History and Generative AI

Professor Mark Humphries, Wilfrid Laurier University, Ontario, Canada

Something is in the air at universities and secondary schools these days.[1] Generative artificial intelligence (AI) has been around for a while, but in late 2022 it reached the mainstream when OpenAI's ChatGPT burst onto the scene.[2] Anecdotally, it seems that most educators first became aware of the technology after hearing about its potential for plagiarism and academic dishonesty. But I think we are witnessing something far more radical and consequential take shape.

During the first half of 2023, there has been an avalanche of new developments in AI, too numerous to name here. But the most important are undoubtedly the release of OpenAI's ChatGPT-4 in March and Meta's open-source model called Llama-2 in July.[3] The point is that the AI race is accelerating. The job market is already being transformed but so too will academia, teaching, and the discipline of History in particular.[4] Yet many historians and educators remain rightly skeptical that a so-called 'plagiarism machine' has any relevance to their work, beyond adding another layer of complexity to authenticating student work.[5]

What many people seem to miss about the AI revolution, though, is that the automation of writing is AI's most visible but least remarkable feature. Most would probably be surprised to hear that its ability to pound-out student book reviews is a byproduct rather than an explicit feature of the technology. Indeed, Generative AI is revolutionary because it has (much to the surprise of its own creators)[6] evolved to a point that we can begin to automate the collection, organisation, and synthesis of information as well as its communication. So, if you are a skeptic, hear me out because I am going to explain why even if we do not like what the revolution portends, we cannot escape the fact that we are already in its midst.

A long time coming

I think of myself as an unlikely convert to AI. Although I have always used a lot of technology in my work as an historian, I have never called myself a digital humanist. In my experience, that term has been reserved for scholars who build text-clouds to visualise word frequencies, print 3D models of artefacts, or make use of augmented reality – amongst many other things. I know this can be exciting, but it has never felt relevant to what I do.

I see myself as a traditional historian: I assign research essays and book reviews in my classes, and I prefer to write books and articles based on detailed archival research. Nevertheless, I have taken thousands of digital images at archives, used OCR software to transcribe documents, and even have an automatic microfilm scanner at home. But for me, technology has always been peripheral; I have just never been enamoured with digital outputs. I like a good detective story, which is what drives me as a scholar.

In this way, I may be typical of my generation: I am comfortable with technology, so much so that I do not think much about it. I was born in 1981 and computers have always just 'been there' as part of the research and writing process. I was educated entirely during the first digital revolution, which brought lots of changes to how we do things as historians, but still allowed the traditional to at least co-exist with the digital. For most of us, computers and then the Internet allowed us to do the things historians had always done, just faster and more efficiently. By the early 2000s, when I went to university, none of this required special knowledge or skills.

An end to the beginning of the Digital Revolution

Generative AI is about to end this détente between the traditional and the modern. AI is disruptive precisely because it significantly reduces the learning curve and time necessary to gather and process information and then generate new content, be that artwork, music, or text. A tool like GPT-4 is, in essence, a digital Swiss army knife: a novice user can now upload an Excel spreadsheet full of historical data and, in plain language, ask GPT-4 to generate graphs and tables or perform complex tasks with the data that would normally require years of experience in coding. All this in seconds and through a conversational, non-technical interface. In essence, generative AI allows us to outsource much of the work involved at becoming good at something like painting, composition, writing, or data analysis.

