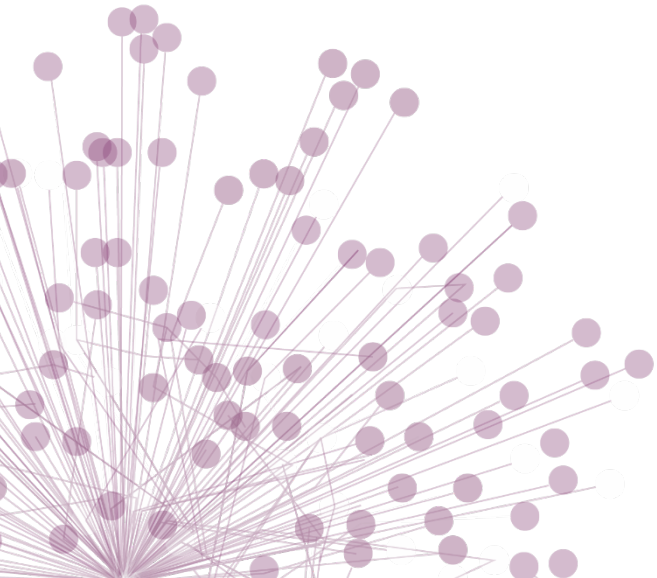
University of Southampton

WEB SCIENCE INSTITUTE

Large Language Models:
Prediction, pollution and projection

WSI Position Paper 2024-05
December 2024

Wendy Hall and Ben Hawes



About the WSI

The Web Science Institute (WSI) brings together the University of Southampton's world-class, interdisciplinary, socio-technical expertise in web science, data science and artificial intelligence (AI) to leverage the unique role of online technologies in tackling global challenges. We work to create collaborations within the University and with industry, governments and third sector organisations that bring interdisciplinary socio-technical insights and impacts to the world's most pressing problems.

https://www.southampton.ac.uk/research/institutes-centres/web-science-institute

Web Science Institute Building 32, Highfield Campus, University of Southampton, SO17 1BJ
wsi@soton.ac.uk (DOI: 10.5258/SOTON/WSI-WP012)

About the Authors

Dame Wendy Hall is Regius Professor of Computer Science, Associate Vice President (International Engagement) and is Director of the Web Science Institute at the University of Southampton.
Wendy is an advisor to the UK government and other governments and companies around the world. In 2023 Wendy was appointed to the United Nations high-level advisory body on AI. Her latest book, Four Internets, co-written with Kieron O'Hara, was published in 2021.

Dr Ben Hawes is a technology policy consultant and an associate director at the Connected Places Catapult. As a civil servant, he worked on the 2017 review of Artificial Intelligence, the national Internet of Things Programme IoT.UK, the 2011 review of Intellectual Property and the Digital TV Switchover plan, among other technology policy programmes.

Technology stock prices have been volatile in recent months. There is a range of views about how and when major technology companies could recoup dividends from the very large investments they have made in Large Language Models (LLMs).

LLMs are generative AI models. They are trained on large volumes of textual data, including from the open internet, to generate credible new text in response to questions and other prompts. Other generative AI models do something similar with pictures, sound and video. LLMs started receiving increased attention and investment after OpenAI released ChatGPT to the public in late 2022. General users were able to play with it, and, importantly, use it immediately for their own purposes. Its capabilities were easy to appreciate. Print journalists (and authors of opinion pieces) acknowledged with horrified fascination the labour-saving and career-threatening implications for them of a technology that creates a draft of an article in seconds.

There is dispute around estimates of how many people are using LLMs, how much and for what. Some reports suggest that - in terms of use by a substantial proportion of the population - they are being taken up faster than personal computers, the internet or mobile internet were. Many institutions, corporations and investors were surprised, even wrong-footed, by ChatGPT's capabilities, and then again by improvements seen in the successor GPT-4.

Major technology companies rapidly adapted their strategies to respond and compete, launching or accelerating their own LLM model development. There is evidently a lot of FOMO, fear of missing out, but underlying that has been excitement that this kind of AI model might offer an accelerated route to achieving general purpose usability and usefulness.

Not everyone has shared the hype. Last year on BBC's Talking Business, one of us pointed out that what was going on looked a little like the early days of the public internet. Early internet users were excited, particularly because it was free to use once connected, which felt novel then. If the experience was clunky by current standards, we could all see it improving quickly and continually. Investors piled in, rightly seeing the potential, but the business models were not there yet, and some of those investors lost everything before it became clear which business models worked.

And then there was Blockchain, which was going to solve many really hard problems, and did not.

The cognitive scientist Gary Marcus has questioned the fervour around LLMs, and recently he has strengthened his position, suggesting that investment in them is a bubble that will burst.[1] There are many questions around LLMs including about the data used to train them, potential misuses, their reliability as information sources, how useful they really are and for what, and what future models could be developed from them.

This is not at all to say that LLMs do not have impressive capabilities or the potential to develop more, only that projecting from where we are now is an uncertain business.

