



Subscribe



# OpenUK - embracing AI openness is essential to foster trust



By **George Lawton** July 9, 2024



Audio mode



Dyslexia mode

**SUMMARY:** OpenUK's Phase Two Open Manifesto suggests raising the bar on AI openness to improve AI safety, but doing so will require confronting concerns about missing out and enabling bad actors.

0 Comments

OpenUK published Phase Two of the [Open Manifesto Report](#) to guide the incoming UK government in driving open source software adoption, increasing employment, boosting the UK economy, and cultivating trust. They recommend building on the UK's lead in setting a high bar for openness in AI.



Subscribe To Our FREE Newsletter

This website uses cookies to ensure you get the best experience on our website.

[Cookie Settings](#)    [Accept all cookies](#)

## Quantifying AI open-washing

At a launch event for the manifesto, Andreas Liesenfeld, Assistant Professor at Radboud University, dug into a recent [paper](#) he co-authored on AI washing that distills some dimensions of AI openness. They broke the concept of AI openness across 14 dimensions across availability, documentation, and access.

At one end of the spectrum, Open AI's ChatGPT is one of the least open by these measures, although it produces good technical results. Meta's various, open weight Llama large language models (LLM) score slightly better but are still closer to the bottom. Mistral's LLMs are in the middle. At the top are LLMs like OLMO, BLOOMZ, and AmberChat, which I have never heard of.

The EU AI Act has exemptions for open source AI, but the definition is still a work in progress. Liesenfeld notes:

**"** *I think it's one of the most important conversations not being had about the European AI Act. It exempts open source, and it's unclear what open source is.*

When his team started looking under the hood, he realized a lot was happening. He explains:

**"** *In generative AI, openness has to be composite. So, it has to be made up of different individual slices of this technology that you have to look at. So, there is not just one individual data point that would suffice when we want to classify this technology as open or not. You have to come up with a spectrum of individual dimensions that you look at.*

At a high level, availability allows someone to audit the training pipeline to retrace the training process, documentation considers how well the pipeline is documented, and

[Subscribe To Our FREE Newsletter](#)

This website uses cookies to ensure you get the best experience on our website.

2. Documentation of reinforcement learning weights, license, code, architecture, preprint, model cards, and datasheets.
3. Access via packages or API.

Many of these things were hard to quantify with a simple yes/no answer. So, they had to develop a gradient to represent many aspects of whether a project was fully open, partially open, or closed. Even then, they felt they were being reductive and missing some nuance. For example, they noticed a tendency to under-report how reinforcement learning data gleaned from users fed into a model's training process.

One big challenge has been the trend toward conflating open weights with open source AI. In this case, a company will publish the weights of their model, but under restrictive licenses or without revealing their training data, which may be biased. Lisander says:

**“** *So by just releasing the weights, you create that kind of interesting win-win situation for at least some end users, and also for the companies who get away with not disclosing large parts of this technology and thereby also avoiding legal exposure. And all the questions that come with it, like ‘Are we going to be held accountable for the training data that went into our system?’ and so forth.*

However, Lisander raises concerns that this dilution of open AI could limit visibility into new decision-making engines that impact governments, businesses, and citizens:

**“** *We need more choice for the people, and we need to put the people's needs center stage for that. And we need to give people options of what technology they want to use when it comes to an open source AI. And I hope the government will push back, on one side, against the influence of big tech, which we have seen is very, very substantial in the EU at least, and on the other side also come to grips with regulating that technology in a smart way and drawing that line really in the right way, where you encourage smaller actors and those who really put in the work to make this system as open as it can be, give them more a*

[Subscribe To Our FREE Newsletter](#)

This website uses cookies to ensure you get the best experience on our website.