In this regard, it is unfortunate that most mainstream media attention has focused on ChatGPT's writing abilities alone, for this obscures its much more interesting ability to perform customised and adaptable task-based reasoning that requires some level of 'thinking'.[7] That they can do this at all (or even fake it) is a bit surprising as the Large Language

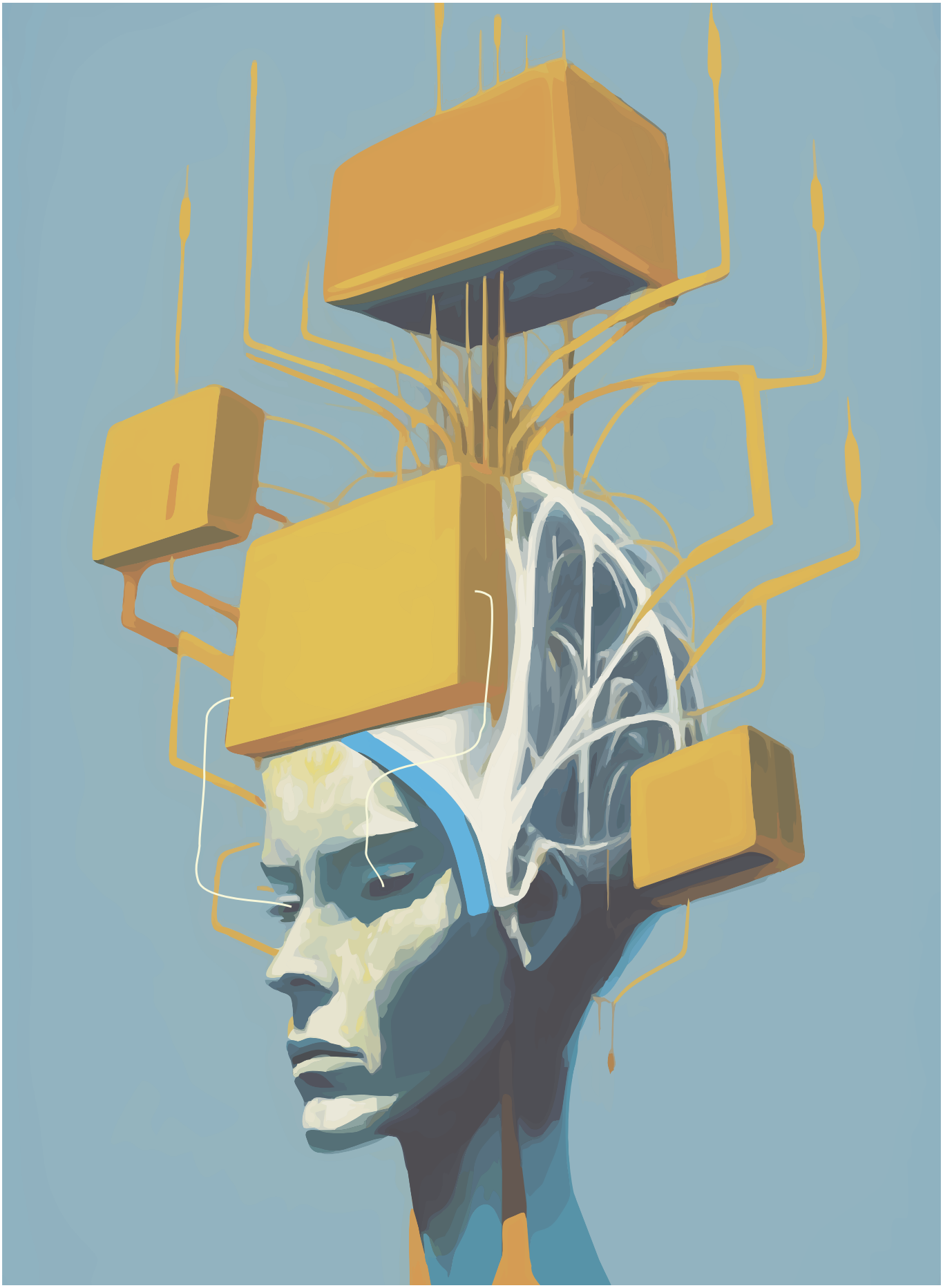

Figure 1. *'Artificial Intelligence,' one of a set of vector illustrations, poster design and watercolour painting on background, Molibdenis-Studio. Source: Shutterstock / public domain.*

Models (LLMs) underpinning AI chatbots are actually trained to respond to prompts by reading tens of billions of pages of text and learning to predict the next word.

