



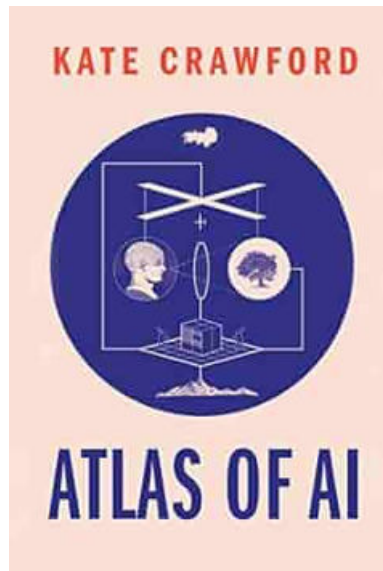
AI Isn't Neutral: How AI Reinforces Bias

By Med Kharbach, PhD

Introduction

I have just finished reading this excellent book by Kate Crawford titled ***Atlas of AI: Power, Politics, and The Planetary Costs of Artificial Intelligence*** and thought to share with you some of my thoughts about the concept of data and its relation the problem of bias in AI systems.

All AI systems are powered by data, it's their foundation and lifeblood. Kate Crawford captures this perfectly with her poignant metaphor: data as the new oil.



To truly grasp the significance of data in today's AI revolution, let's take a step back and explore the historical roots of this transformative technology.

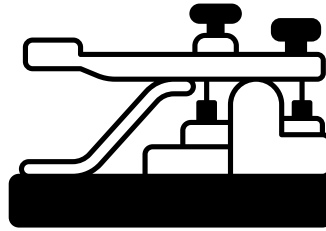
The first seeds of AI were planted by Alan Turing's (1950) seminal paper **Computing Machinery and Intelligence**. Subsequent computer scientists worked and further developed Turing's ideas and in 1956, John McCarthy of MIT invented the nomenclature Artificial Intelligence (Mollick, 2023).



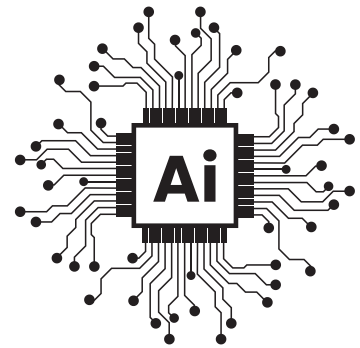
The Internet



Steam



The Telegraph



Source: Canva

AI is a general purpose technology and one common thing that characterizes these technologies is their slow adoption. Let's take the example of the Internet. The Internet was first born in a project called ARPANET in the 1960s and, as Mollick (2023) stated, it took 3 decades for this new invention to take hold in society. Why? Because these general purpose technologies usually require the development of multiple other technologies for them to work well.

The Internet, as a general-purpose technology, required the advent of more powerful and affordable computers in the late 1980s, coupled with the development of web browsers in the early 1990s, to truly gain momentum and revolutionize the way we connect and interact.

The same is true for AI. Although it first emerged in the 1950s, it took decades for the term to become a household name. Like the Internet, AI relies on a robust infrastructure, including high-speed internet, affordable devices with powerful computational capabilities, and, most critically, vast amounts of data to train its models effectively.

Large Language Models (LLMs) like ChatGPT require vast amounts of data to train—billions of words, images, and other content. Such data simply wasn't available before the Internet era. Early training datasets, like the Brown Corpus, WordNet, and ImageNet, were limited in scope and scale.



Source: canva.com

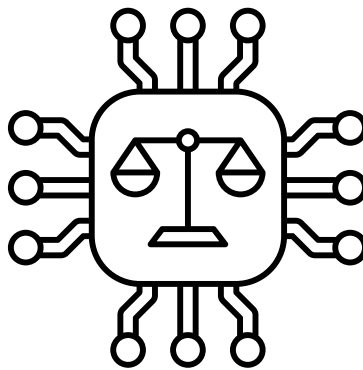
This scarcity of accessible data was one of the key reasons AI experienced the so-called "AI Winter," a period when the field stagnated as other technologies gained prominence. However, the widespread adoption of the Internet transformed the landscape, offering AI researchers an unprecedented wealth of data and computational resources. This shift enabled modern advancements in AI that earlier generations of scientists could only imagine.

