/thoughtworks

Mozilla Study Reveals AI Transparency is Rarely Prioritized Among Tech Builders

March 16, 2023

Berlin

Survey of developers, ML engineers shows limited motivation for prioritizing issues like ethical and legal compliance, despite forthcoming EU AI Act and Digital Services Act

A study investigating tech builders' perspectives on AI transparency found that ethical and legal compliance are rarely motivating agents when designing and deploying AI systems. Over 90% of surveyed respondents ranked ethical guidelines 11th on a 12 point scale.

The study — titled 'AI Transparency in Practice' — was conducted by Mozilla with contributions from Thoughtworks. It surveyed several dozen developers, software engineers, product designers, and machine learning engineers on their motivations, awareness, and challenges regarding AI transparency. Respondents hail from seven countries,[1] and the majority of the 59 participants work in the fields of Natural Language Processing (NLP), Forecasting/Prediction, and Computer Visualization (image, object recognition, facial and biometrics recognition).

Respondents said that although ethical concerns may arise during AI development, there's little incentive to communicate unintended consequences or product limitations to users, such as how these tools could recreate biased and discriminatory results. As a consequence, it largely falls to users to understand how algorithmic-based decisions influence their lives — like why certain content is amplified or banned, or how a medical AI makes a diagnosis.

The survey also found that builders clash with management around implementing transparency measures. In interviews, researchers shared that discussions around AI ethics often aren't allowed to interfere with commercial decisions. Indeed, AI builders ranked "accuracy and target goal achievement" as their first priority, far above "legal compliance" and "compliance with ethical guidelines."

Says Ramak Molavi Vasse'i, AI transparency research lead at Mozilla: "Transparency is not a goal, but a means to achieving trustworthy AI. Yet builders face a significant challenge in providing appropriate explanations to different stakeholders. Meanwhile, our research shows that organizations' current definition of transparency is narrow, failing to take into account the systems' broader societal and environmental impacts."

Says Jesse McCrosky, principal data scientist at Thoughtworks Finland: "The research reveals builders' lack of confidence in how the purported design of AI systems matches the metrics the systems use, which are the heart of what they're truly built to do. About one-third of the respondents doubt whether the metrics they used were appropriate. This helps explain worrying trends in which AI systems discriminate, amplify misinformation and possibly threaten the mental health of their users. Fortunately, responsible design principles and practices can make a difference."

The study also delves into possible solutions for this transparency problem, like regular algorithmic impact assessment audits and the use of well-understood, interpretable-by-design models where explainability is required in the absence of adequate tools.

READ REPORT

More findings from the report:

- Low prioritization of AI transparency. AI transparency is rarely prioritized by the leadership of respondents' organizations, partly due to a lack of pressure to comply with related legislation.

- Major focus on product performance and low business risks. Builders primarily focused on AI systems' accuracy and debugging, rather than providing adequate information on how users are impacted. Ethical discussions were also allowed — if they posed low business risks.

- Minimal communication of unintended consequences. Apart from information on data bias, developers rarely shared information on system design, metrics, or wider impacts on individuals and society. There is limited motivation to provide social and environmental transparency, or proper transparency into unintended consequences or harms.

- Low trust in AI explainability tools. While there is active research around AI explainability (XAI) tools, there are fewer examples of effective deployment and use of such tools, and little confidence in their effectiveness.

- Discrepancies between information shared with stakeholders. Providing appropriate explanations to various stakeholders poses a challenge for developers. There is a noticeable discrepancy between the information survey respondents currently provide and the information they would actually find useful and recommend.

This research is part of a broader transparency project that will examine the specifics of Article 13 of the European Union AI Act and the requirement for AI labeling/declaration (e.g., watermarking of AI-generated content) under Article 52. Learn more around the research findings at the Mozilla Festival virtual session entitled "AI Transparency in Practice Demystified! Ask us anything!" on March 20, 2023.

1 India, England, Scotland, Germany, Netherlands, the U.S. and Cyprus.

- ### -

Press contacts:

Mozilla: Tracy Kariuki: tracy@mozillafoundation.org

Thoughtworks: Linda Horiuchi: linda.horiuchi@thoughtworks.com