Every time ChatGPT-3.5 turbo (the 'free' version) generates part of a word (called a token) it goes through 175 billion computations to determine a probable way to continue the sentence. But that does not necessarily mean it always chooses the most probable (or indeed factually correct) thing to say next. This is a crucial point, because if it did that, its answers would always be identical. Instead, because it is meant to be dynamic and conversational, it is programmed to introduce an element of randomness into the equation, controlled through a parameter called 'temperature'. This makes the LLM sometimes choose less probable outputs. The higher the temperature setting, the more 'original' the response will feel to a human – but it will also increase the chance of factual or contextual errors. Elaborate, detailed prompts with clear instructions and parameters almost always elicit the best responses. This is why some AI scholars have called LLMs stochastic parrots or a version of the mirror test.[8]

I am fascinated by the debate about whether LLMs can truly 'reason' and have been reading and listening to everything I can on inscrutable matrices, 'black boxes', and emergent capabilities – it is fascinating and scary stuff. Few people realise that this is all more alchemy than science at present. Think about this: the people who built these things do not know exactly how they work. In essence, researchers fortuitously discovered when you train a neural network containing a very large number of parameters to predict the next word using an extremely large corpus of text, all by themselves (and quite unexpectedly) these networks just started doing interesting things that their creators did not initially expect nor train them to do. This is the reason that some very smart people in the AI field are starting to get worried: humans built a powerful technology, and we cannot explain how it works or why.[9]

Although some very prominent voices in the academic community might dismiss generative AI's ability to think and reason as a mirage, I think they have yet to provide a convincing rebuttal that addresses the fact that when you ask GPT-4 to do something very complex, it very often succeeds. While Noam Chomsky and others have dismissed feats like passing theory of mind tests or acing standardised tests as an illusion – arguing that LLMs are only simulating reasoning – I am not sure it matters in practice.[10] While the term 'reasoning' may not be strictly accurate, that seems to be a vocabulary issue more than anything. Whatever the correct word for it may be, 'reasoning' does seem to capture the fact that when you ask GPT-4 to do something quite complex – that involves doing something we would normally describe as 'thinking' like parsing a series of difficult to follow instructions – it usually just works.[11] While I think it is important to understand what this involves under the hood, especially for high-risk applications like medicine, it is less crucial when all you want GPT-4 to do is create a table from data in a text file. In 'low stakes' cases like this, it either works or it does not, and when it gets it right, it does not much matter whether it was thinking or is just really good at predicting the next word. At the end of the day, a useable, accurate graph is a useable accurate graph.

AI historians and research assistants

This practical usefulness, specifically the ability to follow instructions and solve problems, is what makes LLMs so powerful. And as historians and History educators, we need to acknowledge that we can harness this power to do useful things for us too. Indulge me for a moment. Like most social historians, I am intrigued by the stories of ordinary people and spend much of my time reading mundane, inglorious, and overlooked sources to generate detailed pictures of ordinary lives lived in historical obscurity. It is an arduous and inefficient process. For example, I spent years generating detailed biographies of hundreds of ordinary shell-shocked soldiers from medical records, war diaries, personnel files, newspapers, and veterans' records to help me understand what happened to them and why. I more recently did a similar thing for around 300 Canadian voyageurs from St-Benoit parish using voyageur contracts, parish records, legal documents, and fur trade records. In both cases, I had to sift through hundreds of different sources, looking at one individual at a time, one source at a time, to generate mini-biographies or life histories. To me, the value of this approach is that it often reveals hidden trends and patterns that are rarely written down in the memoirs, diaries, and letters of elites. The trade-off is that it is time consuming and repetitive, to say the least.