For the first time in human history, immense volumes of raw data—text, images, audio, and video—are aggregated in a single, interconnected space: the Internet. This vast repository has become a goldmine for data miners and researchers offering unprecedented access to material for analysis and training AI models.

Every piece of content shared online, from a casual Facebook comment to an academic paper on a research journal, becomes part of this massive data ecosystem, stored in data centers across the globe, waiting to be harnessed.

The sheer scale of data available today is staggering. As Kate Crawford illustrates, "on an average day in 2019, approximately 350 million photographs were uploaded to Facebook and 500 million tweets were sent. And that's just two platforms based in the United States."(p. 106)

This example offers just a glimpse into the massive data farms generated daily across the Internet. Nearly the entire repository of human knowledge what could be considered humanity's epistemic archive is now stored online. As Crawford aptly notes, "anything and everything online was primed to become a training set for AI" (p. 106), shaping the way AI systems learn and operate in the digital age.



An illustration of the concept AI bias (Canva)

And this is where **the story of AI bias begins**, a story rooted in the vast yet imperfect datasets fuelling modern AI systems. Below, I've summarized this narrative into key points each paired with powerful quotes from Kate Crawford's **Atlas of AI**.

Ways AI Reinforces Bias

In this list, I break down some of Kate Crawford's most compelling arguments about how AI systems are shaped by social, historical, and political contexts—and the consequences this has for education, technology, and society as a whole. Let's dive in!

1. Data is Never Neutral

Kate Crawford lays it out plainly: the idea that data is neutral is a myth. She explains that machine learning systems are trained on images and data scraped from the internet or state institutions without any regard for the context or consent of the individuals involved. In her words, "they are anything but neutral."

Crawford gives the example of images taken from police stations and used in AI training. She highlights how AI systems use these images without asking why the person was in the station or what their story might have been. She writes:

“

A computer vision system can detect a face or a building but not why a person was inside a police station or any of the social and historical context surrounding that moment. Ultimately, the specific instances of data—a picture of a face, for example—aren't considered to matter for training an AI model. All that matters is a sufficiently varied aggregate. Any individual image could easily be substituted for another and the system would work the same. According to this worldview, there is always more data to capture from the constantly growing and globally distributed treasure chest of the internet and social media platforms. (p. 94)

”

What's troubling is how this context gets stripped away when the data enters the pipeline for AI training. These systems aren't designed to consider the social and historical weight of their training data. They reduce complex human realities into abstract numbers, treating deeply significant moments—like being photographed in a police station—as just another data point.

And this has real consequences. AI systems trained on such data carry those biases forward. They may reflect or even amplify systemic inequalities tied to policing or incarceration, simply because they were never designed to ask the right questions: Why was this image taken? Who is in it? What broader systems are at play here?



Illustration of the impact of AI on our planet. I generated this image using Midjourney

2. The Illusion of “Ground Truth”

The idea of "ground truth" in AI sounds reassuring, doesn't it? It suggests a foundation of indisputable facts, a solid starting point for machine learning systems to understand and interpret the world. But as Kate Crawford argues, this concept is far from what it seems. Training datasets, often presented as objective and factual, are actually riddled with contradictions and biases.

She explains it best:

“

Truth, then, is less about a factual representation or an agreed-upon reality and more commonly about a jumble of images scraped from whatever various online sources were available. (p. 96)

”

These datasets aren't carefully curated to reflect diverse perspectives or nuanced realities. Instead, they're cobbled together from whatever happens to be online, which often skews toward the most accessible or widely available content, not the most accurate or representative.

Think about it, when AI systems are trained on these datasets, what they perceive as "truth" is essentially a patchwork of cultural biases, historical gaps, and incomplete narratives. For instance, if an image dataset includes only Western representations of household objects, an AI trained on it might struggle to recognize similar objects from non-Western cultures. Its "truth" is only as good as the dataset it's fed.

This raises big questions: Who decides what gets included in these datasets? What perspectives are missing? And how can we challenge the idea of "ground truth" to create AI systems that better reflect the complexities of the world?

