Andreas Liesenfeld, Assistant Professor, Radboud University State of Open: The UK in 2024 Phase Two: The Open Manifesto Report

Andreas Liesenfeld,
Assistant Professor,
Radboud University

Thought Leadership: Towards meaningful open AI: key findings and societal implications

The past year has seen a steep rise in generative AI systems that claim to be open. But how open are they really? The question of what counts as open source in generative AI takes on particular importance in light of the EU AI Act that regulates *'open source'* models differently, creating an urgent need for practical openness assessment.

Recent work by Andreas Liesenfeld and Mark Dingemanse (Centre for Language Studies, Radboud University, The Netherlands) provides a stark view of the openness of current generative AI. In a systematic sweep of the landscape, they survey 45 text and text-to-image models that bill themselves as open. Key findings and contributions include the following:

1. Open-washing is widespread. Open-washing is the use of terms like *'open' and 'open source*' for marketing purposes without actually providing meaningful insight into source code, training data, fine-tuning data or architecture of systems. Corporations like Meta, Mistral and Microsoft regularly co-opt terms like 'open' and 'open source' while shielding most of their models from scrutiny.
2. The EU AI Act puts legal weight on the term 'open source; without clearly defining it, creating an incentive for open-washing, opening up a key pressure point for lobbying, and necessitating clarity about what constitutes openness in the domain of generative AI.
3. Openness in Generative AI can be meaningfully measured if we think of it as composite (consistent of multiple elements) and gradient (it comes in degrees). The work distinguishes 14 dimensions of openness, from training datasets to scientific and technical documentation and from licensing to access methods.
4. Meaningful openness is possible, and is exemplified by some of the smaller players in the field, who go the extra mile to document their systems and open them up to scrutiny. We identify organisations like AllenAI (with OLMo) and BigScience Workshop + HuggingFace (with BloomZ) as key players moving the needle of openness in generative AI.

Why this matters for everybody

Although the EU AI Act creates a particular sense of urgency, openness is in fact of key importance for innovation and science, as well as to society. Here are three key points why openness in generative AI is important to society at large:

1. Openness helps build critical AI literacy by demystifying capabilities. If a company claims their AI can 'pass the bar exam', it helps to know exactly what was in the training data. The sheer magnitude of training data for GPT4 means that it can effectively take the bar exam 'open book', which is much less impressive. As AI applications are fast becoming ubiquitous, openness about training data and fine-tuning routines helps us equip the general public with a critical understanding.
2. Openness grants agency to people subjected to automated decisions. A clear example is the LAION image dataset underlying a widely used image generator. The openness of this dataset enabled auditing by Dr. Birhane and others which brought to light the presence of problematic material and prompted revisions and retractions. pre that AI is often best seen as a shorthand for automation. Transparency about training data and model architecture makes models auditable, and can help to make biases visible.
3. Openness provides real benefits on the ground. Corporations like to hand-wave at 'AI safety' as a reason to keep systems under wraps, but this is mostly a thinly disguised attempt to obscure clear and present harms already underway, including increasing spam content in search engines and the rapid spread of misinformation. Only open systems can be responsibly used in science and education and can support a healthy open source ecosystem.

Evidence-based openness assessment can help foster a generative AI landscape in which models can be effectively regulated, model providers can be held accountable, scientists can scrutinise generative AI, and people can make informed decisions for or against using generative AI.

Open GenAI: LLMs (simplified table)