But what if we had an AI agent, not one that would think and write for us, but one that would help us with the monotonous parts of the research process, specifically retrieving and linking information across disparate datasets much faster and more efficiently than any human could hope to do? This is already possible using GPT-4 and what is called a customised vectored database filled with transcribed historical documents. Vectored databases attach something called 'embeddings' to chunks of text or images. These long, floating-point numbers look and act like GPS coordinates, representing the meaning or content of blocks of text and images in an imaginary semantic three-dimensional space. Just as GPS coordinates tell us how near we are to a given point on the Earth, embeddings can be used to retrieve the data from the database that is nearest or most relevant to a given query. This data might be several pages of text, a few sentences, or even single points like a birth year and name. When the user generates a query, GPT-4 uses this database, rather than its training data, to answer questions, which makes it more accurate and useful.[12]

These interactions do not use the web-based ChatGPT that you are probably familiar with, but a GPT-4 Application Programming Interface (API). APIs allow one computer program to call another computer program over the Internet, send it information, and get a response. In the context of generative AI, APIs are used to send information to a generative AI model like GPT-4, ask it to do something with that information, and then get the results. These interactions are also highly customisable: you can give the AI specific

instructions, change its temperature settings (which affect the consistency and randomness of the output), and require output in a specific format (like a table). You can also send it customised data and tell it to confine its analysis to that information, limiting 'hallucinations' (false information that is not based on real data or events). It sounds complicated, but once you learn the ropes, it is more powerful and flexible as well as a lot faster and more efficient than using ChatGPT.

This sort of program is nothing new and is very similar to the types of agents and 'Chat-with-my-PDF' type plugins that people have been developing since the GPT-3.5 API was released to the public in February 2023.[13] Yet historians have some unique needs that those off-the-shelf solutions do not address. In sifting through documents, we need to pay attention to chronology and dates and ensure that citations are accurate. While this does not work with ChatGPT (either 3.5 or 4), it is all possible with properly structured vectored databases and search functions, but they need to be customised for the ways historians ask questions and the types of results we expect to see. In one sense, this is just another type of search function, but it is also more than that. Here is a clear use case where search often fails: how many times have you spent hours looking for a specific document that you know you have seen in your files but cannot quite remember the wording or what folder it was in? It would save a lot of time and effort to be able to describe its contents in plain language to an AI assistant that could then go find it based on a somewhat vague description. Semantic searching using vectored databases allows you to find things that are similar in concept, but not necessarily in wording.

Let's look at a more expansive example. Imagine you are looking for information on how soldiers responded to fear in battle by going through a database of their letters. You would quickly find they did not always use words like 'scared' or 'afraid' and often employed euphemisms – not always, but enough that it makes keyword searching databases difficult. While the obvious solution is to just 'read all the letters' manually, this is not practical if you are faced with tens of thousands of individual documents. To solve this sort of problem, we might normally construct a sample of some sort, but this would mean going through an awful lot of irrelevant data to find a few precious examples that may or may not be representative (depending on how you chose your sample). No matter how you slice it, it is a very time consuming and labour-intensive task. This is also what historians have always done: there was never any alternative. The unpleasant truth is that all too often, though, all that work results in only a few lines in whatever article or book chapter you are working on.

But with an embedded search, paired with an AI agent, you can quickly find all the relevant examples in a data set – without having to categorise or code them first. It is a technical process, but surprisingly easy to accomplish. In a nutshell, your query is given an embedding via an API like OpenAI's Ada-002 to represent its conceptual and semantic content. Provided you have the information you need loaded into a vectored database, the AI will then try to match your query to the embeddings from letters that are close to it in an imagined semantic space. Done properly, the results will be topically and conceptually relevant and you can iterate through all the examples in a very large database quickly – in mere seconds, in fact.

