Democracy in the Age of AI*. Cambridge: MIT Press.

According to Buchanan and Imbrie, AI has its evangelists, Cassandras, and warriors. The evangelists see nothing but AI's potential to drive humanity toward a bright, prosperous future. The Cassandras, on the other hand, are skeptics whose warnings are rarely heeded. The final category, the warriors, believe technology primarily serves as a weapon to wield against adversaries. The title of the book itself reveals the authors' intention: to persuade their readers that AI and its companion technologies represent a "new fire". This metaphor is vividly illustrated in a chapter describing the wonder and terror experienced by J. Robert Oppenheimer, the father of the American atomic bomb, when he first witnessed his creation.² Like atomic energy, AI comes bearing both promise and a fear of the unknown. Indeed, this new form of technology consistently raises profound conundrums regarding its role in humanity's progress.

Perceptions

At the beginning, we needed to develop an understanding of the perceptions some of our students might have about AI. Through our discussions, it became apparent that, at least for some students from non-technical fields (including myself), their view of AI was influenced by popular media. Many references were made to movies featuring characters like Arnold Schwarzenegger portraying relentless, bloodthirsty robots programmed to kill humans. Another common reference was *The Matrix*, where, in a fictional future, human beings serve merely as enslaved batteries providing energy to their machine masters.

For most people, algorithms, which are the language and equations required to power AI, are difficult to comprehend. This lack of understanding, or rather the inability to grasp the complexity of the "algorithmic wall", causes much of the public to perceive AI as possessing limitless capabilities, almost as if it were "magic". This misconception often leads the entertainment industry to create images of terrifying robots intent on destroying humanity.

However, as we delve deeper into the technical aspects of AI, we also become more aware of the human foundations underlying the technology. When examining the core of AI, specifically its so-called neural network, the

2 See Kai Bird and Martin J. Sherwin. 2006. *American Prometheus: The Triumph and Tragedy of J. Robert Oppenheimer*. New York: Vintage Books.

class gradually began to understand that engineers are essentially attempting to replicate the way the human brain functions. It is noteworthy that the “neural networks” that enable AI’s computational sophistication depend on the number of layers that the chipset can process. Data availability lies at the core of this process, and the system’s efficiency relies on both the capacity to store information and the speed at which it processes data.

As we continued to explore the technical dimensions, we encountered another AI platform known as generative adversarial networks (GANs). In this framework, the processor’s power enables two opposing sets of arguments to challenge each other. Through the mistakes made and the subsequent corrections that follow, the process evolves into what is commonly known as “machine learning”. The machine becomes trained to think dialectically, continually improving its output.

One of the most intriguing aspects of this technology is its ability to generate images that can deceive even the most discerning human observers. There are now websites displaying human faces that do not exist, as well as similar sites featuring images of houses that have never been built. This generative capability of AI has even extended to mimicking human voices, making it increasingly challenging to distinguish truth from fabrication.

Real/Unreal

This development has pushed us to the brink of what we can consider human creativity, while also presenting a deeply unsettling conundrum: the ethical considerations brought about by such technology. What is real? What is unreal? While these questions may be more familiar to philosophers, for the many who are not, AI amplifies the sense of confusion that people experience. The power to create images, sounds, faces, and voices is now easily accessible, as long as there is a sufficiently large dataset available to entrepreneurs.

The book highlights a pioneering group of computer engineers and data scientists driven by the ambition to create an autonomous artificial being capable of recognizing elements in its environment. This form of autonomy would mark the initial steps toward developing technology that can think independently. Achieving this, however, requires vast amounts of data.

In the early days of artificial intelligence during the 2000s, data

scientists, primarily from West Coast American universities, sought ways to collect as many datasets as possible to improve their models. Scientists like Li Fei-Fei were among the pioneers pushing the boundaries as they sought access to large volumes of images to feed into algorithms, enabling systems to recognize objects. This process required a substantial collection of images and human labor to differentiate one object from another.

Over time, as processors have become faster, AI systems have developed an increasing demand for images that they can process independently. The only resources capable of satisfying this demand are images of humans and their personal information. Social media platforms such as Facebook and Instagram have become rich sources of unimaginable amounts of data, turning end users into consumers at every step.

