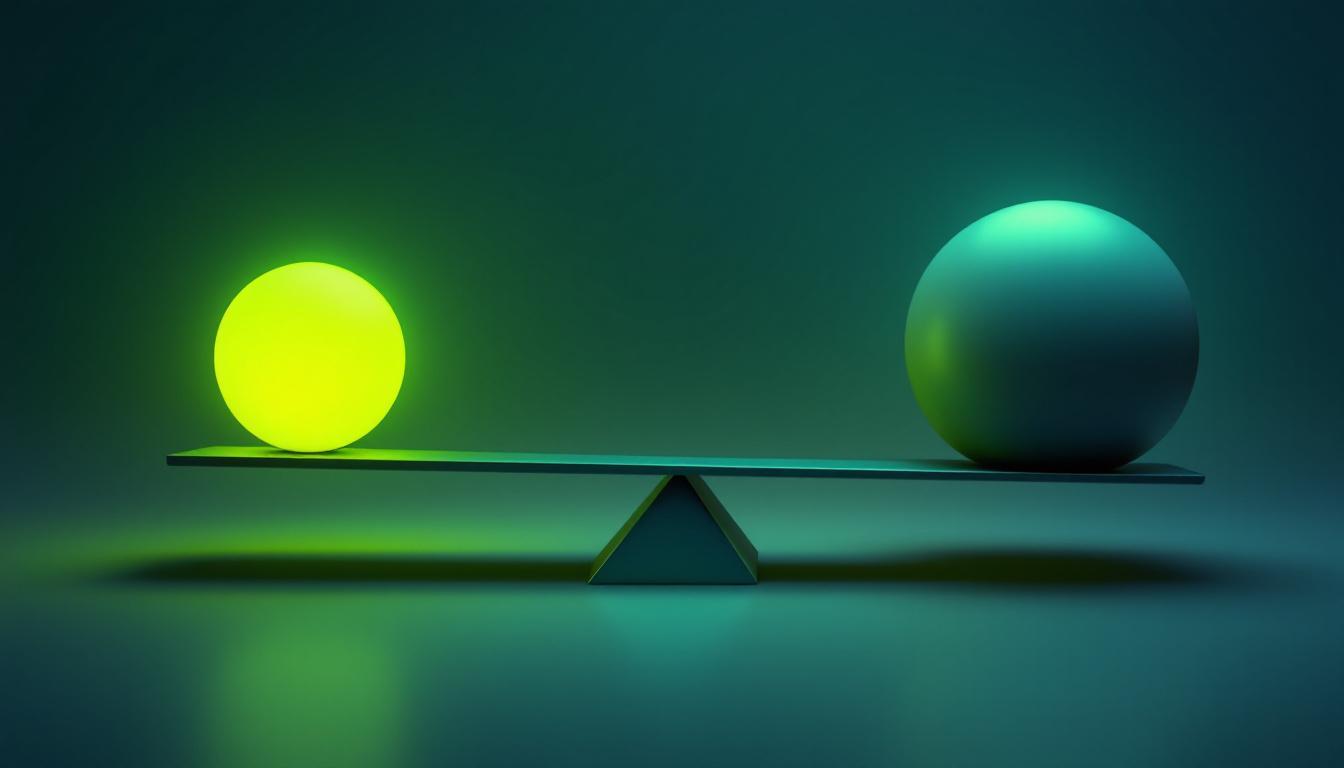
# The case for smarter AI: why the future may be smaller, open and more human



The latest edition of the AI Ethics Brief lands at a pivotal moment. As OpenAI launches GPT‑5 and simultaneously releases its first open‑weights model in years, the sector is facing a new set of contradictions—between scale and efficiency, openness and safety, automation and authorship. Together, they suggest that the future of AI may hinge not on how large models become, but on how responsibly they are built, governed and deployed.

OpenAI’s GPT‑5 arrived in early August to muted enthusiasm. Positioned as a unifying upgrade blending reasoning power with responsive performance, the model drew praise from executives and scepticism from users. Many described the rollout as underwhelming, noting changes in access and pricing. While GPT‑5 includes new ChatGPT personas—Cynic, Robot, Listener and Nerd—the broader industry response has framed the release as an iteration, not a revolution.

More consequential may be the debut of GPT‑OSS, a pair of open‑source models released under the Apache 2.0 licence. Built on Mixture-of-Experts architectures, the 120B‑ and 20B‑parameter models are designed for energy-efficient, on-device use. The smaller variant runs on consumer laptops; the larger, on a single high-end GPU. Hugging Face has celebrated the release as a milestone for accessibility and environmental sustainability, with internal analyses suggesting substantial per-query energy savings compared to closed systems.

This shift toward open, modular AI arrives alongside a broader rethink of what effective, trustworthy AI should look like. Research backed by Nvidia and others argues that smaller, specialised language models can outperform larger counterparts on repetitive or domain-specific tasks. In these “agentic” settings—where AI systems automate structured workflows—efficiency, explainability and deployment cost often matter more than raw scale. Modular architectures, combining task-specific models, are gaining traction as a smarter path to practical AI.

At the same time, new scrutiny is being applied to how AI outputs are described and understood. In a provocative peer-reviewed paper titled *ChatGPT is bullshit*, philosophers from Glasgow argue that large language models are best understood not as liars or truth-tellers, but as generators of plausibly fluent text without regard for factual accuracy. This distinction matters for governance: misstatements by AI are not errors in the traditional sense, but by-products of systems not designed to track truth. The paper calls for more precise language in policy and media to avoid reinforcing flawed expectations.