Let's expand this a bit more. With enough tweaking and database building – and access to an LLM with state-of-the-art visualisation capabilities – you could also have it do some additional lateral processing, pulling up biographical information on the author of a letter from their personnel file and census records. An AI agent could also go through unit war diaries and find out exactly what a soldier's unit had been doing in the days leading up to the point when the letter was written. It could even go through newspapers to find out what the recipient may have been reading when they got the letter in the mail. The possibilities for this type of lateral, recursive research are limited only by the availability of digitised documents – or the ability of the historian to digitise the documents themselves. The LLM technology is also not there yet, but it is truly in sight. Unlike the promise of truly self-driving cars, which has always led the actual technology, in this case LLMs already have these capabilities; its just a matter of coding, deploying, and training them to do history-specific tasks. That will take time, but it will be measured in months and years, not decades.

The future of historical research

In the very near future, imagine harnessing this AI reasoning power to do something that would be impractical for a single human or team of humans to do in a lifetime. Here I think of the remarkable Canadian *Programme de Recherche en Démographie Historique* (PRDH) database which documents the lives of everyone that lived in New France/Lower Canada (modern Quebec) from the 1620s to the 1860s.[14] This project has been going on for decades and is a truly invaluable resource that has enabled countless studies. As I keep experimenting with this technology, its clear that it will soon become feasible to construct a similar database for Canada as a whole, covering any combination of records from 1842 through the 1930s. When you think about it, this has been theoretically possible for years; the only limiting factors were the enormous number of human research assistants and the amount of time required. Yet AI reduces these costs exponentially by promising to automate processes that are time consuming and repetitive, compressing days of work into seconds or minutes. Rest assured, genealogy websites have been experimenting with this sort of technology for some time and will soon make this a core aspect of their business. They are already using it to transcribe and index the 1931 census in partnership with Library and Archives Canada.

Uncertain futures

So, what will this generative future hold for the humanities and History in particular? A caveat: beware of anyone who claims to have the answer (including me), for they are probably wrong, at least about the specifics. But perhaps not in the way you suspect. What has astonished me during the first half of 2023, is that in an article I published in February, many of the things that I predicted were years away