The book introduces the concept of “surveillance capitalism”, where personal information is not only used against individuals but also sold for profit to anyone with the means and capacity.³ One clear example of this is the case of Cambridge Analytica, an obscure data company that purchased information from Facebook to influence American voters during an election campaign.⁴

It is clear that the authors themselves are extremely cautious about the possibilities of AI. As the book has been recently published, the authors are acutely aware of the growing geopolitical tensions between the two largest economies in the world. There are parallels to be drawn between the present day and the Cold War, when the Soviets competed with the West over who possessed the most advanced technology. This form of détente now characterizes the relationship between China and the United States, as they race to develop increasingly powerful platforms driven by processors and AI.

Evidence of this competition is apparent in the intense rivalry between tech companies in both countries and in the efforts by each nation to undermine the other’s capacity to develop these technologies. The underlying fear fueling this paranoia is the perception that data fed into AI platforms developed either by the US military’s Defense Advanced Research Projects Agency (DARPA) or the People’s Liberation Army could be used to enhance

3 See Shoshana Zuboff. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.

4 For more on this see Christopher Wylie. 2020. *Mindf*ck: Inside Cambridge Analytica’s Plot to Break the World*. London: Profile Books.

weaponry that each side could deploy against the other. The possibility that these platforms might not only be used for espionage but could also develop the capacity to physically impact the real world, influencing life or death decisions, is very real.

One of the chapters in the book, aptly titled “Killing”, brings this issue to the forefront. What struck me was the use of the term “lethal autonomous systems” and its intention to make the act of killing more “efficient”. While there is a persistent fear that these war machines could turn against humanity, it is even more concerning to consider how these machines might actually achieve autonomy. We are reminded of Skynet, from the Terminator series, launching nuclear missiles at human cities and destroying civilization as we know it. However, the movie does not explain how we came to place such faith in these machines in the first place.

Killing from afar

This question weighed heavily on me. The books made me reflect on the extensive use of unmanned aerial vehicles (UAVs) in warfare. If, at some point, these machines were to develop the capacity to identify not only inanimate objects but also living targets, who would be held accountable if a UAV accidentally killed an innocent child? Fearing that this topic might complicate our discussion of the book’s contents, I decided to introduce a related issue that, while not directly covered in the text, I believed to be relevant.

Journalist Mark O’Connell spent time trying to understand the many tech billionaires living within the Silicon Valley bubble who aspire to live “forever”.⁵ Their attempts to cheat death and their obsession with cryogenics provide fascinating insights into how they aim to “perfect” the human body. If we imagine the human body as a machine, even though it is an organic one, it still cannot escape the ravages of time. Therefore, if the human body could be augmented, as many of these “transhumanists” claim, it might eventually overcome humanity’s ultimate imperfection: death. However, since we have

5 See Mark O’Connell. 2017. *To Be a Machine: Adventures Among Cyborgs, Utopians, Hackers, and the Futurists Solving the Modest Problem of Death*. New York: Granta Books. Curious readers can avail themselves of more tomes on transhumanism, the topic of his book with this short article written for the Guardian newspaper see <https://www.theguardian.com/books/2018/may/10/mark-oconnell-five-books-to-understand-transhumanism>.

not yet achieved that goal, we continue to be vulnerable to sickness, pain, and mortality.

This is where machines come into play. Examining the transhumanist influence on AI development, it could be argued that this technology represents an effort to surpass human capacity for retaining and processing information, and perhaps even for making decisions. In theory, this could make us more precise and less prone to errors, reducing both our physical and moral burdens. In a world fractured by wars and conflicts, AI is becoming increasingly necessary as nation-states and governments seek to defeat their adversaries in the most efficient way possible. The automation of violence now drives this pursuit, as any delay in a soldier or combatant making the final decision can be a significant disadvantage. Consequently, AI has been tasked with making decisions as accurately and swiftly as possible. Since human qualities are perceived as a hindrance in this context, every human element must be minimized to the greatest extent feasible.

