The Ethics of Generative AI

Should We Use and Promote Generative AI?

Johannes Link, HeiGIT gGmbH, July 17, 2024

Initial Assumption: Ethics Matter

- We have to consider ethics in our work
  - If we - as a charitable, non-profit organization - do not, who will?
  - Our decisions and voices make a difference
- Ethics are never efficient nor do they increase productivity
- Ethics always means balancing of interests
- Ethics are usually not objective

What are the **benefits**? What are the **costs**? And **to whom**?

It’s about Generative AI

… not about ML in General

- ML can be very useful for classification, extra- and interpolation, prediction, simulation etc.
  - ML is often superior to classical algorithms
  - ML also comes with downsides and dangers
  - There’s no clear boundary between ML and Gen AI:
    Deep Learning, Transformers, Embeddings, Vector DBs are used on both sides.

- Gen AI (LLMs, stable diffusion, etc.) is a different beast:
  - Resources Usage
  - (Potential) Collateral Damages
  - (Unfounded) Promises of Value

(Potential) Benefits of LLMs - 1

The stuff that (mostly) works

- Text Classification and Information Extraction: **LLMs work noticeably better than classical approaches**
  - As long as the tasks do not require too much local context
- Automatic Translation
  - If you don’t know the target language you cannot recognize LLMS’ hallucinations.
  - Specialized Models (e.g. DeepL) are considerably better than generic ones

(Potential) Benefits of LLMs - 2

The bold claims

- Replace Human Reasoning and Planning

  „dramatic breakdown of function and reasoning capabilities of state-of-the-art models trained at the largest available scales“
  from recent paper **Alice in Wonderland…**

- Revolutionize Medicine, Engineering etc.
  - The really useful applications are plain ML and pattern recognition systems that support experts in their field

- Take over routine work
  - Full automation is rarely possible due to intolerable error rates
  - Therefore, AI-supported workflows **need a human in the loop**.
    The human now has a dull control task instead of doing the real thing.

(Potential) Benefits of LLMs - 3

Support our creative efforts

- Improving texts. **Cheap proofreading**.
  - Created texts are average at best. And sometimes wrong.
  - Good proofreading tools exist that do not use LLMs.
- Creating Images and Graphics
  - Allows superficially good-looking illustrations if you cannot draw and cannot afford someone who can
  - Adding **shallow images** to presentations and text is not an improvement
- Code generation
  - Makes me faster, if - and only if - I'm knowledgeable enough to check if the output is correct.
  - Generated code is average at best, but full of security holes.
  - **Writing code is not the bottleneck** of software development
- Generate boilerplate texts - Often good enough, but **should we have those texts in the first place?**
- „Democratizing Creativity"
  - Making art is already a fundamentally democratic process. Anyone can do it!
  - **Democratization** is a term to justify that Gen AI threatens the income of many creative workers

(Potential) Benefits of LLMs - 4

Hidden Drawbacks

- Support Teaching and Learning
  - Explain complex stuff to students
    - LLM explanations are often shallow and sometimes wrong
  - Explain software and other complicated systems
    - You'll have to check all explanations for correctness and sufficient depth
  - 2nd order effect: LLMs take away learning opportunities.
    Where do the experts of tomorrow come from?
- Better UIs: Chat bots and spoken language interfaces
  - Natural language is rarely the best interface for a tool
  - The additional degrees of freedom of free text can distract and confuse users
  - Often an improved GUI can eliminate the need for NL input

Benefits of LLMs: Conclusion

- Most working applications of LLMs are a mere convenience for experts
- Big gap between actual benefits of GenAI and what the companies promise. The bold claims are mostly marketing.

„they do a poor job of much of what people try to do with them, they can't do the things their creators claim they one day might, and many of the things they are well suited to do may not be altogether that beneficial.“

from article **"AI isn’t useless. But is it worth it?"**

„AI Is Overhyped, Wildly Expensive, and Unreliable. […] but investors may continue to get rich anyway“

from recent **research paper by Goldman Sachs**

„Generative AI is a climate disaster“

- MS, Google, Amazon invest billions every few weeks for new massive **hyper scale data centers**
  - MS, Google, Amazon etc. have **bet their future on Gen AI**
  - Google emissions up by 48% in 5 yrs - originally wanted to be on 0 in 2030
  - MS emissions up 30% since 2020
    - More than 5 GW server capacity, more than Portugal’s electric consumption
  - Shift to renewable energy for data centers => Renewable energy not available for other purposes
  - Jevons Paradox aka **Rebound Effect**: More efficiency leads to more usage instead of savings
- Data centres require water, and are often located in areas with already too little water

  „estimating that AI could account for up to 6.6bn cubic metres of water use by 2027 – nearly two-thirds of England’s annual consumption.“
- Dedicated hardware
  - Specialized AI chips (GPUs) have a life span of about 2 years => Lots of **additional electronic waste**
  - Nvidia is currently the 2nd most valuable company in the world due to its monopoly in GPUs.
    They're also completely intransparent as for their GHG footprint and offer no life cycle assessment (LCA) for their products.