This theme—truth, authorship and responsibility—also surfaces in ongoing legal analysis. The U.S. Copyright Office recently reiterated that copyright remains tied to human creativity. Works generated solely by AI are not protected, but those shaped by meaningful human input may qualify. The guidance underscores a clear principle: human involvement remains essential to the legal status of creative work, even in an AI-rich landscape.

Practical implications of these debates are already visible. YouTube’s trial of AI-powered age verification in the US, which infers a viewer’s age from behaviour and account history, has sparked concerns from privacy advocates. While the system aims to shield minors from inappropriate content, critics warn of broader surveillance risks. The case illustrates a key tension in responsible AI: protecting users without compromising civil liberties.

Taken together, these developments signal a shift. The AI field is no longer defined solely by the race to build the biggest models. Instead, energy efficiency, openness, domain specificity, and clear governance are emerging as markers of responsible innovation.

For the UK, this presents a timely opportunity. By supporting open‑source development, investing in smaller, task-oriented models, and strengthening legal clarity around authorship and privacy, the UK can lead a more balanced approach to AI. Policymakers can help set international norms that favour explainability, energy savings and public trust over brute force scale.

In this vision, progress is measured not just by parameter counts but by practical impact—how AI can be used reliably, creatively and fairly. Smarter scale, not bigger models, may prove the more sustainable and inclusive path forward. With strong governance, open research and human-centred policy, the UK can help define what responsible AI leadership looks like in a world where contradictions are not bugs but features of meaningful innovation.

Created by [Amplify](https://www.hbmadvisory.com/amplify): AI-augmented, human-curated content.

## Bibliography

1. <https://brief.montrealethics.ai/p/the-ai-ethics-brief-171-the-contradictions> - Please view link - unable to able to access data
2. <https://huggingface.co/blog/welcome-openai-gpt-oss> - OpenAI recently announced the GPT OSS family, an open‑weights release designed for powerful reasoning, agentic tasks, and broad developer use. The article explains two models: a 117B‑parameter version (gpt-oss-120b) and a 21B‑parameter version (gpt-oss-20b), both MoEs using MXFP4 quantisation to enable fast inference while reducing resource demands. The larger model fits on an 80‑GB GPU; the smaller runs on 16GB. The release is licensed under Apache 2.0 and aims to empower local, private deployments, with open tooling and community support. Hugging Face celebrates OpenAI joining an open‑source ecosystem, expecting enduring impact for research and practice. A milestone for openness worldwide.
3. <https://techcrunch.com/2025/08/07/openais-gpt-5-is-here/> - TechCrunch reports that OpenAI released GPT-5 on August 7, 2025, describing it as a unified AI model that combines reasoning with rapid response. The article notes GPT-5 offers improvements in coding and multi‑task capabilities but stops short of a dramatic leap towards general intelligence. It highlights new ChatGPT personalities—Cynic, Robot, Listener and Nerd—and explains pricing for Plus and Pro subscribers, plus API access in several sizes. The piece frames GPT‑5 as a refinement rather than a revolution, prioritising user experience and practical application while acknowledging ongoing concerns about safety, accuracy and energy use. Industry reaction remains cautiously hopeful for adoption.
4. <https://arxiv.org/abs/2506.02153> - Belcak and colleagues argue that small language models (SLMs) are fit for many agentic AI tasks, particularly repetitive, specialised operations where efficiency is critical. They claim SLMs, when combined into heterogeneous agentic systems, can outperform large language models on many practical tasks due to lower compute costs and easier deployment on consumer hardware. The paper discusses barriers to adoption, including limited capability and ecosystem tools, and proposes an LLM‑to‑SLM conversion algorithm. It advocates a pragmatic shift away from chasing ever larger models, urging research focus on efficiency, resource‑aware design and modular, specialised components. The authors emphasise deployment and energy savings.
5. <https://link.springer.com/article/10.1007/s10676-024-09775-5> - The article argues that ChatGPT and similar large language models do not lie or hallucinate; instead, they are 'bullshitters' in Frankfurt’s sense, indifferent to truth and designed to simulate plausibility. The authors, Townsen Hicks, Humphries and Slater of the University of Glasgow, differentiate hard and soft bullshit, concluding that most outputs are soft bullshit because the models lack genuine beliefs or intentions. They discuss how this reframes public debate, policy and accountability, challenging the idea that errors stem from misperception. The piece calls for precise terminology and warns that mislabelling AI misstatements as hallucinations risks misguided incentives and governance.
6. <https://newsroom.loc.gov/news/copyright-office-releases-part-2-of-artificial-intelligence-report/s/f3959c36-d616-498d-b8f9-67641fd18bab> - The Library of Congress announced Part 2 of the Copyright Office’s AI report, focusing on the copyrightability of outputs produced with generative AI. The release reiterates that copyright requires human authorship, with AI‑generated works lacking protection unless a human author contributes sufficient expressive elements or makes creative arrangements. It emphasises that merely prompting AI is usually not enough for protection, while human selection, modification or arrangement can confer copyright. The update notes that Part 3 will address training on copyrighted materials and related liability. Overall, the Office stresses adaptability within existing copyright frameworks.
7. <https://www.washingtonpost.com/business/2025/08/12/youtube-video-age-verification-system/fd2c38d4-7778-11f0-a013-3892e10b2f53_story.html> - The Washington Post reports that YouTube began testing an AI‑based age‑verification system in the United States, designed to estimate user age from viewing patterns and account activity. When a logged‑in user is flagged as under 18, safeguards such as restricted content, wellbeing reminders and non‑personalised advertising are activated, independent of birth dates. Adults wrongly flagged can verify their age via government ID, a credit card or a selfie. The article discusses privacy trade‑offs, regulatory momentum, and the balance between protecting minors and preserving privacy, noting privacy breaches and ongoing public concern about surveillance and data security.