Written evidence submitted by Professor Michael Wooldridge (UAIG0032)

The much-trumpeted recent advances in AI such as Large Language Models (LLMs) like ChatGPT have led to considerable speculation about possible applications. This brief document summarises what I see as the key opportunities and risks associated with the use of AI in government. Comments apply to all varieties of AI, but LLMs in particular.

Opportunities

- *Productivity*. The UK has a well-documented productivity problem, and the public sector is no exception. We could argue about the causes, but the potential for LLMs to be productivity tools for white collar workers are tantalising. LLMs can be used to automatically summarise and collate documents, compare documents, extract key points, rewrite text for different audiences, identify logical inconsistencies, automatically generate draft presentations, and more. These are *routine intellectual labour*: they don't require deep human insight, but are time consuming, and are the core activity of thousands of whitecollar roles. I believe this is the single most important possible application of this technology.
- *Bringing corporate memory to life.* Large organisations such as HMG have enormous repositories of "unstructured" data – ordinary written text such as emails, meeting minutes, reports, policy documents, and so on. For the most part this data languishes on our computers unread. Imagine feeding this enormous repository of text to an LLM, and then asking questions about it. A routine but time-consuming question like "What committees approved the decision to lift the lifetime allowance on pensions in 2024?" could be answered in seconds. You could go on to ask what the main objections were, what the perceived advantages were, and so on. In this way LLMs have the opportunity to *bring corporate memory to life* – and to leverage a data repository that is currently largely unused.
- *New government services*. The World-Wide Web enabled a new class of government services, reducing public workload and costs (e.g., Self Assessment tax returns). AI will enable further new opportunities for government services, and I recommend attention being given to the many possibilities. Consider for example an AI program that helps someone understand and plan for their pension. A dream would be a single AI program that provided a single uniform interface to all government services. Such a program could even be personalised to different audiences (teenagers, elderly, etc).
- *Automated workflows*. Consider a typical local government activity such as processing planning applications. The public hate it (they find the process convoluted, expensive, time consuming and opaque), and local governments hate it because it's expensive and time consuming (and it makes the public hate them). It's entirely plausible to consider automated feedback systems, trained on previous planning applications, that would give immediate feedback on draft applications, for example highlighting areas where a proposal needs improvement. The ultimate dream would be to fully automate such processes, although this seems a somewhat distant possibility.

Risks

- *Over inflated expectations.* The recent advances in AI are genuine and as an AI researcher I find them truly exciting: but that does not mean AI is a silver bullet for our nation's problems. It doesn't even mean that AI is ready for widescale rollout. It is

important to have a sober understanding of where AI can safely and productively be used, and to manage expectations. It is *very* hard to put AI into practice.

- *AI gets things wrong, a lot*. The public perception of AI is that it is super-intelligent, and capable of incredible feats of reasoning and problem solving. It isn't. It is primarily a system that pattern matches on enormous quantities of text, picking up on scenarios that seem to match the one at hand. In particular, LLMs have no conception of truth or falsity, and as a consequence they get things wrong – a lot. These problems are deeply baked into the technology, and no truly reliable fixes seem likely to be available in the short term. This is likely to be a *serious* impediment to the rollout of AI services in government: an AI program that advises me no income tax is payable, only for me to discover that this is not the case when I submit my tax return, will cause intense frustration (and likely legal cases).
- *Use of private data*. AI requires data; lots of data. The more data, and the higher quality data, the better. HMG is sitting on an enormous repository of data, which is incredibly valuable. This could surely be leveraged for an enormous number of socially beneficial applications. Some of this data is relatively uncontentious – making available data about London's public transport system would enable a host of new products and services, and stimulate innovation. But releasing data about UK citizens or organisations to Silicon Valley would, I think, be an enormous mistake. Once we hand over our data, whatever promises are made and whatever covenants are placed on it, *we have lost control.* UK AI services should be developed in the UK, on secure, trusted, UK-owned data centres. More generally, there are a raft of issues around the use of personal data for training AI, which raise enormous sensitivities. Enormous care needs to be taken to navigate these issues if we are to build beneficial AI programs without running roughshod over our rights with respect to private data.
- *New opportunities for cybersecurity attacks*. Large Language Models are powerful tools, and preventing them from being used to enable criminal activities is a major ongoing challenge. Unfortunately, the "guardrails" provided to prevent inappropriate use of such technology are fragile, and not hard to circumvent. The moment an LLM-based government service is released, it will immediately come under attack: some attackers will be nothing more than vandals, but more consequential attacks also seem inevitable. Apart from the risk that public models are used to (e.g.) enable terrorist attacks by providing recipes to build IEDs, there is also a concern that such attacks may expose government data used to train the model. As with the more general problem of cybersecurity, AI security will be a game of cat-and-mouse for the foreseeable future; and attacks of ever-increasing ingenuity on government AI systems seem inevitable.
- *Public trust*. Some people are excited about AI; but many more are worried about it. They are worried about their jobs, about their privacy, and they may even be worried (wrongly) about existential threat. However well motivated the use of AI in government is, I think it is likely that the government use of AI will therefore be met by scepticism (at best) and hostility and anger at worst. These fears – however misplaced – need to be taken seriously, and transparency is absolutely essential to build trust.

January 2025