3. Skewed Inferences from Biased Data

The way AI learns is both fascinating and deeply flawed when its training data isn't diverse enough. Kate Crawford highlights this issue with a simple but powerful analogy:

“

Consider the task of building a machine learning system that can detect the difference between pictures of apples and oranges. First, a developer has to collect, label, and train a neural network on thousands of labeled images of apples and oranges. On the software side, the algorithms conduct a statistical survey of the images and develop a model to recognize the difference between the two classes. If all goes according to plan, the trained model will be able to distinguish the difference between images of apples and oranges that it has never encountered before. If all apples are red and none are green, then a machine learning system might deduce that 'all apples are red.

But if, in our example, all of the training images of apples are red and none are green, then a machine learning system might deduce that "all apples are red." This is what is known as an inductive inference, an open hypothesis based on available data, rather than a deductive inference, which follows logically from a premise. (pp. 96 - 97)

”

It's a clear example of how AI makes assumptions based on the patterns it observes in the data it's given. The problem is, these assumptions don't just stay confined to apples. When datasets reflect only a narrow slice of the world, AI systems start making faulty generalizations in areas that can have real-world consequences.

For example, if a facial recognition system is trained mostly on images of lighter-skinned individuals, it might struggle to correctly identify or differentiate darker-skinned faces. The result? A system that works well for some people but fails—or worse, discriminates—against others.

4. The Echo Chamber of Benchmark Datasets

AI's reliance on standardized benchmark datasets might sound like a step toward consistency and collaboration, but as Kate Crawford points out, it often creates a dangerous echo chamber:



Illustration of the concept of AI bias. I generated this image using Midjourney

“

These benchmark datasets become the alphabet on which a lingua franca is based, with many labs from multiple countries converging around canonical sets to try to outperform one another. (p. 98)

”

The result? AI systems trained on the same narrow datasets, reinforcing biases and stifling creativity. This echo chamber perpetuates existing biases and limits the potential for AI to truly innovate or represent broader perspectives. If everyone is training their models on the same dataset with the same gaps and blind spots, how can we expect AI to evolve beyond its current limitations?



An illustration of the concept of AI training (Canva)

5. Text and Language Are Politically Charged

Language is anything but neutral, and as Kate Crawford emphasizes. It carries the weight of time, place, culture, and politics.

“

There is no neutral ground for language, and all text collections are also accounts of time, place, culture, and politics. Further, languages that have less available data are not served by these approaches and so are often left behind.” (p. 103)

”

This reality dismantles the myth of “neutral” AI, especially when it comes to language models. AI language systems are trained on massive datasets pulled from the internet, places like Reddit, Wikipedia, and forums. These sources are far from balanced; they reflect the socio-political biases, cultural assumptions, and power dynamics of the people who created them.

Take the example of a chatbot trained on Reddit data. While Reddit offers a wealth of language for training, its content often skews toward certain cultural or ideological perspectives. A chatbot trained on this data might adopt toxic, discriminatory, or exclusionary language, simply because that's what it "learned" from the dataset. The AI isn't inherently toxic, it's reflecting the biases of the platform it was trained on.

6. Data as a Resource: A Colonial Metaphor

Kate Crawford's metaphor comparing data to a natural resource isn't just clever—it's a powerful critique of the extractive logic that underpins data mining in the AI age. She writes:



An illustration of the concept of colonial power (Canva)

“

Data began to be described as a resource to be consumed, a flow to be controlled, or an investment to be harnessed. The expression “data as oil” became commonplace, and although it suggested a picture of data as a crude material for extraction, it was rarely used to emphasize the costs of the oil and mining industries: indentured labor, geopolitical conflicts, depletion of resources, and consequences stretching beyond human timescales. Ultimately, “data” has become a bloodless word; it disguises both its material origins and its ends. And if data is seen as abstract and immaterial, then it more easily falls outside of traditional understandings and responsibilities of care, consent, or risk. As researchers Luke Stark and Anna Lauren Hoffman argue, metaphors of data as a “natural resource” just lying in wait to be discovered are a well-established rhetorical trick used for centuries by colonial powers. (p.113)

”

This analogy cuts deep. Like colonial powers extracting oil, minerals, or land without regard for the people or ecosystems involved, today’s tech companies treat data as raw material, free for the taking. The human experiences, emotions, and identities embedded in that data are stripped away, leaving behind a sanitized commodity that powers AI systems.