I brought this perspective into the classroom by using examples from the current Israeli offensive against the Palestinians. In a series of exposés published in 2023 by the independent news portal *+972 Magazine*, journalist Yuval Abraham revealed the inner workings of a unit within the Israeli military that develops automated systems used to target the Palestinian civilian population.⁶ This process involves treating the West Bank and the Gaza Strip as a “laboratory” where the Israeli Defense Forces (IDF) fine-tune their weaponry.

Since its inception, Israel has established a substantial military-industrial complex, selling weapons to the highest bidder. AI and related technologies have accelerated Israel’s ability to do this, further reinforcing its strategy of containing the Palestinian territories. Anthony Loewenstein highlights this by asserting that Palestinians live within a “panopticon”, constantly monitored by advanced technology that “listens” to their lives while their metadata is processed by algorithms to monitor the behavior of the “inmates”.⁷

The population is then sorted into categories, and for those deemed

6 See Yuval Abraham’s. 2023. “‘A mass assassination factory’: Inside Israel’s calculated bombing of Gaza.” *+972 Magazine* 30 November 2023. <https://www.972mag.com/mass-assassination-factory-israel-calculated-bombing-gaza/>.

7 See Antony Loewenstein. 2023. *The Palestine Laboratory: How Israel Exports the Technology of Occupation Around the World*. London: Verso.

dangerous, their fate is sealed at the tip of an incendiary device. Guided by an AI system known as Pegasus, Yuval revealed that Israel's "mass assassination factory" tracks targets, and when they reach a certain proximity, a bomb is dropped on that person. However, there is collateral damage, with allegations suggesting that the devices are intended to kill not only the suspected terrorist but also their family members. All of this occurs through the precision of AI.

The human, as the final line of defense, whose thumb rests on the release button, is reduced to a mere formality in the decision-making process. Instead, responsibility is handed over to a cold, emotionless algorithm that is perceived as more efficient and reliable than its human operator. According to +972, the October 2023 attacks and the subsequent kidnapping of Israeli citizens, as well as non-Israelis, have reinforced the resolve of this small Middle Eastern country to eliminate forces it perceives as hostile. Since Hamas is deeply embedded within Palestinian society, technology has become the primary tool for identifying Hamas operatives living among the Palestinian population.

Since Hamas is largely the only viable entity providing state services, this makes any Palestinian a potential enemy. By enhancing surveillance technology, the IDF has been able to identify possible threats through messaging and communication applications. This increased "lethality" allows the system to assess the likelihood of each monitored individual becoming an "enemy". However, since all Palestinians are viewed as potential enemies, the threshold becomes blurred, leading to inaccuracies. An enemy could just as easily be a policeman, a civil servant, or even an ordinary citizen making use of hospital services provided by the Hamas administration.

The question then remains: who ultimately takes responsibility for the lives and deaths of these "potential" targets or enemies? Does accountability rest with the individual military personnel or with the AI, the ghost in the machine?

Thinking in Dark Times

The class discussion often failed to produce concrete answers, but my colleague quickly suggested adopting the works of a thinker whom countless political scientists have relied on for guidance during challenging times. Hannah Arendt's ideas are so expansive that they provide her followers with a steady framework to navigate the complexities of the human condition, even

in the 21st century.

Arendt warned that modernity, equipped with technology designed for widespread and efficient destruction, not only transforms the conduct of warfare but also alters humanity's collective sense of morality. Her concept of the "banality of evil" is particularly insightful. Based on her observations of Nazi functionary Adolf Eichmann⁸ and her teacher Martin Heidegger,⁹ Hannah Arendt exposes the consequences of "unthinking" when excessive faith is placed in systems rather than in the capacity for thought and conscience. Unthinking, however, does not imply the cessation of thought but rather something akin to "thoughtlessness". Arendt's portrayal of Eichmann presents him as "bland" and "banal" as he dutifully fulfilled his administrative role, efficiently sending Jews to extermination camps. Eichmann's numerous victims remained nameless and were kept at a distance from his gaze.

Similarly, AI offers an even greater promise of efficiency and speed, with the potential to replace the human decision-maker or, at the very least, exert such overwhelming influence that humans become dependent on it for decisions involving life and death. Consequently, the operator, the human who ultimately pulls the trigger or activates an advanced weapon system, becomes just another Eichmann-like cog, waiting for confirmation from the algorithm and losing fundamental human attributes.

