

How Generative AI Fits into Knowledge Work

Mari Sako

Since OpenAI released ChatGPT in November 2022, we have seen a step change in excitement about generative AI, coupled with concerns about its safety. Given this inflection point, we need to pay renewed attention to its impact on the future of knowledge work carried out by professionals. This is because compared to earlier types of AI, generative AI gets closer to the core activities of professionals, namely giving advice to and treating clients.

And yet, how and how fast professionals' work will change is not well understood. Instead of leaving the issue to be part of "unintended consequences",ⁱ this column argues that we can influence how generative AI will become embedded in the work we do as professionals.

Professionals in a variety of fields – including medicine, audit, accounting, law, and data science – are essentially in the business of diagnosis and treatment, connecting the two via inference. Put simply, professionals have a claim to classify a problem (diagnose), reason about it (infer), and take action on it (treat).ⁱⁱ To date, AI had affected all areas of professional work, but primarily diagnosis; analysis of medical data (e.g. in radiology) or accounting and legal data (e.g. in due diligence) are good examples. But now, generative AI is moving the needle towards affecting all parts of professional work. This is exciting but also threatening for professionals.

This column first conceptualizes what professionals do. We then put a spotlight on generative AI and emerging use cases in professional work, with a view to raising key questions that begin to address when machines do better than human professionals, and in what ways machines complement humans. Much of the human-machine interaction is in the hands of the professionals themselves.

What do professionals do?

Knowledge workers think for their living. They create value with their expertise and critical thinking. A subset of them, professionals, evolved differently across

time and location.ⁱⁱⁱ But professional work in different fields has three common modalities: diagnosis, inference, and treatment. Take doctors as an example. In diagnosis, doctors ask questions of a patient and carry out tests with a view to diagnosing a patient's symptom. In inference, doctors derive a prognosis from their medical knowledge. In treatment, doctors prescribe medicine and/or carry out an operation.

While the language used is medical, other professionals essentially have the same three modalities. Litigation lawyers may firstly conduct discovery (i.e. diagnose relevant facts in a case), exercise legal reasoning to derive the best way to argue a case (i.e. infer) and represent the client in court (i.e. treat). Investment bankers help their corporate clients collect relevant financial information in due diligence (i.e. diagnosis) before using their finance knowledge (i.e. inference) to recommend the best financial structure for mergers and acquisitions. And data scientists clean and explore data (i.e. diagnose), build models to analyze and interpret (i.e. infer), before presenting data visuals for a specific audience.

In all these professional contexts, inference that connects diagnosis to treatment is based on expert knowledge that has a theory component and a practice component. A good doctor is good not just because she studied at a top medical school, but also because with years of practice she can refine her diagnosis and treatment from having seen other patient cases. Skilful performance in professional work typically depends not only on theoretical knowledge obtained during formal training, but also on tacit knowledge and heuristics that the performer finds difficult or impossible to articulate fully. As Michael Polanyi said, “we know more than we can tell.”^{iv} Just as tacit knowledge underpins riding a bicycle, staying afloat in water, and playing the violin, it is essential in many aspects of professional judgement.

Generative AI use cases

Readers of this *Communications* magazine require no reminder about the technological advances that lie behind the surge in generative AI. In particular, the discovery of a network architecture, the Transformer that is based solely on attention mechanisms, dispensed with recurrence and convolutions entirely.^v With massive compute power and big data, large language models (LLMs) are being deployed primarily to generate text-based data.

Various use cases of generative AI models are emerging, including knowledge retrieval, clinical decision support, and summarization of key findings in medicine; legal research and the generation of contracts and other documents

such as deposition summaries in law; and co-piloting of code generation in data science.

Many of these use cases involve professionals potentially exploiting the co-occurrence of two or three modalities of professional work. For example in audit and accounting, an AI model may flag up anomalies and instances of non-compliance in tax law. And it is a short step from this diagnosis to treatment, in which potential instances of likely non-compliance may be pre-empted before they occur. Prompt engineering in LLMs also makes it more likely that diagnosis and treatment co-occur, for instance self-diagnosis and self-care in health.

Generative AI and professionals: what we know

Although the technology is still nascent, some patterns are emerging about how the performance of generative AI can be improved with or without humans in the loop, potentially overtaking human performance.

First, in comparing the performance of LLMs and that of humans, the gap is closing with updated model versions. Specifically, GPT-3.5 had already passed medical, law, and business school exams despite mediocre performance. GPT-4 does a better job to ace the bar exam, and it has a reasonable chance passing the CFA exam for financial professionals.^{vi} While GPT-4 finds quantitative parts of various exams more challenging, this can be addressed by equipping it with the ability to execute Python code. Passing exams, just like detection e.g. of very small cancer tumors, is a matter of improving accuracy, which machines do well by being thorough and consistent.

Second, domain-specific LLMs trained by professionals engaging in reinforcement learning from human feedback (RLHF) perform better than LLMs trained only on general-purpose text corpi such as Wikipedia. In legal research, for example, WestLaw Precision by Thomson Reuters and Lexis+ by LexisNexis are powered by LLMs trained by domain experts, i.e. attorneys in these cases. RLHF may be also applied to further training of a general-purpose LLM. For example, the law firm Allen & Overy trained the GPT-3.5 model with lawyer prompts and responses that were kept within the firm.

What we don't (yet) know

But there is less reliable evidence on how professionals' usage of generative AI affects the quality of their work. This is not just because professionals are still experimenting with early adoption. It is also because professional training,

professional control, and the division of labor between juniors and seniors within a profession is yet to be worked out. Evidence below give rise to further questions.

