

Dataetisk forskning, analyser og aktiviteter: april 2025



ChatGPT image prompt: "This lively scene captures a Workers' Day celebration in a park where four workers sit on the grass, enjoying beers and chatting while a robot collects trash nearby. With a politician addressing an engaged crowd against a backdrop of red flags and a "WORKERS' DAY" banner, the impressionistic painting blends warm hues of reds and oranges with the cool greens and blues of the park, creating a vibrant and balanced composition filled with subtle detail."

Tre højdepunkter fra denne måned

- [En ny forskningsartikel](#) udforsker udfordringerne med at bygge bro over den faglige kløft mellem tech-eksperter på den ene side og etikere på den anden side.
- Google DeepMind har udgivet [en rapport](#) om, hvordan man kan nærme sig generel kunstig intelligens på ansvarlig vis.
- En [ny artikel](#) diskuterer problemer med almindelige metoder til at måle kvaliteten af generativ AI.

Forskningsartikler

Lamberti, P. M., Bombaerts, G. & Ijsselstein, W. (2024). "Mind the gap: bridging the divide between computer scientists and ethicists in shaping moral machines." *Ethics and Information Technology* **27**(1): 2 DOI: [10.1007/s10676-024-09806-1](https://doi.org/10.1007/s10676-024-09806-1).

Lazar, S. (2025). "Governing the Algorithmic City." *Philosophy & Public Affairs* **53**(2): 102-168 DOI: <https://doi.org/10.1111/papa.12279>.

Mickens, J., Radway, S. & Netravali, R. (2025). "Guillotine: Hypervisors for Isolating Malicious AIs." *arXiv preprint* [arXiv:2504.15499](https://arxiv.org/abs/2504.15499).

Veluwenkamp, H. (2025). "What responsibility gaps are and what they should be." *Ethics and Information Technology* **27**(1): 14 DOI: [10.1007/s10676-025-09823-8](https://doi.org/10.1007/s10676-025-09823-8).

Analyser, værktøjer & debat

[Advanced AI video generation may lead to a new era of dangerous deepfakes](#) (Ada Lovelace Institute)

[Affirming the Scientific Consensus on Bias and Discrimination in AI](#) (Open letter)

[AI Liability Along the Value Chain](#) (Mozilla)

[An Autonomy-Based Classification – AI Agents, Liability and Lessons from the Automated Vehicles Act](#) (Interface)

[Auditing the use of AI to assess risk. The case of Viogén](#) (Eticas Foundation)

[Crowdsourced AI benchmarks have serious flaws, some experts say](#) (TechCrunch)

[Europe's Tech Sovereignty Demands More Than Competitiveness](#) (Project Syndicate)

[Human Rights are Universal, Not Optional: Don't Undermine the EU AI Act with a Faulty Code of Practice](#) (Tech Policy Press)

[Taking a responsible path to AGI](#) (Google DeepMind)

[The Shocking Far-Right Agenda Behind the Facial Recognition Tech Used by ICE and the FBI](#) (Mother Jones)

[What we learned from reading ~100 AI safety evaluations](#) (AI Policy Perspectives)

[What Will Remain for People to Do?](#) (Knight First Amendment Institute)

Begivenheder

-