Debts, risks and liabilities

Not being magical, LLMs do not generate textual information from nothing. They have been trained on text. Much of the text that has been used for training LLMs is copyright and has been used for training without the permission of copyright-holders. Reasonably enough, many copyright-holders believe their rights have been infringed. Some LLM developers are now licensing content they use for training, but significant claims have not yet been resolved, and no international legal consensus or widely accepted licensing processes have yet emerged.

Some training datasets include information about individuals, triggering potential data protection liabilities. Google and X have both suspended use of some EU residents' personal data for training their models. Meta has announced plans to use posts by users of WhatsApp, Instagram, Threads and Facebook, which has generated resentment among some users.

Model developers may face more than claims for compensation if they use data for training without a legal right. The US Federal Trade Commission and private litigants have in some cases sought model deletion as a penalty: “the deletion of models trained on unlawfully used or possessed data”.[2]

As more datasets become unavailable, some model developers have reported a shortage of data for training. However, using synthetic data for training instead can create additional problems.

From the copyright and privacy perspectives, LLMs rely more on other people’s work and data than their developers seem to want to acknowledge. From the perspective of truth and reliability, LLMs are too original. LLMs can generate content that is convincing, but factually wrong. When that happens, it is not an untypical glitch. It happens because they are trained to write text that is a credible imitation of previous texts. They are optimised for formal consistency with existing texts and not for truthful representation of reality. Unless measures are taken to check their outputs before use, there is no anchor to that reality.

These errors of fact are described as “hallucinations”. This is an unhelpful term because it suggests a self within the LLM that could understand the difference between true and false information.

If outputs need to be checked in case they are false, how useful are they? That depends what you use them for. In many areas of work, when creating the first draft of a document the value of immediately filling the empty page will be well worth the cost of having to check the facts. But if you use an application to deliver information through a medical advice chatbot or to prepare for a legal trial, you will need to have much lower tolerance for inaccurate and wholly invented information.

Some commentators propose that an information service that uses a LLM, for instance a search engine, should have obligations to deliver factual answers: a duty to tell the truth. A group of researchers recently proposed “the creation of a legal duty

to minimize careless speech for providers of both narrow- and general-purpose LLMs and derived commercial applications."[3]

Like visual and audio generative AI techniques, LLMs can also be used deliberately to generate or distribute false information. Much recent concern around deepfakes has been about video and audio that convincingly and falsely present people, including political or media figures. However, misleading text can also be harmful. LLMs make it vastly easier to generate text to use for misinformation and fraud, at large scale and low cost, and spread it around through multiple channels.

False information generated by LLMs may then go into datasets that are used to train future models, with implications for those developing the models, and for the rest of us.

LLM-generated misinformation can undermine those models, as a recent article sets out. "We find that indiscriminate use of model-generated content in training causes irreversible defects in the resulting models, in which tails of the original content distribution disappear. We refer to this effect as 'model collapse' and show that it can occur in LLMs."[4]

Inaccurate information flooding from LLMs onto the internet can also degrade the communal information sphere. The potential harms from this are hard to model, but there are obvious and large scale potential threats to trust and efficiency, for internet users. In relation to the open internet, LLMs might come to be seen as both parasites and polluters.

As well as impacting the information environment, internet search powered by LLMs may also deter users from going to the online sources from which the LLMs are drawing information, and so reduce the traffic that those sources rely on to sustain publication.

LLMs are costly to train, and already have a significant and increasing collective energy footprint.

These are all possible sources of future liabilities that are difficult to quantify now. Some of those liabilities could yet fall to the corporations developing the models. Others appear to be externalities, which is to say liabilities that may fall on everyone else.

Compounding uncertainty

It generally takes time to explore what a technology is good for and how to work with it, and longer to develop the skills and organisational forms to deliver in practice in economically productive ways. The lags have been very different for different technologies. While analysts do their best to find common patterns and factors, there is no single trajectory with predictable timing for the journey from technical innovation to market uptake, and social and economic benefits.

Mobile phone and internet use grew fast in comparison with earlier technologies. After the bursting of the internet investment bubble of the early 2000s, the exponential growth of a small number of global internet platforms has made success look easy, even inevitable. Their success created expectations that further innovations driven by the same companies with the same focus on accumulating data and compute and funding research will follow a similar upward path.

Google and Facebook have been sustained by much the same virtuous circle. They collected data about users activity and used it to improve their computing models and their service to users. Improved services drew in more users for more time spent online, which allowed the platforms to collect more data, and so on. The platforms used some of that data to direct advertising, paid for by client companies. This positive feedback and the user lock-in secured by network effects delivered growth, revenues and rising stock values, which allowed further investment in computing capability, technical innovation and further accumulation of data. The corporations achieved winner-takes-all positions in their markets, with the power to out-compete or buy promising new entrants. To date, market regulation has not developed tools to address dominant data holdings, so this type of data-driven market power in technology sectors has come to seem unchallengeable.

Users also saw the benefits from ongoing improvement in the services, although many now feel that the user experience is no longer improving. As Cory Doctorow has memorably described, dominance in each market allows platforms to pay less attention the quality of their core service to users.[5] However, so far the platforms continue to dominate their markets, and advertising revenues continue to deliver.