Actual power consumption of Gen AI

- CMU Study (https://arxiv.org/pdf/2311.16863.pdf)
  - Generate a **single image** -> 1 full phone charge
  - 1000 **generated texts** -> A 5th phone charge
  - Most carbon footprint from use, not from training
  - Large models are much more energy-intensive than small models,
    but small models require more training
  - "For very popular models, such as ChatGPT, it could take just a couple of weeks for such a model's usage emissions to exceed its training emissions"
- Alex de Vries (in **„The growing energy footprint of artificial intelligence"**)
  - According to Google search with full AI support would use 10x as much energy

Human Costs of „AI“

Those who benefit are not the ones who pay the price

- **Ghost labour**: work that appears to be performed by a computer, but is actually delegated to underpaid contractors, working in horrible conditions, with few labour protections and no benefits.
  e.g. "OpenAI used Kenyan workers on less than $2/hr to make ChatGPT less toxic“
- Using „AI“ for creating art deprives us of our **most humane activity**: Doing something just because we like it.
- Gen AI removes „the special thing“ from text and art. **Everything becomes alike**.
- Using existing art to train LLMs is a classical **tragedy of the commons** problem:
  Something that’s ok for individuals becomes exploitation when done by corporations.
- Training data is used without the consent of the original authors.
  It is **intellectual property theft** - Legally, however, the jury is still out.
  - It is harming communities of small artists everywhere.
    Now people are afraid to be taken advantage of when they share their work freely.
- **Opportunity costs**: How could we benefit humanity with all the money and endeavour we spent on LLMs?
  - Example: The EU is dramatically cutting funding for FOSS initiatives because "lots of budget are allocated to AI"

More Collateral Damages and Dangers

- **Deepfakes** are used to harass and abuse people
- The training data is unknown (to the public) but it's **conservative and biased**
  - LLMs drown us in superficial, culturally one-side and often wrong „information“
  - Using AI in decision making strengthens existing biases and discrimination
- Gen AI is capital-intensive. **It's making the rich richer**.
  - Automation through Gen AI is about weakening the position of employees
  - Downward spiral of quality to what can just be tolerated
- Lack of **Accountability**
- **Privacy** and Security Hazards

Summary 1

The basics

- The existing value of Gen AI is mostly in **automating and supporting simple tasks**, providing some convenience or speeding up grunt work
- Most of the **promised value** has not been realised - and **cannot be realised** with current technology
- The **costs and dangers** of Gen AI however **are genuine**
  - Environmental
  - Societal & Human

Summary 2

Personal conclusions

- As an ethically acting organisation, we should **no longer promote commercial Gen AI** or try to integrate it into our products.
- Instead, we should use our reach to **call out the costs and damages**.
- Using commercial LLMs is promoting. Let's stop using them!
- **Some of the many problems** are rooted in Big Tech's way of pushing "AI" forward and **could be avoided.**
- **Critical and responsible research** is essential:
  - What are the real (future) benefits?
  - How and to what degree can we minimise damages? How can we get independent from Big Tech?
  - Is the long-term vision worth the costs?

Sources

- https://www.theguardian.com/technology/article/2024/jul/02/google-ai-emissions
- https://spectrum.ieee.org/ai-energy-consumption
- https://www.cell.com/joule/abstract/S2542-4351(23)00365-3
- https://tinyml.substack.com/p/the-carbon-impact-of-large-language
- https://www.algolia.com/blog/ai/the-pros-and-cons-of-ai-language-models/
- https://spectrum.ieee.org/chatgpt-for-coding
- https://www.aisnakeoil.com/p/ai-scaling-myths
- https://www.goldmansachs.com/intelligence/pages/gs-research/gen-ai-too-much-spend-too-little-benefit/report.pdf
- https://tante.cc/2024/06/18/ethische-nutzung-von-ki-beim-tag-der-abschlussarbeit-am-kit/ (German)
- https://arxiv.org/html/2406.02061v1
- https://www.bloodinthemachine.com/p/ai-is-not-democratizing-creativity
- https://www.infoq.com/presentations/responsible-development-ai-hype/
- https://www.youtube.com/watch?v=3LIvHF-IX9Y (German)
- https://www.theregister.com/2024/07/17/foss_funding_vanishes_from_eus/
- https://techwontsave.us/episode/229_generative_ai_is_a_climate_disaster_w_sasha_luccioni