

AI is Expensive



LSE's Alison Powell explains the real costs of AI, and how these might be mitigated.

AI is everywhere; AI is here. The story around AI implies that it is here because it's making things efficient: AI is better at detecting cancerous tumours in some scan images than radiographers, AI is faster at finding legal judgements within case law, AI can make office work more efficient by drafting emails or summarizing information from the web.

However, the story of this efficiency leaves out a discussion of some of the costs of AI. AI is expensive, not cheap. The efficiencies that are promised do not necessarily involve less work and fewer costs – just different work and different costs, some of which will only reveal themselves in time. These costs include:

1. Increased inequities (including inequitable labour between people 'in front of' and 'behind' the screen; inequitable opportunities for learning resulting from embedding of AI systems in learning and information infrastructures; environmental and material inequities resulting from the use of scarce natural resources to power ubiquitous technologies)
2. Shaky institutions that struggle to do things differently
3. Costly need for highly skilled review of AI outputs

First, we need to understand better what makes AI expensive. Then, we need to consider what factors can actually lead to real efficiency. My research has been examining both the deceptive stories that shape the way AI is described, as well as sociotechnical design considerations that must be taken into account when determining whether the cost of AI is worth it.

The First Cost: Errors, and correcting them

The story of AI's 'efficiency' focuses on the idea that computers are more efficient and less prone to error than humans. Yet in areas where AI has been celebrated as being 'better' than humans, many errors are occurring.

- For example, experts tested AI models that provided US election information, and all [performed poorly](#). Research has also indicated that finding – much less correcting – these errors requires highly-specific expertise, or work 'in front of the screen'.
- In the case of radiography, one area where enthusiasm about AI led to [widespread use](#), the original training data for the AI that worked more effectively than average radiographers was based on many years of [data](#) generated by skilled human radiographers. AI scans still make errors. These errors are likely to be different, and if future training data sets no longer include as many readings from skilled human radiographers, this may generate errors that are more difficult to identify or to guard against. While using an AI might reduce short-term costs, medium or long term costs are high, since they require radiographers to [become AI experts](#) and to identify when the automated system is not working.
- In legal scholarship, an assessment of AI tools used to summarise legal precedents has discovered that 1 in 3 results of a Westlaw AI project are [incorrect 'hallucinations'](#). Identifying the errors requires a lot of domain expertise, which, like radiography, means more pressure on experienced and skilled people. The risk here is that case review will take more time rather than less, or that really dangerous errors will create doubt about the foundations of the legal profession.

Reducing AI errors by increasing the data used to train systems or making the models more complex increases other costs.

The Second Cost: Fragile systems and inequalities

Reliance on AI can make systems fragile by reducing the range of ways that organizations and groups manage information, make decisions, and take action. It can also reduce the robustness of organizations by making many kinds of work more routine and some kinds of work more cognitively challenging.

- The kinds of tasks easiest to automate using AI often involve [doing MORE of some simple and straightforward tasks](#). That is, more summaries of meetings, more data visualisations, more basic insights. However, most AI assistance needs interpretation from skilled and expert people, as well as much more data cleaning, labelling and processing work from less skilled people. This produces global inequalities of labour.
- ‘Behind the screen’, workers with few labour rights [work in conditions shaped by ‘gamification’ or constant competition](#). In ‘front of the screen,’ fewer people are performing increasing numbers of complex tasks without variation. In both cases, these labour relations limit autonomy, competence and wellbeing at work.
- New research on the [use of AI by scientists](#) suggests that using AI can narrow the ways that researchers think about ideas, creating ‘scientific monocultures’ that preempt necessary new thinking. Similarly, research on the use of AI in mathematics instruction suggests that students whose learning is guided by the AI retain less than other students – when the ‘crutch’ of the system is removed, they struggle to complete problems. This lack of flexible thinking can have long term implications in a range of areas.

The Third Cost: AI costs more energy than you can imagine

- The economist Maria Mazzucato recently reported that “about [700,000 litres of water](#) could have been used to cool the machines that trained ChatGPT-3 at Microsoft’s data facilities”.
- The environmental politics of data centres is at the heart of research conducted by my former student Sebastián Lehuedé, who describes how the extraction and diversion of resources to [data centres makes global politics more unstable](#), as technology companies site their data processing centres in places with cheap electricity or access to water.
- This creates competition between data centres and housing, with capacity limits on the UK’s electrical grid already meaning that in West London, data centres are [competing](#) with new housing developments for access to electricity.

Embedding AI more extensively increases these resource constraints.

Deciding on how to use Expensive AI

The real costs of AI, as I've shown here, might come months or years after a system is put in place. They can include increased inequalities between people working 'behind the screen' to transform aspects of the world into computable data and people working 'in front of the screen' to identify errors generated by AI systems. A third cost includes the damage to environments and the diminishment of resources necessary for survival.

Mitigating or avoiding these costs requires re-thinking what makes for efficiency. In my work I develop three features of '[holistic technology design](#)' that should be applied to AI:

1. **Reciprocity:** the capacity for people to effectively reciprocate in relation to the potential impact of technology introduction, in different contexts, beyond consultation or 'acceptance'. This involves communicating an understanding of the full range of impacts of technologies rather than embedding them in social systems 'by default'.
2. **Temporality:** considerations of differently-timed impacts of technology – these include emergent inequities due to the spatial or temporal displacement of labour.
3. **Reversibility:** capacities to roll back or change technology systems in order to address various forms of inequity – this might mean removing or deciding not to introduce AI into a system or work setting where it might increase costs.

Using these principles to begin collaborations with environmental and well-being economists and specialists in public sector institutions can avoid the high costs of AI. So many stories about AI innovation are told by those who stand to benefit most

This post represents the views of the author and not the position of the Media@LSE blog, nor of the London School of Economics and Political Science. The author draws on research presented here in a [public event](#) at Inspire, University of Edinburgh, 5 June 2024.

Featured image: Photo by [Massimo Botturi](#) on [Unsplash](#)