Think about it: when you share a photo, write a blog post, or even comment on a friend’s post, that content becomes fuel for AI models. Yet, no one asks for your consent, no one compensates you, and rarely does anyone acknowledge the humanity behind the data. By framing data as “just a resource,” tech companies sidestep the ethical responsibilities tied to its use, reducing your lived experiences to something transactional.

Crawford’s critique also highlights the broader consequences of this extractive approach. Just as colonial practices caused long-lasting harm, economic disparity, cultural erasure, environmental degradation—data extraction has its own fallout. It perpetuates inequality, reinforces bias, and often exploits the very individuals who are least aware of how their data is being used.



An illustration of the concept of exploitation (Canva)

7. AI and Social Inequities

Kate Crawford highlights a grim reality about AI: it's not just a technological marvel but also a system capable of perpetuating and amplifying social inequities. She writes:

“

A decade ago, the suggestion that there could be a problem of bias in artificial intelligence was unorthodox. But now examples of discriminatory AI systems are legion, from gender bias in Apple's creditworthiness algorithms to racism in the COMPAS criminal risk assessment software and to age bias in Facebook's ad targeting.(p. 128)

”

From gender biases in hiring algorithms to racial disparities in criminal justice tools, these systems are only as fair, or unfair, as the data they're trained on. Crawford underscores how AI systems trained on biased datasets often end up reflecting and reinforcing the inequalities already embedded in society.

For example, credit scoring systems have been shown to disadvantage women, offering them lower credit limits than men with the same financial profiles. Similarly, criminal justice algorithms like COMPAS have been found to unfairly label Black individuals as high-risk for reoffending, even when their actual likelihood of committing another crime is no greater than that of their white counterparts.

“

Every dataset used to train machine learning systems, whether in the context of supervised or unsupervised machine learning, whether seen to be technically biased or not, contains a worldview. (p. 135)

”

The issues aren't just about numbers, they affect lives. AI's failures can lead to women being denied opportunities, marginalized groups being unfairly targeted, or entire communities being excluded from the benefits of technology. Even voice recognition systems, which often struggle to understand female voices or accents, reveal how AI can fail those who fall outside its narrow training parameters..