First, take the implications of generative AI for professional training. There is evidence that less experienced professionals benefit more from the use of GPT-4 recommendations than more experienced professionals.^{vii} For example, using a co-pilot to generate codes is like having a personal trainer to become data scientists. But does this lead to a virtuous circle of junior professionals using generative AI to accelerate their training? Or will the use of generative AI in early career stages lead to skipping important exploration, including making mistakes from which one learns?

Second, professionals care deeply about the quality of work they do, as they should. And yet, there is worrying evidence that professionals tend to regard generative AI as a source of inaccuracy, leading to lower quality work.^{viii} Shouldn't generative AI enable professionals to achieve the same quality of work but in less time? One piece in this puzzle is the difficulty in assessing the quality of professional work, particularly when quality is not just a matter of accuracy. For example, a document summary should be also complete and nuanced. Moreover, in creative activities, ChatGPT may be able to generate more new ideas of varying quality than humans.^{ix} But there is also evidence that the best humans still outperform ChatGPT in creative thinking tasks.^x

Third, human-machine interactions remain complex in professional settings, and is likely to evolve as the machine performance improves over time. There is evidence that experienced professionals tend to ignore machine recommendations when they judge that accuracy of the outputs is not high enough. Such “algorithmic aversion” is juxtaposed against human reactions when faced with high quality AI assistance. Studies found that access to high quality AI induced workers to exert less effort, a state of “falling asleep at the wheel”^{xi} or “asleep at the keyboard.”^{xii} Paradoxically therefore, maximizing performance from human-machine interaction may require lower quality AI than is technologically feasible. This raises the question: what is an optimal degree of dependence, without over-reliance or under-reliance, on AI to validate and explain the LLM outputs? How can we ensure that professionals remain vigilant (“awake”) at the keyboard when using co-pilots that “auto-complete”?

Future of professional work

Amidst all excitement about macro-projections, such as that generative AI could add between \$2.6 trillion to \$4.4 trillion annually to the global economy,^{xiii}

this column characterized professionals as doing diagnosis, inference, and treatment. Generative AI not only affects all three modalities of professional work. It also enhances the possibility of two or three modalities co-occurring, transforming the business of advisory work. Consequently, regulatory advice and regulatory compliance might morph into one, as might self-diagnosis and self-treatment in healthcare.

Professionals are already taking things into their own hands to influence the nature of interaction between humans and AI machines. For example, Hollywood script writers won their case in controlling the misuse of AI-generated film content; ongoing court cases involve the alleged infringement of copyrights held by programmers whose codes are available on Github's Copilot and OpenAI's Codex AI Program.

But we need to go beyond these concerns about intellectual property to address specific questions that touch all professionals. (i) Given the current state of technology, what is the best mode to validate machine outputs without falling into the trap of over-reliance ("asleep at the keyboard") or under-reliance (due to algorithmic aversion)? (ii) In validating AI model outputs, what quality measures can be developed over and above the data science measures of accuracy? (iii) As AI performance improves, is there a less-than-perfect performance level – like imposing speed limits in road transportation -- at which human professionals are most satisfied and motivated to perform consistently well? Professionals should lead in addressing these questions, to the benefit of themselves and society at large.

- ⁱ Cusumano, M. A. 2023. Generative AI as a New Innovation Platform. *Communications of the ACM*, 66(10): 18-21.
- ⁱⁱ Abbott, A. 1988. *The System of Professions: An Essay on the Division of Expert Labor*. Chicago: University of Chicago Press, Chapter 2.
- ⁱⁱⁱ Sako, M. 2013. The business of professionals. *Communications of the ACM*, 56(7): 30-32. Sako, M. 2020. Artificial intelligence and the future of professional work. *Communications of the ACM*, 63(4): 25-27.
- ^{iv} Polanyi, M. 1966. *The Tacit Dimension*. New York: Anchor Day.
- ^v Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
- ^{vi} Lakshmi Varanasi "GPT-4 can ace the bar, but it only has a decent chance of passing the CFA exams. Here's a list of difficult exams the ChatGPT and GPT-4 have passed", *BUSINESS INSIDER* (updated November 5, 2023), <https://www.businessinsider.com/list-here-are-the-exams-chatgpt-has-passed-so-far-2023-1?r=US&IR=T>
- ^{vii} Brynjolfsson, E., Li, D. and Raymond, L.R. 2023. **Generative AI at Work**. National Bureau of Economic Research working paper.
- ^{viii} ThomsonReuters. 2023. **Future of Professionals: How AI is the catalyst for transforming every aspect of work**. ThomsonReuters.
- ^{ix} Girotra, K., Meincke, L., Terwiesch, C., & Ulrich, K. T. 2023. Ideas are dimes a dozen: Large language models for idea generation in innovation. **Available at SSRN 4526071**.
- ^x Koivisto, M., & Grassini, S. 2023. Best humans still outperform artificial intelligence in a creative divergent thinking task. *Scientific reports*, 13(1): 13601.
- ^{xi} F. Dell'Acqua. 2022. **Falling asleep at the wheel: Human/AI collaboration in a field experiment on HR recruiters**, Working paper.
- ^{xii} H. Pearce et al. 2022. **Asleep at the keyboard? assessing the security of github copilot's code contributions**. IEEE Symposium on Security and Privacy (SP): 754-768.
- ^{xiii} McKinsey & Co. 2023. **The Economic Potential of Generative AI: the Next Productivity Frontier**. <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>