Qeios

Peer Review

Review of: "WeAudit: Scaffolding User Auditors and AI Practitioners in Auditing Generative AI"

Fady Albashiti[1]

1. LMU Klinikum, Munich, Germany

The use of AI applications is no longer limited to professional, industrial practitioner and is used by many people in everyday life. This use could certainly be associated with many benefits, but could also cause harms. In order to be able to identify the harms and mitigate or better eliminate them, a process and appropriate tools are required to audit AI applications. This is important in order to create trustworthy, responsible and auditable AI and to raise the human confidence in such systems.

In this article Deng et al. describe the methods and tools they have applied to empower and support user auditors and AI practitioners along the process of auditing Gen AI, in this case the “Stable Diffusion”, an open source Text2Image system.

The main question Deng et. al raising in this work is: “**How might we develop tools and processes to effectively scaffold end-user engagement in AI audits, while ensuring their findings are useful and actionable for AI practitioners?**”

To answer this question, Deng et al. have performed a formative study with 11 end users and seven AI practitioners to identify possible design goals to support user-engaged AI auditing. Then they have developed *WeAudit*, a workflow and web-based system accordingly. A user study with 45 user auditors and a later evaluation with 10 AI practitioners were performed subsequently.

The methods used to conduct the studies are well-known methods from the HCI field (e.g. usability tests, think-aloud method). This work builds on prior research, and theoretical models, frameworks such as models of information foraging and sensemaking, recommender systems and crowdsourcing.

The reported findings provide details how *WeAudit* supports user auditors and their engagement in AI auditing processes (Figure 5, 6) in noticing otherwise overlooked AI harms and reflecting upon them and

 doi.org/10.32388/QVHAW6

in reporting findings in ways that AI practitioners find actionable. Moreover, several opportunities were identified to enhance and further develop *WeAudit* features.

The findings show, that the engagement of user auditors (and also AI practitioners) is crucial and should be a part of the development process of GenAI algorithms and applications. The involvement should be at early stage (where GenAI algorithm and application on the level of Minimal Viable Product) an versions (e.g. as Beta Tester, Fail-Fast/Fail-Often). The cohort of user auditors should be representative regarding e.g. Gender, Nationalities, Occupation. Deng et. al stated that it is planned to evaluate an improved version of *WeAudit* to with a broader population, not only college students. Also, the data used to train, test and validate the GenAI algorithms and applications should be of representative origin to avoid a data-driven harmful AI biases and discrimination. This requirement should be addressed by the developers of GenAI algorithms or applications and AI practitioners and could be part of future research. Deng et. al designed and evaluated *WeAudit* taking the T2I "Stable Diffusion" as a use case. The remaining question in sense of sustainability and reusability of *WeAudit* is, whether the implemented features are general enough or could be generalized to audit further GenAI applications. Further question is, if we could imagine to audit GenAI algorithms and applications with AI tools. This question is triggered by the mentioned "intelligent task routing" (by automatically providing users auditors tasks they are most interested in). This is an interesting approach, but it should be assured, that at the end of the day all the hypothetical/theoretical paths/routs are considered and no AI harms or biases still uncovered or overlooked.

The Article is well structured, the redline is clear and the references between the sections is well positioned and make the reading flow easier, the figures are helpful to understand the *WeAudit* System and workflow. The chosen approaches and design methods are stable and well-known in the HCI field, especially in software development.

In this relatively young field (AI Auditing) this work is of high relevance and a very welcome contribution to highlight the importance of having frameworks and tools to audit AI applications and to consider / include end users /user auditors in this process and make their feedback accessible to AI practitioners to develop trustworthy and responsible AI. AI Auditing shouldn't be considered as a technical matter or limited to industrial professionals or AI practitioners or developers. AI Auditing should be a task for the society as a whole.

Recommended for Publication

Declarations

Potential competing interests: No potential competing interests to declare.