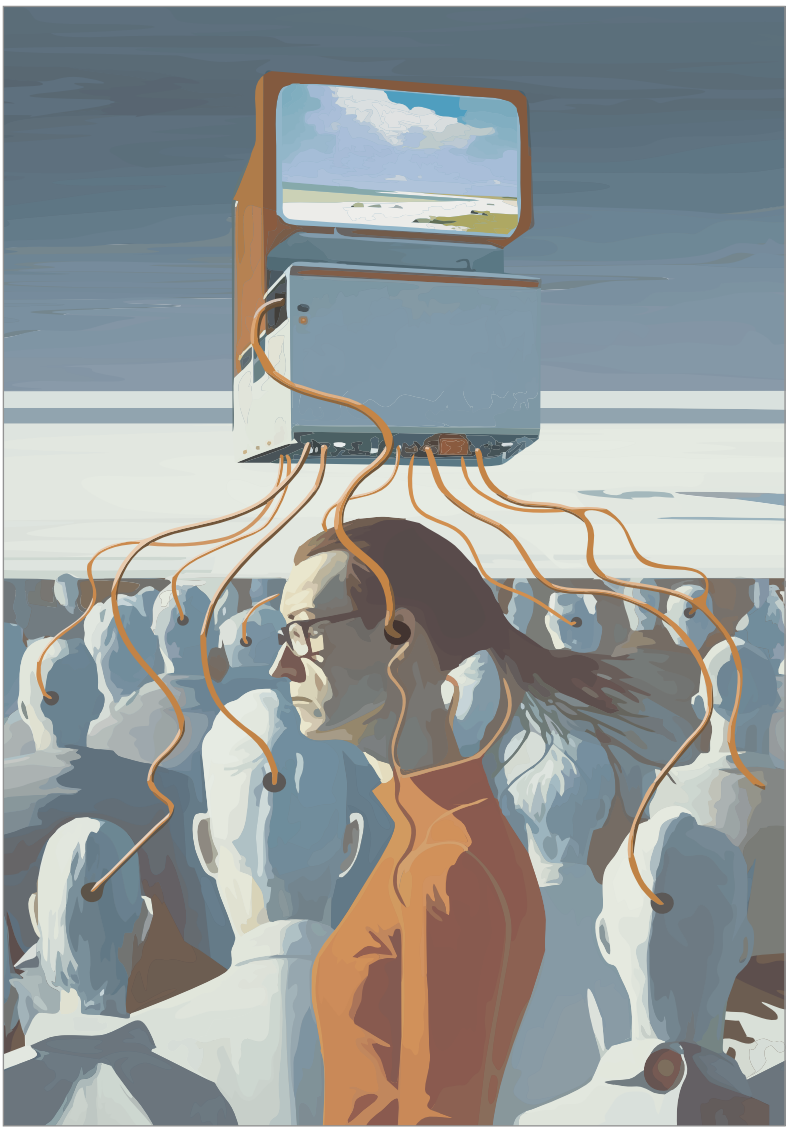

Figure 2. *'Artificial Intelligence,' one of a set of vector illustrations, poster design and watercolour painting on background, Molibdenis-Studio. Source: Shutterstock / public domain.*

materialised weeks or months later. The timeframe only continues to accelerate; as I write I am already testing my own local Llama model based on seventy billion parameters in my office – something I would have told you was years away six months ago.[15]

What is clear, however, is that as AI tools are integrated into Google, social media, and Microsoft Office over the next few months, we are going to have to feel our way forward. We will obviously have to wrestle with questions of authorship, authenticity, and attribution – things that we already know well. While AI adds a new dimension, they are largely old wine in new bottles. We have always needed to be clear and transparent about our methodology and the authorship of the works we publish, and that will undoubtedly soon include statements about AI use in publications. A trickier question is that very soon our graduates will enter a work-world where they are expected to make efficient use of both AI and their own human talents. Figuring out how to get them there is a new problem.

Conclusion

Perhaps somewhat ironically, this is why I am now more optimistic about the future of the humanities and the teaching of History, than I have been in many years. You see, the efficient use of AI requires efficient operators with a good grasp of language, a good general knowledge about the world, and the critical thinking skills necessary to get the best results from the machine and fact check hallucinations. Here we can take the basic skills we have always practised and taught and dress them up for a new era with the latest AI jargon. University Deans and School Principals will no doubt be ecstatic. But my sense is that we also must make some very real changes, whether we want to or not.

What is also becoming very clear to me is that the continued development of AI research is also going to require collaboration with humanists, and in particular, historians. Making LLMs both more reliable and more accurate is becoming less of a technical computing problem than it is a methodological training problem. To get better, LLMs will need to get larger and have more computing power, but they also need to learn to 'think' more efficiently and effectively.

As historians and History educators, we claim to specialise in training students to find relevant sources, assess their trustworthiness and authenticity, to examine them in the proper context, and to arrange them hierarchically. Footnotes (as opposed to in-text citation) are designed to lay the research process bare; they are, in essence, all about interpretability. They are meant to show not only where information came from, but to demonstrate how we arrived at our conclusions. These are exactly the issues facing AI developers today. This may sound idealistic, but it is precisely why History has so often served as a training ground for a whole range of other professions. AIs may well be our next crop of students.

As this revolution takes hold, we need to be prepared to rethink how we teach research, critical thinking and writing. We need to acknowledge that AI can be a useful tool in this process and pioneer effective ways to maximise its potential. It likely means developing new assignments and new approaches to evaluation. It also means incorporating AI into our own work as well. This need not be as frightening as it sounds, as at least some of it will happen organically. But we also need to be aware of the pitfalls of AI, that it is alchemy, and is highly fallible. In this sense, it is a strange and unusual beast. Not since humanity relied on the horse have we been asked to work with something that we cannot fully understand and that may (or may not) do our bidding. ♦

[1] This article is adapted from posts on the author's substack, *Generative History*, https://generativehistory.substack.com (accessed 16 August 2023).

[2] OpenAI, ChatGPT, https://chat.openai.com (accessed 16 August 2023).