There is an opaque aspect to their success. Because they provide the product – targeted advertising opportunities – and the market mechanism, it can be difficult to verify the value of that product from the outside. To date, though, advertisers pay and stock markets continue to have confidence in the core advertising business of both.

It is hardly surprising that these companies hope that collecting more data and investing in compute and innovation will generate a similar virtuous circle with LLMs, and some elements of that are in place. Users try out models. Data from that use enables refinement of the model, and collection of more data. But other parts of the virtuous circle – user lock-in sustained by network effects and advertising revenue – do not appear to be there in the same way. The models will continue to improve, at least in some respects, but they are not continually paying back while they improve.

To date, no dominant revenue model for LLMs has emerged that delivers as the internet platform advertising business model has and does. This is not by any means to say it never will. Microsoft may be able to leverage its vast data about how organisations operate to deliver LLM-powered business services that prove transformative for business customers.

Massive live experiments are in train, with Meta, Apple, Google, Microsoft and OpenAI taking different approaches. So far none of those approaches obviously offers the network effects and winner-takes-all outcomes seen in Web2 platform markets. What and where is the consumer or user surplus that can be captured and

monetised? Perhaps that winner-takes-all dynamic will emerge in limited market sectors (as it has in internet platform sectors) but not in LLMs for general purpose uses. That might be better for consumers and for the AI and internet ecosystems.

Even when it builds on great past success by a company, technology investment is a wager based on a prediction, not a transaction for a known outcome. Stock markets manage a great deal of private sector technology investment, taking data from many sources to give prices based on estimated future corporate values and revenues. Markets are also fallible prediction machines, with a record of drifting away from reliance on facts from time to time, before correcting, sometimes violently. Like LLMs, they sometimes extrapolate from past data to get things wrong.

To the extent that corporations are incentivised to increase shareholder value, rather than directly to deliver the best products and services, they can tend to promise more than they can deliver. This is generally priced in, but can cause collective over-excitement where there are more unknowns, including around new trends and technologies.

Researchers can also exaggerate progress in the hope of attracting more funding. A recent paper “Questionable practices in machine learning” described 43 “questionable research practices… which fall short of outright research fraud.”[6] Competition for funding in AI is intense, increasing the temptation to exaggerate achievements.

In part because of the problems with accuracy, it may be particularly difficult to assess the breadth of potential application areas for LLMs, or the depth of their usefulness in each area. They might become indispensable at stages in processes, like automated translation and some types of coding. Wider application may yet require combination with other techniques, to compensate for the areas where LLMs fall short. They may not continue to advance as the general purpose technologies they appear to be at first encounter.

Perhaps we are unduly impressed by technologies that imitate us well. Surely if something is like us it must be genuinely clever? LLMs may have done this better than anything else so far. But it is not certain that an imitation machine, if improved, will eventually become something else. The bet seems to be that models will improve until a different level of performance is reached, demonstrating significant advances in completing tasks and solving problems, constituting a general purpose intelligence. From where we are now, that may be more a matter of faith than science-based forecasting.

In his book *Rebooting AI*, Gary Marcus points to limitations that could prevent an approach depending solely on LLM models from delivering further fundamental advances.[7] He suggests that future major technical advances in AI and subsequent resulting productive gains may be more likely to come from combining other AI techniques with LLMs. Google among others has introduced tools that check the factual basis of LLM outputs, which seems a welcome development.

Improvements in LLMs have kept enthusiasm going, but not everyone shares it. We do not expect a repeat of the bursting of the dotcom bubble in the early noughties.

There are some similarities, but the context is different. The major players here are large and very well resourced. They could lose investor confidence, be forced to revise strategies and investments, and still survive. But a correction of some kind seems possible. It is exciting to watch, but there is a risk that expecting too much too soon from LLMs could result in failure to push forward research in other AI techniques, and limit the richness of AI that gets developed.

Perhaps it is time we talked a little less about AI in general, as if it were a single wave of technology development, and more about the different techniques and their potential future trajectories and combinations. In technology investment, hype helps until it doesn't.

---

[1] Why the collapse of the Generative AI bubble may be imminent, Substack, August 2024.

https://garymarcus.substack.com/p/why-the-collapse-of-the-generative

2 The Deletion Remedy, Daniel Wilf-Townsend, September 2024. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4933011

[3] Do large language models have a legal duty to tell the truth? Sandra Wachter, Brent Mittelstadt, Chris Russell, August 2024.https://www.sciencedirect.com/science/article/pii/S0388000124000615

4 AI models collapse when trained on recursively generated data, by Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson and Yarin Gal, Nature, July 2024. https://www.nature.com/articles/s41586-024-07566-y

[5] The 'Enshittification' of TikTok, Or how, exactly, platforms die, by Cory Doctorow, Wired, January 2023. https://www.wired.com/story/tiktok-platforms-cory-doctorow/

[6] Questionable practices in machine learning, by Gavin Leech, Juan J Vazquez, Misha Yagudin, Niclas Kupper, Laurence Aitchison July 2024. https://arxiv.org/pdf/2407.12220

[7] Rebooting AI: Building Artificial Intelligence We Can Trust, Gary Marcus and Ernest Davis, Pantheon Books, 2019.