As researcher Elke Schwarz explains, these automated killing systems "...challenge the sensory and evaluative authority of the human in the loop" (Schwarz, 2018: 288). She further notes that the system, comprising both hardware and software, offers its human operators "superhuman" abilities. Taking unmanned aerial vehicles as an example, these killing machines excel in endurance, as drones do not blink or suffer from pilot fatigue. They also surpass human capabilities in data collection and analysis, processing vast amounts of data. This combination, paired with the apparent capacity for greater precision, turns the drone into more than just an instrument, making it a sanitized guide in the practice of killing (Schwarz, 2018: 288).

For Schwarz, the dilemma posed by autonomous systems is not just that

8 See Hannah Arendt. 2006. *Eichmann in Jerusalem: A Report on the Banality of Evil*. New York: Penguin.

9 See Hannah Arendt. 1994. "Heidegger the Fox." *Essays in Understanding, 1930-1954: Formation, Exile, And Totalitarianism* 361-366. ed. Jerome Kohn. New York: Harcourt Brace.

they enhance our capacity to kill while eroding our moral sense but that they achieve this by gathering and processing vast amounts of data, transforming them into “meta-data”. It is this decontextualized meta-data that determines who lives or dies. The individual is of no consequence unless their behavior aligns with what an algorithm categorizes as “combatant” or “non-combatant”. Schwarz states, “...algorithmic data analysis programs...serve as a justification for the legitimate killing of persons who may have done nothing to warrant lethal harm” (Schwarz, 2018: 288). In essence, these systems have assumed the roles of judge, jury, and executioner.

What role does a discussion like this play in our understanding of AI? Would such conversations ultimately influence the rapid pace of technological advancement? Would technologists, programmers, computer scientists, and engineers even pay attention? Most students in the class were able to grasp the essence of the discussion, asking whether there were “rules of engagement” established to better manage the more lethal aspects of AI. We reflected on the applicability of Isaac Asimov’s laws of robotics, but these considerations did not resolve the dilemma we faced. Simply formulating laws to prevent technology from harming humanity was insufficient.

Powerful adversaries such as China and the United States have signed agreements pledging not to use AI for lethal purposes. However, these codes of conduct have not prevented Israel, a close US ally, from developing technologically advanced weaponry. During the seminar, I began to consider the role of the university in this context. Many higher education institutions are already emphasizing AI, promoting greater volumes of research, development, and application. This trend is evident given the governmental push to provide financial incentives through public funding. Cash-strapped universities are taking advantage of these opportunities, accessing substantial state resources as governments move toward “sovereign” AI, weaponizing technological advancements in the name of “national interests”.

In today’s global landscape, we seem to be experiencing a new kind of Cold War, both different and similar. As individual states pursue AI advancements, and as the international environment becomes increasingly multipolar, it is expected that relations between states will become more strained, with AI emerging as a critical asset. Universities and research institutes will likely become more deeply integrated within this ecosystem. At this critical juncture, universities should not simply aim to profit from the

influx of research funding aimed at developing advanced AI systems. Instead, they should reaffirm their role as places of conscience by encouraging society to rethink its approach to this new technology. How can this be achieved?

To think

Perhaps the simplest thing to say right now is to encourage people “to think”. It really is that straightforward. Exploring Arendtian thought, Paul Formosa (2010) differentiates the types of thinking available to discerning students. The emphasis is placed on what she calls a “critical thinker”. While there are “professional thinkers”, such as her former professor Martin Heidegger, who engage in thinking as part of their vocation, these “authority figures” are often perceived as being too enamored with the sound of their own thoughts. What is needed are thinkers who actively engage with others, allowing their logic to be either improved or disproved. Formosa describes this approach as one that “...exposes itself to the test of free and open examination, and this means that the more people participate in it, the better... other citizens are needed as a necessary touchstone, though not a replacement for one’s own thinking” (Formosa, 2010: 91-92).