[3] Samantha Murphy Kelly, 'ChatGPT's future: 5 jaw-dropping things GPT-4 can do', *CNN Business* (16 March 2023), https://www.cnn.com/2023/03/16/tech/gpt-4-use-cases/index.html (accessed 16 August 2023); Meta, 'Meta and Microsoft Introduce the Next Generation of Llama', *Meta Newsroom* (18 July 2023), https://about.fb.com/news/2023/07/llama-2 (accessed 16 August 2023).

[4] Casey Noenickx, 'Workplace AI: How artificial intelligence will transform the workday', *BBC* (17 May 2023), https://www.bbc.com/worklife/article/20230515-workplace-ai-how-artificial-intelligence-will-transform-the-workday (accessed 16 August 2023).

[5] Rose Horowitch, 'Here Comes the Second Year of AI College', *The Atlantic* (7 August 2023), https://www.theatlantic.com/ideas/archive/2023/08/ai-chatgpt-college-essay-plagiarism/674928 (accessed 16 August 2023).

[6] Mark Sullivan, 'The scary truth about AI chatbots: Nobody knows exactly how they work', *Fast Company* (17 May 2023), https://www.fastcompany.com/90896928/the-frightening-truth-about-ai-chatbots-nobody-knows-exactly-how-they-work (accessed 16 August 2023).

[7] For a good example see: Ian Bogost, 'ChatGPT Is Dumber Than You Think', *The Atlantic* (7 December 2022), https://www.theatlantic.com/technology/archive/2022/12/chatgpt-openai-artificial-intelligence-writing-ethics/672386/ (accessed 16 August 2023).

[8] The best readable overview of how LLMs work is: Stephen Wolfram, 'What Is ChatGPT Doing … and Why Does It Work?', *Stephen Wolfram Writings* (14 February 2023), https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/ (accessed 16 August 2023).

[9] Cade Metz, 'Microsoft Says New A.I. Shows Signs of Human Reasoning', *The New York Times* (16 May 2023), https://www.nytimes.com/2023/05/16/technology/microsoft-ai-human-reasoning.html (accessed 16 August 2023). The actual paper is: Sébastien Bubeck et al., 'Sparks of Artificial General Intelligence: Early experiments with GPT-4', (arXiv preprint arXiv:2303.12712v5 [cs.CL], 13 April 2023) https://arxiv.org/abs/2303.12712 (accessed 16 August 2023).

[10] Noam Chomsky, Ian Roberts, and Jeffrey Watumull, 'The False Promise of ChatGPT', *The New York Times* (8 March 2023), https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html (accessed 16 August 2023).

[11] Ethan Mollick describes many of the complex tasks GPT-4 can do in detail at: Ethan Mollick, 'It is starting to get strange', *One Useful Thing* (13 August 2023), https://www.oneusefulthing.org/p/it-is-starting-to-get-strange (accessed 16 August 2023).

[12] For an introduction to vector databases see: Adrian Bridgwater, 'The Rise of Vector Databases', *Forbes* (19 May 2023), https://www.forbes.com/sites/adrianbridgwater/2023/05/19/the-rise-of-vector-databases/?sh=5ff4166e14a6 (accessed 16 August 2023).

[13] For an introduction to using AI agents on customised data see: Tom Davenport and Maryam Alavi, 'How to Train Generative AI Using Your Company's Data', *Harvard Business Review* (6 July 2023), https://hbr.org/2023/07/how-to-train-generative-ai-using-your-companys-data (accessed 16 August 2023).

[14] Drouin Institute, *Programme de Recherche en Démographie Historique*, https://www.prdh-igd.com (accessed 16 August 2023).

[15] Mark Humphries and Eric Story, 'Today's AI, Tomorrow's History: Doing History in the Age of ChatGPT', *Active History* (1 March 2023), https://activehistory.ca/blog/2023/03/01/todays-ai-tomorrows-history-doing-history-in-the-age-of-chatgpt/ (accessed 16 August 2023).