Ethical AI in a Rapidly Evolving Landscape - deep dive TPAC meeting minutes

Slides: Lightning Talk - Ethical AI

Participants

-
- Humera Minhas, eyeo
- Anssi Kostiainen, Intel
- Yen-Lin Huang, ministry of digital affairs, Taiwan
- Cornelius Witt, eyeo
- Joshua Ssengonzi, DefendDefenders/Article19
- Laszlo Gombos, Samsung Internet
- Kris Anne Kinney
- Guy Teller

Meeting minutes

- Deep dive format: first intro; then open format and discussion
- Intro Humera
    - Biking accident last week
    - Question: why did the bike not prevent the accident?
    - E-bike does not stand for e-thical
    - Bigger question about ethical AI and decision making (e.g., when it comes to self-driving vehicles
    - ML guidelines are there and are based on values - https://www.w3.org/TR/webmachinelearning-ethics/
    - Core values that we need in the system
    - Now: making out of theory the actual practice
    - ‘Not wearing a helmut’ is a bit similar to running experiments or developing prototypes (‘just doing it’..)
    - Other guidelines come from ICO (UK data protection regulator)
    - Many open questions remain, like engineering-vs-product problem? Who takes ownership? Open source vs. closed source? How can I make my ML/AI-system more sustainable and unharmful for the environment?
    - One of the many questions we want to answer: where do we start?
    - Experience at eyeo - first lessons learnt: enablement and sharing knowledge, defining ownership and establishing processes (e.g., security/privacy/legal reviews)
- Christos:
    - Starting with values
    - Values that should be included when working with and developing AI
    - In practice, things to do: identify patterns that should be avoided (e.g., biases), identify values you care the most about, explore how your products/solutions are performing against those values.

- Question: Are you thinking or applying any ethic-related values in your product development process in regards to AI? What are we doing now and what are we willing to do? (Add your answers below)
    - Anssi: Adopted legal principles at Intel to operationalize these principles. Also W3C guidance developed.
    - Consultation to clients who want to increase accessibility. Being honest and confronting moral questions as they arise
        - In accessibility, it's ethical and also legal
    - IBM: Prioritizing ethics over revenue. Some use-cases or prototypes might be not acceptable.
        - Also relevance of reputational losses/considerations
    - Privacy of users needs to be safeguarded. Also, preventing biases. E.g., protection certain (vulnerable) groups.
        - Bias - examples also in AI when gen-AI creates images, e.g., doctors are disproportionately white males, etc
    - David Boulton: RE Shifting perspective: Identifying the patterns you want to avoid:
    - Short circuiting learners attention and participation in differentiating and disambiguating their learning needs by AI anticipating their needs and providing them learning resources before they have expressed needs for them. This comes down to valuing their agency and participation (which is more vital to their life in the long run) over their short term performance in specific learning objectives.
    - Data bias - does it originate from already existing biases in team
        - E.g., hiring practices
    - Important to upscaling own skills by talking to and with experts
    - Relevant to also do impact assessments, to get data on the current status quo
        - Reference also to new EU Data Act
    - One concern is that AI does best support what learners need to learn. How about the capacity of AI for what learners need? Otherwise, the learner is disserved in the long run.
    - Security is not a state, it's a process
        - Same for ethics in AI? Needs continuous testing
    - David Singer (Apple): Society at large vs. individuals
        - E.g., personalization of AI systems as "they know" the user ('bubble' problem)
        - E.g., the confidence of certain AI-generated answers (also on questions relating to historic events, that we actually cannot say for sure.
        - Ansii: we should educate people - it's not always a technical solution, we need to educate people.
        - This is something similar to fake news - do you believe it or not? Do you research it? When you write code, you do QA it.

        - Need for validating information before “trusting”
- User approach to LLM need to be clearer, better aligning between the user and the AI system have the same understanding on the required outcome
- Yen-Lin Huang/Ministry of digital affairs, Taiwan: initiating online/onsite deliberating process as alignment assemblies with Collective Intelligence Project for exploring the exact ethic topic that the publics really care. The result will determine the direction of AI Act in Taiwan. Secondly, encourage decentralized/open-source LLMs in Taiwan.
    - Finding out what public thinks and which attitudes citizens have on AI
    - All results will be open-sourced
    - One way to overcome bias could be in decentralized/open-source LLMs
- Explainability of AI is needed, otherwise, it is a black box
- Values and principles:
    - Global north with stronger democracies
    - Other parts of the world with less democratic structures
        - Potential for using AI tools for surveillance or control, etc
        - Need to take these use cases into account when looking at authoritarian regimes
        - Looking at it from ethical AI and principle perspective. But what about organizations that see these values differently etc.? For example, surveillance governments and the use of tools such as Pegasus
- Context hygiene needed and logging integrity needed (something that cannot be tempered)
- Retail perspective:
    - “Data is the new oil” but data is being handed out and shared / collected by others
    - Potential need for some user license agreement, also in terms of protecting vulnerable consumers. However, the problem about monitoring certain clients and what they buy and potentially also monitor “regular” customers
- Delegating to AI for automation for use cases where people do care less about is another risk.
- David: The great opportunity of AI ethics may well be that it demands we reexamine the ethics of politics and business we take for granted.
    - The important takeaway is that the threat is not AI. The threat is AI used to weaponize predation. If we don’t end the commonly accepted predatory ethics running rampant in our politics and capitalism, AI will be used to manipulate and enslave us (in ways many to most will never realize). Again, I am not talking about some Ultron-like artificial being, I am talking about AI serving the

same ethics that abuse us now in our commonplace politics and commerce. Think of the next big iteration of Cambridge Analytica (the tech that powered Brexit and Trump). Think of the next big iteration of AI-individuated data vampires.
https://learningstewards.org/musks-asteroid/