## Patience could be an AI virtue

Perhaps it's time to take a step back to examine what is going on under the cover of darkness. Sure, the models are getting better, and this is raising panic among businesses and government officials who want to keep pace with AI front-runners. However, the current tools also consume a lot of energy and raise questions about trust. Patience is required to navigate the next wave of innovation to build a better future for all of us, argues Neil Lawrence, DeepMind Professor of Machine Learning at Cambridge:

**"** *Although big tech is claiming they've done these extraordinary things, all they've done is taken university ideas and supercharged them with larger compute and larger data... But because they could afford to do that, they never stopped to think. One of the really important things in open innovation in this space is by saying that 'you have to do all this massive training in order to be at the cutting edge,' which I think is not true because they didn't stop to think they're excluding the notion that actually there are other ways of doing this. There are cheaper ways of doing this, and we are already starting to see that in the open ecosystem.*

*As soon as the [Meta] Llama weights came out, people started showing things that no company had done because the scale and breadth of the innovation ecosystem is greater even than Google and OpenAI can muster. What strikes me as bizarre about that is, at the time, I was on the open on the AI Council, shortly before it was disbanded, and we were telling the Government, when they were sort of panicking about this, 'let's build BritGPT.' We were saying there's going to be open models, and they were saying, 'ridiculous, that's never going to happen.' We've seen this pattern again and again. One company will decide that their best option is to release models in some open way to disrupt the incumbents. That will happen. Don't make stupid decisions now.*

Lawrence argues that a better strategy is to empower individuals to use these new tools with confidence since they are the ones who are aware of problems and exceptions. It's

[Subscribe To Our FREE Newsletter](#)

This website uses cookies to ensure you get the best experience on our website.

On the surface, CEOs are entranced by the notion that artificial general intelligence (AGI) will replace us quite soon. That's nonsense, declares Lawrence, who has been thinking deeply about these things for decades. We are grappling with the bandwidth asymmetry between humans and machines, not AGI. He says:

**"** *This is what's affecting society, not just now in the last two years when everyone started wetting themselves about ChatGPT, but before this, with social media and other forms of machine learning with relatively simple algorithmic decision-making systems that had access to enormous amounts of our data. So, they don't have to be more intelligent than us. If they're looking at quantities of data that are many millions of times more.*

Rather than focusing on the latest algorithms, we need to invest more time figuring out how to include more voices from the front lines in building better ones. Lawrence observes:

**"** *That's the strength of the open source community, and if deployed properly in the areas we're talking about, that can be the strength of the UK, our government, and our education system, that we realize that bringing these different voices to the conversation. It may require more effort at convening at the beginning, but it does mean everyone knows what they're doing, and it doesn't matter that they're from different cultures and different languages. They're working towards a shared goal.*

## What about the baddies?

An important consideration in all of this is that better AI could also empower bad people to do more bad things. It's already starting with deep fakes, better ransomware campaigns, and more effective cyber-attacks. But Lawrence finds it concerning that tech companies are having private conversations about these matters with policy makers who lack the expertise to push back. Instead, we need to empower experts on the ground with the best tools available to fix problems while simultaneously stopping the baddies. He explains:

[Subscribe To Our FREE Newsletter](#)

This website uses cookies to ensure you get the best experience on our website.

*just as we're saying with the manifesto, education, public service is a really great place to start because it's a place that's been eroded.*

*When you look at the challenges we face, you can see that within the UK and probably in the wider world, there's been a massive loss of trust in professional expertise and undermining of teachers, undermining of civil servants, and undermining of all these people that we now need to step up and experiment and understand how to best use and deploy these technologies that are sort of wading in and saying, 'You people aren't doing a good job. We're going to replace you with process.'*

*Well, if we accept that, if we accept that we're replacing everyone with process, then AI wins because AI is just process. If we want to bring the humanity back into the equation, we have to empower and retrust people in society to build and deploy these techniques.*

## My take

AI can produce some amazing results... and even more amazing failures. We are still coming to grips with how they break, waste resources, and sow mistrust. All these things require more sunshine and openness to examine and navigate. AI openness nutrition labels are a good starting point.

It will not be an easy path. A lot of money is betting against it. But there is time to be patient. Although many may paint AI as a race, it's certainly not in our best interest to race to the bottom, either.

*Image credit - © Warchi - Canva.com*

## Get your weekly enterprise AI digest

Complete the form below to receive the top enterprise AI stories from diginomica, every week.

[Subscribe To Our FREE Newsletter](#)