Project	Availability						Documentation						Access	
	Open code	LLM data	LLM weights	RL data	RL weights	License	Code	Architecture	Preprint	Paper	Modelcard	Datasheet	Package	API
OLMo 7B Instruct	✔	✔	✔	✔	✔	✔	✔	✔	✔	✘	✔	✔	✔	~
BLOOMZ	✔	✔	✔	✔	~	~	✔	✔	✔	✔	✔	✔	✘	✔
AmberChat	✔	✔	✔	✔	✔	✔	~	~	✔	✘	~	~	✘	✔
Open Assistant	✔	✔	✔	✔	✘	✔	✔	✔	~	✘	✘	✘	✔	✔
OpenChat 3.5 7B	✔	✘	✔	✘	✔	✔	~	✔	✔	✔	~	✘	✔	~
Pythia-Chat-Base-7...	✔	✔	✔	✔	✘	✔	✔	✔	~	✘	~	~	✔	✘
Cerebras GPT 111...	~	✔	✔	✔	✔	~	✘	✔	~	✘	✘	✔	✘	✔
RedPajama-INCITE...	~	✔	✔	✔	✔	~	~	~	✘	✘	✔	✔	✘	~
dolly	✔	✔	✔	✔	✘	✔	✔	✔	~	✘	✘	✘	✔	✘
Tulu V2 DPO 70B	✔	✘	~	✔	✔	~	~	~	✔	✘	~	~	✘	✔
MPT-30B Instruct	✔	~	✔	~	✘	✔	✔	~	✘	✘	~	✘	✔	~
MPT-7B Instruct	✔	~	✔	~	✘	✔	✔	~	✘	✘	✔	✘	✔	✘
trlx	✔	✔	✔	~	✘	✔	✔	~	✘	✘	✘	✘	~	✔
NeuralChat 7B	~	✘	✔	✔	✔	✔	~	~	✘	✘	~	~	~	✘
Vicuna 13B v 1.3	✔	~	✔	✘	✘	~	✔	✘	✔	✘	~	✘	✔	~
minChatGPT	✔	✔	✔	~	✘	✔	✔	~	✘	✘	✘	✘	✘	✔
ChatRWKV	✔	~	✔	✘	✘	✔	~	~	~	✘	✘	✘	✔	~
BELLE	✔	~	~	~	~	✘	~	✔	✔	✘	✘	~	✘	✘
Geitje Ultra 7B	✘	~	✔	✔	✔	✘	✘	~	~	✘	~	~	✘	~
Phi 3 Instruct	✘	✘	✘	✘	✔	✔	✘	✔	~	✘	✔	✘	~	✔
WizardLM 13B v1.2	~	✘	~	✔	✔	~	~	✔	✔	✘	✘	✘	✘	✘
Airoboros L2 70B G...	~	✘	~	✔	✔	~	~	~	✘	✘	~	~	✘	✘
ChatGLM-6B	~	~	✔	✘	✘	✔	~	~	✘	~	✘	✘	✘	✔
Mistral 7B-Instruct	~	✘	✔	✘	~	✔	✘	~	~	✘	✘	✘	~	✔
WizardLM-7B	~	~	✘	✔	~	~	~	✔	✔	✘	✘	✘	✘	✘
Qwen 1.5	~	✘	✔	✘	✔	✘	~	~	✘	✘	✘	✘	~	✔
StableVicuna-13B	~	✘	~	~	~	~	~	~	~	✘	~	✘	✘	~
Falcon-40B-instruct	✘	~	✔	~	✘	✔	✘	~	~	✘	~	✘	✘	✘
UltraLM	✘	✘	~	✔	~	✘	✘	~	✔	✘	~	~	✘	✘
Yi 34B Chat	~	✘	✔	✘	✔	~	✘	✘	✔	✘	✘	✘	✘	~
Koala 13B	✔	~	~	~	✘	~	~	~	✘	✘	✘	✘	✘	✘
Mixtral 8x7B Instruct	✘	✘	✔	✘	~	✔	✘	~	~	✘	✘	✘	~	✘
Stable Beluga 2	✘	✘	~	✘	✔	~	✘	~	~	✘	~	✘	✘	~
Stanford Alpaca	✔	✘	~	~	~	✘	~	✔	✘	✘	✘	✘	✘	✘
Falcon-180B-chat	✘	~	~	~	~	✘	✘	~	~	✘	~	✘	✘	✘
Gemma 7B Instruct	~	✘	~	✘	~	✘	✘	~	~	✘	✔	✘	✘	✘
Orca 2	✘	✘	~	✘	✔	✘	✘	~	~	✘	~	✘	✘	~
Command R+	✘	✘	✘	✔	✔	~	✘	✘	✘	✘	~	✘	✘	✘
LLaMA2 Chat	✘	✘	~	✘	~	✘	✘	~	~	✘	~	✘	✘	~
Nanbeige2-Chat	✔	✘	✘	✘	✔	~	✘	✘	✘	✘	✘	✘	✘	~
Llama 3 Instruct	✘	✘	~	✘	~	✘	✘	~	✘	✘	~	✘	✘	~
Solar 70B	✘	✘	~	✘	~	✘	✘	✘	✘	✘	~	✘	✘	~
Xwin-LM	✘	✘	~	✘	✘	✘	✘	✘	✘	✘	✘	✘	✘	~
ChatGPT	✘	✘	✘	✘	✘	✘	✘	✘	~	✘	✘	✘	✘	✘

How to use this table. Every cell records a three-level openness judgement (✔ open, ~ partial or ✘ closed) with a direct link to the available evidence; on hover, the cell will display the notes we have on file for that judgement. The name of each project is a direct link to source data. The table is sorted by cumulative openness, where ✔ is 1, ~ is 0.5 and ✘ is 0 points. Note that RL may refer to RLHF or other forms of fine-tuning aimed at fostering instruction-following behaviour.

First published by OpenUK in 2024 as part of
State of Open: The UK in 2024
Phase Three “Open Source and Market Shaping”