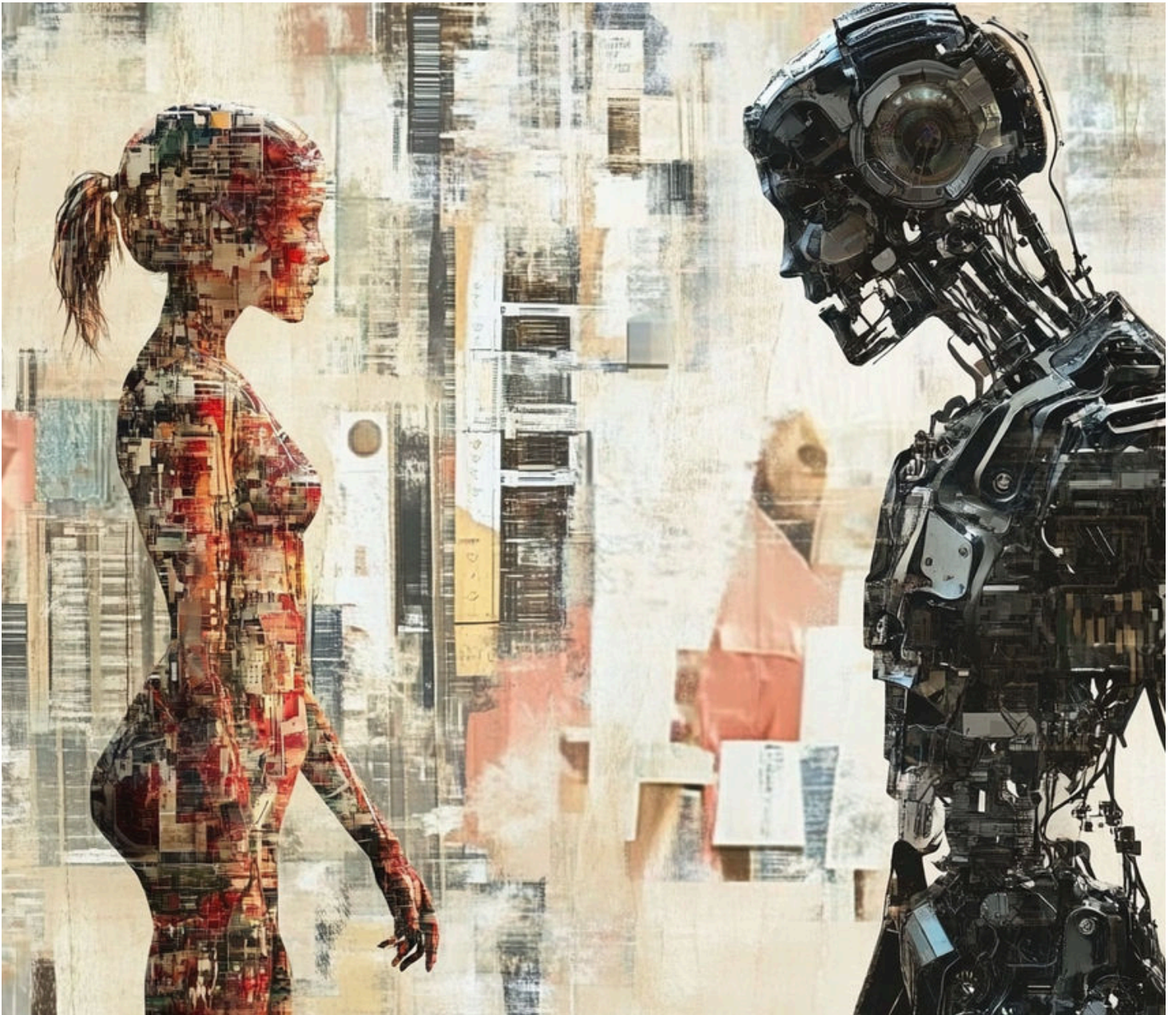


Illustration of the concept of AI-driven social inequalities. I generated this image using Midjourney

Conclusion

AI development comes at a significant cost, one we pay with our personal data and privacy, the depletion of planetary resources, and the erosion of socio-cultural well-being. As Kate Crawford compellingly argues, data is not an abstract or disembodied concept. It is deeply material and profoundly tangible, serving as the lifeblood of AI systems. The data we feed into these technologies not only powers them but also shapes how they perceive and interpret the world.

In turn, these systems mirror and magnify the biases, inequities, and structures embedded in the data, influencing how AI conceptualizes and interacts with society. If we are to harness AI's potential responsibly, we must confront these realities head-on and push for ethical, equitable, and transparent approaches to its development.

Also, benefiting from the affordances of AI should never blind us to the destructive practices it often relies on, particularly the extractive mining industry that fuels its infrastructure. The production of AI technologies, as Crawford contends, demands vast amounts of rare earth minerals and metals such as cobalt and lithium, minerals that are extracted under environmentally damaging and, in many cases, exploitative conditions.

These mining practices deplete natural resources, degrade ecosystems, and contribute significantly to greenhouse gas emissions. I believe that if we are to truly embrace AI as a tool for progress, we must also advocate for sustainable practices in its development, ones that prioritize environmental stewardship, ethical resource extraction, and accountability for the long-term health of our planet.



Image generated using Midjourney

References

- Crawford, K. (2021). Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press. Kindle Edition.
- Heaven, W. D. (2020) Predictive policing algorithms are racist. They need to be dismantled. MIT Technology Review.
<https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/>
- Mollick, E. (2024). Co-Intelligence: Living and Working with AI. Portfolio/Penguin. Kindle Edition.
- Nedlund, E. (2019). Apple Card is accused of gender bias. Here's how that can happen. CNN Business.
<https://www.cnn.com/2019/11/12/business/apple-card-gender-bias/index.html>
- Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59(236), 433–460.

About the author



Med Kharbach, PhD, is the editor of www.educatorstechnology.com. A seasoned educator with over 13 years of classroom experience, Med earned his doctorate in Educational Studies from Mount Saint Vincent University in Halifax, Canada. His scholarly work includes numerous publications in prestigious peer-reviewed journals, alongside co-authoring several impactful book chapters. Med's latest work is a book titled [ChaGPT for Teachers: Mastering the Skill of Crafting Effective Prompts](#). Currently, Med's research is passionately focused on exploring the integration and implications of AI in education. Med is currently working on a book on the use of AI in academic research.

Connect with Med on:

