

hypertrophic AI – can libraries lead the pushback?

hypertrophic (adj.):
abnormally enlarged or overgrown

Dr. Argie Kasprzik
ZBW – Leibniz Information Centre for Economics
Focus on Open Science, Budapest, November 6th, 2025

LAI (*large-scale AI*) comes at a **price**. **Always**.

We often do not realize that we are paying it.

Also, the biggest price is not paid by us but by **others**.

Do we really want that?

costs for humanity: LAI accelerates climate change



electricity



air

water



both the training and
the running of LAI uses up
an amount of resources
which often exceeds
that of small states,
and the trend goes upwards

electronic waste



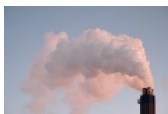
costs for humanity: some examples for resource use

training GPT-4: 50 GWh – 200 flights NY-SFO – 50x more than GPT-3

training GPT-3: 5,4 million liters of water;



LLaMa-3: 22 million liters – as much as needed for farming 2 tons of rice



prognosis 2030: LAI will make up ~21% of global electricity use
and 4x to 6x the water consumption of Denmark

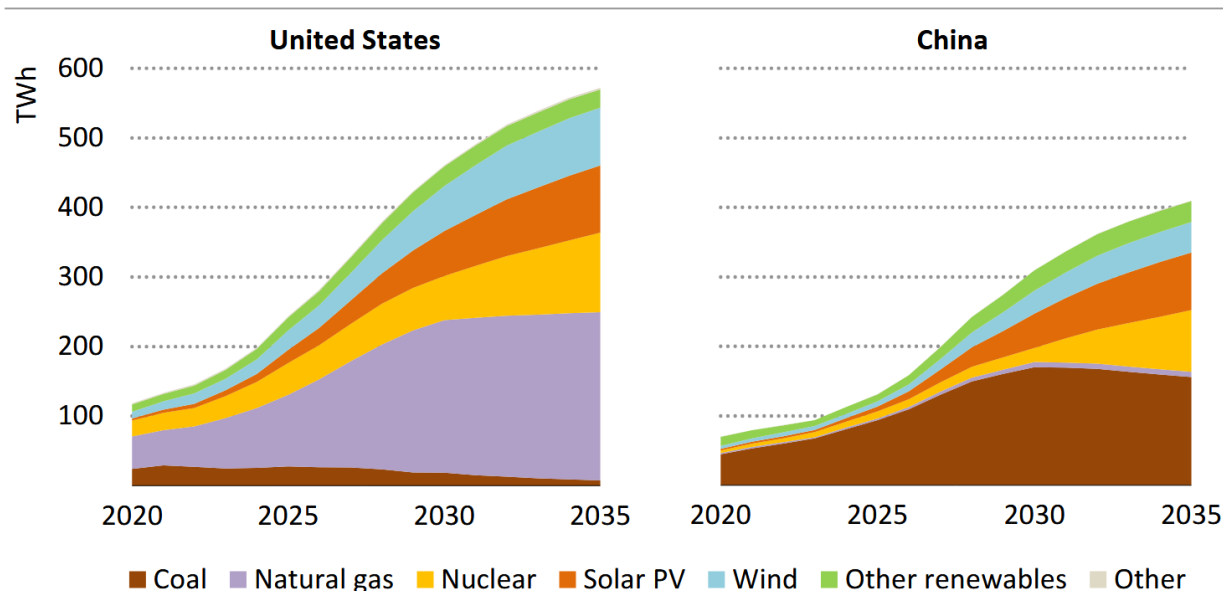
2030, example USA: 4x more deaths than 2023 due to LAI-related air pollution;
60.000 more cases of asthma (predominantly in poorer communities)



2030: up to 2,5 million tons of electronic waste more ~ one smartphone per human



Figure 2.21 ► Electricity generation for data centres in the United States and China in the Base Case, 2020-2035

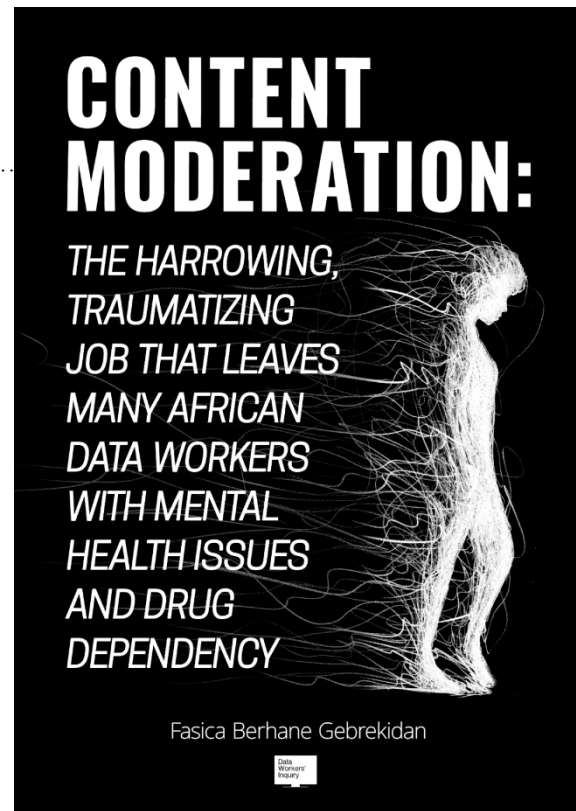


IEA. CC BY 4.0.

Natural gas is set to continue to dominate the near-term data centre electricity supply in the United States, with coal predominant in China

costs for humanity: exploitation

- Large AI continues to depend substantially on **humans** sorting, cleaning, and **annotating** data
- this kind of work is often **outsourced** by big tech to **less privileged** countries (e.g. Kenya, Argentina), under **exploitative conditions** – **no safety at work**, hourly wages of less than 2 USD
- many people doing that kind of work are **suffering** from health problems and post-traumatic stress disorder (PTSD)



costs for humanity: gatekeeping and undue influence

knowledge is power – **asymmetry gets worse**

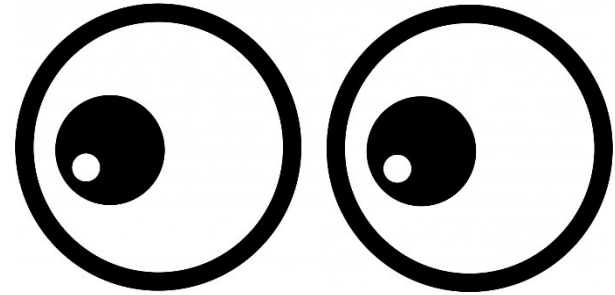
- paid-subscription ("premium") plans → access to (allegedly) high-quality **information and tools becomes a luxury**
- if you do not have the financial means **you pay with your data / privacy**
- big tech build monopolies and act as **gatekeepers** – control who gets access and exploit downstream actors
- big tech obtain a **potential influence** that they should not have **in use cases that lie in the public interest** (e.g., research, education, infrastructure)



costs for humanity: privacy, infrastructure – two examples

often unsolicited, invasive **use of data without respect for** personal privacy, copyright, terms of licences, provisions to exclude AI-related traffic

- example: experiments by OpenAI / researchers of University of Zürich in online community (Reddit)
- example: strain on Open Access infrastructure by crawling / bots

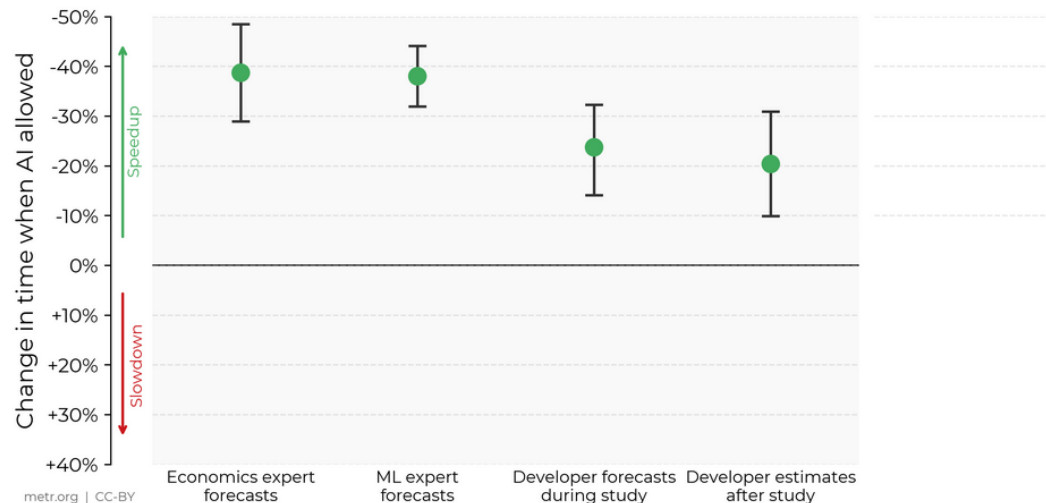


costs for humanity: change management

- reports from the community suggest that more and more managers actually enforce the use of AI tools
- potentially negative impact of misguided change management on productivity (!), psyche, and motivation of staff ↓ →

Against Expert Forecasts and Developer Self-Reports, Early-2025 AI Slows Down Experienced Open-Source Developers

In this RCT, 16 developers with moderate AI experience complete 246 tasks in large and complex projects on which they have an average of 5 years of prior experience.



Tech Industry 1w · Follow
Uber · vibesuxx

I hate vibecoding, change my mind

Vibecoding sounded cool in the beginning but now it's driving me crazy. The junior on my team keeps churning out AI-generated CRs like a machine.

I waste hours reviewing the same lazy, copy-paste AI junk. Pretty sure they're pasting code straight from ChatGPT or whatever, tossing feedback back into it, and resubmitting without even bothering to look.

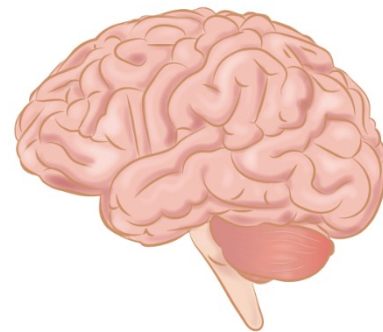
Meanwhile, my manager won't shut up about AI adoption, so now it's basically a free pass to flood the repo with garbage. No care for the codebase, zero respect for anyone's time.

Anyone else stuck in this mess? I'm seriously done with this crap.



costs for humanity: deskilling

first (non-representative) studies suggest that there is a
"cognitive debt" to pay if humans **overrely on AI assistance**



- **essay writing task** (Kosmyna et al. 2025): 3 groups (LLM, search engine, brain-only); LLM group had neural connectivity reduced by 55%, reduced neural integration later on, **could hardly summarize their own essays**; positive metacognitive effects for brain-only group when using LLMs in a second session
- **cancer detection** (Budzýn et. al 2025): decrease in non-AI-assisted detection rate of colon cancer **even for experienced doctors** when having used AI assistance for about 3 months

costs for humanity: information (service) quality

- many applications have **not yet reached enough maturity** – they suffer from bias, unilateralism, political influences, logical inconsistency, no grounding in established/documentated facts
 - alleged **time saved vs. time spent** (by others) finding subtle mistakes (e.g., fact-checking, debugging code, etc.)
- tech firms often push their solutions into their services anyway
 - consequences: **AI slop** (low-quality content), **enshittification** (vendor lock-in, then degradation of service)
 - another potential consequence: MAD (model autophagy disorder)



towards an alternative approach

Civilized AI

1. “use AI to make the world better”
 2. “use the smallest AI that works”
(automation)
 3. “don’t depend on corporate AI”
 4. “evaluate & create data sets”
 5. “be open and transparent”
- perfect task for libraries!*





Timnit Gebru (she/her)

@timnitGebru@dair-community.social

2h

Friends, for something to be open source, we need to see

1. The data it was trained and evaluated on
2. The code
3. The model architecture
4. The model weights.

DeepSeek only gives 3, 4. And I'll see the day that anyone gives us #1 without being forced to do so, because all of them are stealing data.

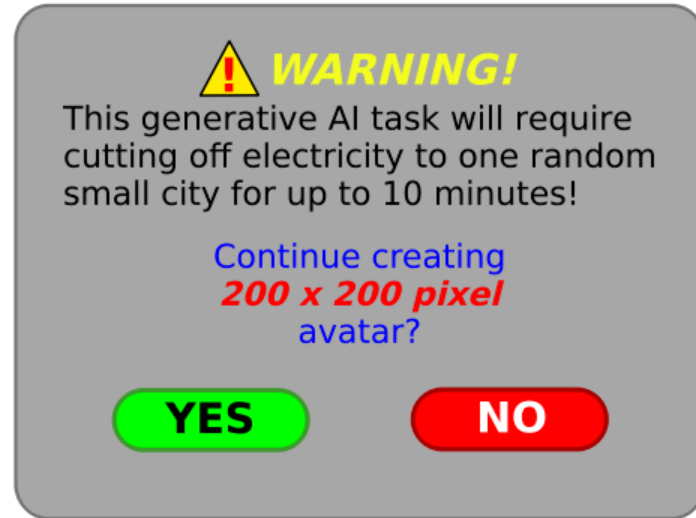
screenshot: <https://dair-community.social/@timnitGebru/113909880610412733>

greenwashing / open weight vs. REAL open source

- companies like OpenAI and Meta regularly **claim to** have published "open source" models (LLaMa, GPT-OSS)
- but are **called out** by the **open source community** – those are only **open weight**
- ... there are actually not many models which are fully open source
- however: **SwissAI with new LLM Apertus** comes close!
 - trained on 15 trillion tokens across more than 1,000 languages
 - 40% of the data is non-English
 - "the entire development process, including its architecture, model weights, and training data and recipes, is openly accessible and fully documented"



What can I do as an individual?



What can I do as an individual?

push back against hypertrophic AI!



- **less is more** – is this particular use necessary?
do I even need an AI-based tool for that?
- **deliberate use of AI-based applications** – using commercial tools
sends a signal to tech firms that further inflating their services is worth it
- **educate and raise awareness** in others – including your leaders!
- look out for **your own rights** – do I really want to
 - use built-in AI tools? if not, **switch them off / opt out** wherever possible!
 - relinquish my input (written, audio, video) to an AI-based tool
e.g. in video calls, for a summary? if not, **object / opt out!**



What can we do as institutions?

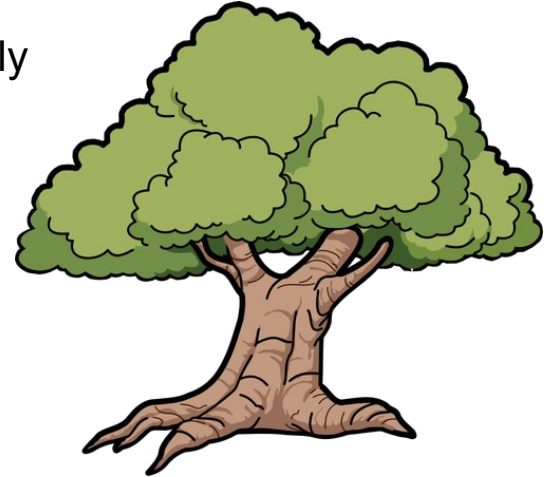
- raise **awareness**, educate (and act accordingly)
- mindful **change management**, use **human expertise** strategically
- activism!
 - push towards **openness** (Open Data, Open Source)
 - push towards **improvement of legal situation**

context: we live in a world in which

- people (have to) generate content **in order to make a living**
- enterprises curate information and develop services **in order to make a profit**

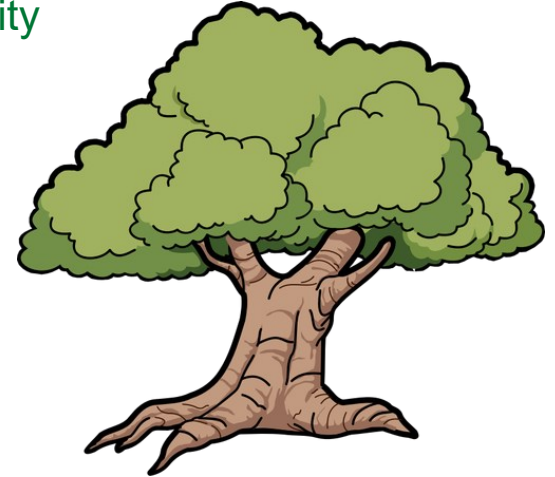
and not for the **benefit of all / the public good** ; point of view: content should be **open and free to use for all** –

including as training data – but **ONLY for non-commercial players!** → **fair use** for purposes of education and research

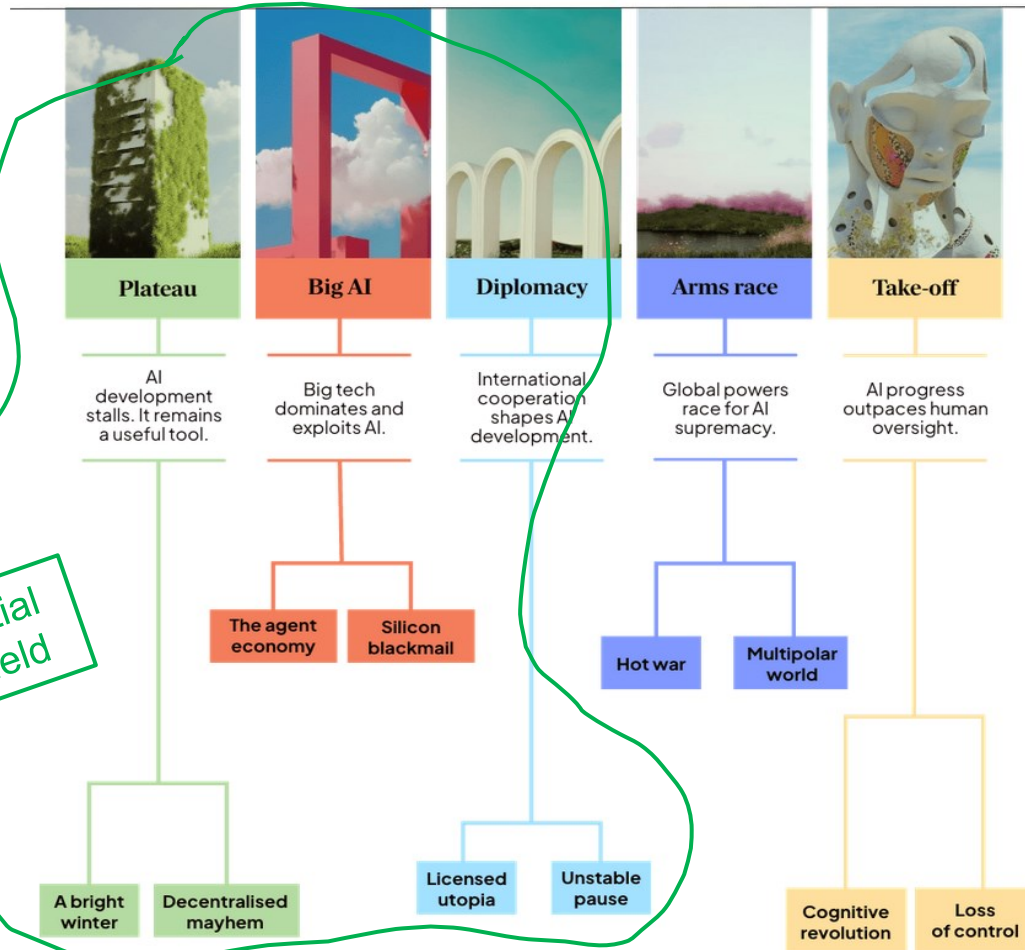


What can we do as institutions?

- cooperation!
 - "evaluate & create data sets" !
 - reuse models / services that use as few resources as possible, are run locally if possible, and preserve privacy / data security
- for centres of competence:
 - develop models / services that
 - use as few resources as possible,
 - can be run locally,
 - and in a way that preserves privacy / data security and make that process transparent for reusability
 - inject human expertise at strategic points
 - optimize own software / hardware for carbon footprint



Five Futures for AI: 2025 - 2032



Message



Thank you!

contact: a.kasprzik@zbw.eu

what my team does with AI at ZBW:

<https://www.zbw.eu/en/about-us/knowledge-organisation/automation-of-subject-indexing-using-methods-from-artificial-intelligence>

link list see next four slides ↓

screenshot:
<https://mastodon.online/@streetartutopia/115186042059851789>



Leibniz-Informationszentrum
Wirtschaft
Leibniz Information Centre
for Economics



link list (last checked 2025-11-03)

overview / general / energy

- <https://savethe.ai>
- <https://heated.world/p/ai-is-guzzling-gas>
- <https://www.iea.org/reports/energy-and-ai>
- https://ai-impact-risk.com/ai_energy_water_impact.html
- in "An MIT Exploration of Generative AI": <https://mit-genai.pubpub.org/pub/8ulgrckc/release/2>
- MIT Technology Review: <https://www.technologyreview.com/supertopic/ai-energy-package/>
- IFLA AI Entry Point for Libraries (incl. risks): <https://repository.ifla.org/items/f197f327-dc49-4743-bb57-0a373505da8b>
- link list about drawbacks of all kinds: <https://fmjansen.com/posts/case-for-ai/>
- <https://arstechnica.com/information-technology/2025/07/ai-in-wyoming-may-soon-use-more-electricity-than-states-human-residents/>
- https://www.linkedin.com/posts/bgamazay_openai-has-announced-o3-which-appears-to-activity-7276250095019335680-sVbW

water additional links

- <https://grist.org/technology/amazon-data-centers-water-positive-energy/>
- <https://www.tomshardware.com/tech-industry/artificial-intelligence/using-gpt-4-to-generate-100-words-consumes-up-to-3-bottles-of-water-ai-data-centers-also-raise-power-and-water-bills-for-nearby-residents>

air pollution and electronic waste additional links

- <https://www.theguardian.com/technology/2024/sep/15/data-center-gas-emissions-tech>
- <https://www.technologyreview.com/2025/05/20/116272/ai-natural-gas-data-centers-energy-power-plants/>
- <https://www.sciencealert.com/scientists-predict-ai-to-generate-millions-of-tons-of-e-waste>

link list (last checked 2025-11-03)

reactions of big tech / greenwashing

- <https://grist.org/accountability/microsoft-employees-spent-years-fighting-the-tech-giants-oil-ties-now-theyre-speaking-out/>
- <https://www.theatlantic.com/technology/archive/2024/09/microsoft-ai-oil-contracts/679804/>
- <https://www.scientificamerican.com/article/what-do-googles-ai-answers-cost-the-environment/>

annotation by humans / exploitation

- Humans in the loop: Labour in the AI supply chain <https://www.youtube.com/watch?v=GEuy4USHMGY>
- <https://data-workers.org/Fasica/>
- <https://www.theguardian.com/media/2024/dec/18/kenya-facebook-moderators-sue-after-diagnoses-of-severe-ptsd>
- <https://www.ihrb.org/latest/content-moderation-is-a-new-factory-floor-of-exploitation-labour-protections-must-catch-up>

privacy / other legal and ethical aspects

- <https://desfontain.es/blog/privacy-in-ai.html>
- <https://static1.squarespace.com/static/5e449c8c3ef68d752f3e70dc/t/6710039559ef840f59365bc8/1729102742546/Stopping+Big+Tech+from+Becoming+Big+AI.pdf>
- <https://opentools.ai/news/openai-turns-rchangemyview-into-ais-persuasion-playground>
- <https://retractionwatch.com/2025/04/29/ethics-committee-ai-llm-reddit-changemyview-university-zurich/>
- <https://arstechnica.com/ai/2025/03/devs-say-ai-crawlers-dominate-traffic-forcing-blocks-on-entire-countries/>
- <https://coar-repositories.org/news/updates/open-repositories-are-being-profoundly-impacted-by-ai-bots-and-other-crawlers-results-of-a-coar-survey/>

link list (last checked 2025-11-03)

degradation of information landscape and deskilling

- https://en.wikipedia.org/wiki/AI_slop ; <https://en.wikipedia.org/wiki/Enshittification>
- <https://edition.cnn.com/2025/01/13/business/enshittification-internet-meta-nightcap/index.html>
- "Self-Consuming Generative Models Go MAD": <https://arxiv.org/pdf/2307.01850> ; <https://news.rice.edu/news/2024/breaking-mad-generative-ai-could-break-internet>
- <https://link.springer.com/article/10.1007/s00146-025-02199-9>
- <https://www.forbes.com/sites/bernardmarr/2023/05/16/the-danger-of-ai-content-farms/>
- <https://www.wired.com/story/linkedin-ai-generated-influencers/>
- two studies about "cognitive debt": <https://arxiv.org/abs/2506.08872> ; [https://doi.org/10.1016/S2468-1253\(25\)00133-5](https://doi.org/10.1016/S2468-1253(25)00133-5)
- <https://www.arthurperret.fr/blog/2024-11-14-student-guide-not-writing-with-chatgpt.html>
- <https://open.devinci.fr/ressource/etude-2024-impact-ia-generatives-etudiants/>
- https://advait.org/files/lee_2025_ai_critical_thinking_survey.pdf
- <https://www.theverge.com/openai/686748/chatgpt-linguistic-impact-common-word-usage>
- <https://arstechnica.com/ai/2025/04/researchers-find-ai-is-pretty-bad-at-debugging-but-theyre-working-on-it/>
- <https://metr.org/blog/2025-07-10-early-2025-ai-experienced-os-dev-study>
- "Google's AI is Destroying Search, the Internet, and Your Brain" <https://archive.ph/tNSO4>

link list (last checked 2025-11-03)

alternative approaches

- <https://hidde.blog/ethical-ai/>
- Suominen, Osmo. Building Civilized AI. AI Sauna, 6 May 2024. <https://www.youtube.com/watch?v=oT8FP1JH5vE&t=1322s> ; https://docs.google.com/presentation/d/e/2PACX-1vRA1o11pODoJ0FmFc8dRj-xNZRU7IsxzDACkiYt6d-Bdfql1ujw3GpSedTQnXDG0MrRg3_WAI1GQS/pub
- <https://greenscreen.network/en/blog/within-bounds-limiting-ai-environmental-impact/>
- <https://www.thegreenwebfoundation.org/publications/report-ai-environmental-impact/>
- <https://www.sciencedirect.com/science/article/pii/S2666498424001340>
- <https://medium.com/@nagidraavid/small-language-models-vs-large-language-models-understanding-the-differences-and-implementations-fc91ff208541>
- <https://tenbluelinks.org/>

openness / fair use

- <https://opensource.org/ai/open-source-ai-definition>
- <https://www.forkable.io/p/metaspas-new-llama-4-ai-models-arent#%C2%A7openness-is-binary>
- <https://shujisado.org/2025/01/27/why-is-the-llama-license-not-open-source/>
- <https://shujisado.org/2025/01/27/significant-risks-in-using-ai-models-governed-by-the-llama-license/>
- <https://blog.archive.org/2025/05/14/protect-fair-use-especially-now/>
- <https://osai-index.eu/guides/open-llms-education>
- <https://www.swiss-ai.org/apertus>

general (digital carbon footprint, independence from corporate providers)

- <https://www.digitalcleanupdate.org/>
- <https://codecarbon.io/>