Critical thinking is, of course, only one aspect of the entire process. It is an essential function of any university that values its role in developing intellectual capacity. Cultivating this faculty is of the utmost importance, given the modern tendency to relinquish decision-making to mechanized forms of “thoughtlessness” embodied in technological systems. The soldier is not devoid of thought; rather, his confidence in his own reasoning has been overtaken by his faith in the machine. What the modern warrior, operating the UAV guided by “smart” technology, has ultimately lost is his sense of judgment. This loss of judgment marks the downfall of Eichmann.¹⁰

Through ideologies and what can be described as the automation of thought, humanity has lost its ability to think critically and, more importantly, to exercise independent judgment. It is crucial to instill in students the understanding that judgment, thought, and critical thinking are inherently interconnected. Formosa argues that “...critical thinking

10 Arendt explores this idea extensively in her magnum opus, *The Human Condition*. To put it simply, modernity reflects a condition where ideologies dominate, and the human faculty becomes nothing more than an obstacle to profit, efficiency, and management. It represents the ultimate capitulation of humanity’s intellectual and moral capacities, leaving them in the hands of “systems”.

has a further conditioning effect as it sets the foundation for (but does not guarantee) representative judgment... by acclimatizing the thinker to regularly considering their opinions from a plurality, which leads to both impartiality and an acute awareness of the nuances of difference” (Formosa, 2010: 93).

This plurality stems from an Arendtian engagement with the world, characterized by “natality”, where individuals act in harmony with others to create a world that is dynamic and constantly evolving through interaction.

Moving forward

As we approached the end of the semester, I hoped that the students were able to appreciate the final point I raised. From what I could gather, they seemed to enjoy the banter and the lively conversations that emerged from our individual readings of the chapters and articles. What became clear towards the end of the class was that, despite being perceived as a “professional” thinker, I was still able to encourage them to consider the possibility of becoming critical thinkers themselves.

The students, despite their heavy lunches, busy schedules, and assignments outside of class, gradually came to understand the importance of the question I posed at the beginning of this reflection. Who will ultimately take responsibility, and who must bear that burden? How do we prevent it from committing harm? As we explored these questions further, the class began to realize that the issue was not solely with AI but with ourselves.

During my years of teaching in Taiwan, I have often sensed a palpable feeling of vulnerability, not only due to the threat of physical aggression from China but also because of the fear that Taiwan’s open system could be exploited and overwhelmed by misinformation and half-truths. The power of AI and false messaging lies in its ability to create divisions within society. Therefore, we emphasized in class the importance of being especially vigilant about the “fake news” that surrounds us. Another issue raised in class touched on the dangers of ‘confirmation bias’ where we only seek out information which suits our own world view. This will inadvertently lead towards the inability to engage with those whose perspectives are different from ours leading to ‘tunnel vision’.

As the semester concluded, it felt as though there were still many unresolved issues. Nevertheless, what mattered most to both me and my

colleague was that we had not only succeeded in fostering a sense of responsibility but had also encouraged a deeper critical engagement with the world, two important elements if our democracy is to thrive and survive.

References

- Formosa, Paul. 2010. "Thinking, Conscience and Acting in Times of Crises." *Power, Judgment and Political Evil: In Conversation with Hannah Arendt* 89-106. Burlington: Ashgate.
- Schwarz, Elke. 2018. "Technology and moral vacuums in just war theorizing." *Journal of International Political Theory* 14, 3: 280-298.

Further Readings

- Gonzalez, Roberto J. 2024. *War Virtually: The Quest to Automate Conflict, Militarize Data, and Predict the Future*. Berkeley: UC Press.
- Schwarz, Elke. 2018. *Death machines: The ethics of violent technologies*. Manchester: Manchester University Press.

論機器與人：人工智慧時代之判斷與思考

梁家恩

台灣國立中正大學認知科學博士學位學程副教授

摘要

在此簡短卻饒富反思的文章中，我回顧了自身協同教學一門人工智慧課程的經驗，及其衍生的若干議題。透由數篇指定閱讀文本的引導，學生及授課者皆得以省思人工智慧所帶來更加廣泛的挑戰，尤其發人深省的是針對人工智慧倫理及責任概念深入探討之必要性。

關鍵字

人工智慧、無人機、暴力、戰爭、人權、倫理、判斷
