

Collaborative Learning Discussion 1

Codes of Ethics and Professional Conduct

Case Study: Malicious Inputs to Content Filters

Summary Post

This topic has opened up many questions surrounding ethical, professional and legal standards on particular topics.

FCC (2019) has a list of requirements which seem to have been opposed by a group, who had been labeled as activists. Both organizations appeared to have differing ideas on what is and what is not acceptable behavior. Although it's not modern, the Belmont report aims to protect all individuals and groups by ensuring ethical procedures are adhered to (The Belmont report, 1979) and its principles seem to be the foundation that BCS (2022) may have been built on.

Majola (2023) shared a study by Caliskan et al. (2017) which found that machine learning models used in natural language processing tend to reflect biases present in the training data, thereby potentially perpetuating social biases. As a result of the activists' access to the machine learning model and blocker plus deciding to keep the model running, it is possible that the data is also poisoned (Goldblum, et al. 2023) and therefore, the requirements within CIPA (FCC, 2019) are not being adhered to.

Mutegi (2023) also supports this point by stating that “opting to disable the accounts of activist groups while keeping the existing model intact, Blocker Plus's management not only failed to unblock topics outside the scope of The Children's Internet Protection Act, but also exacerbated the issue”.

Nomusa (2023) also stated that “This highlights the need for constant evaluation and scrutiny of the algorithms and datasets used in content filtering systems”, which could help mitigate ethical and professional concerns. It has also become clear that potential ethical and data privacy issues may have taken place if foreign citizens were involved in the US schooling system according to other global data protection regulations (GDPR, 2018; CCPA, 2023; PIPL, 2023). Ibrahim (2023) also highlighted this issue by suggesting that the collection of user input draws attention to the possibility of data misuse and privacy.

References

BCS. (2022) Code of Conduct for BCS Members. Available at: <https://www.bcs.org/media/2211/bcs-code-of-conduct.pdf> [Accessed on 02 July 2023]

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183-186. doi: 10.1126/science.aal4230

CCPA. (2023) California consumer privacy act. Available at: <https://www.oag.ca.gov/privacy/ccpa> [Accessed on 19th June 2023]

FCC. (2019) Children's Internet Protection Act (CIPA). Consumer Guide. Available at: https://www.fcc.gov/sites/default/files/childrens_internet_protection_act_cipa.pdf [Accessed on 02 July 2023]

GDPR. (2018) General data protection regulation. Available at: <https://www.gdpr.org> [Accessed on 19th June 2023]

Goldblum, et al. "Dataset Security for Machine Learning: Data Poisoning, Backdoor Attacks, and Defenses." *IEEE transactions on pattern analysis and machine intelligence* 45.2 (2023): 1–1. Web. Available from: <https://ieeexplore-ieee-org.uniessexlib.idm.oclc.org/document/9743317> [Accessed on 1st July 2023].

Ibrahim, M. (2023) Collaborative discussion 1. Available from: <https://www.my-course.co.uk/mod/forum/discuss.php?d=166942#p264057> [Accessed on 01 July 2023].

Majola, N. (2023) Collaborative discussion 1. Available from: <https://www.my-course.co.uk/mod/forum/discuss.php?d=166942#p264057> [Accessed on 01 July 2023].

Mutegi, B. (2023) Collaborative discussion 1. Available from: <https://www.my-course.co.uk/mod/forum/discuss.php?d=166942#p264057> [Accessed on 01 July 2023].

PIPL. (2023) Personal Information Protection Law. Available at: <https://pipl.com/privacy-center>. [Accessed on 19th June 2023].

The Belmont report. (1979) Ethical Principles and Guidelines for the Protection of Human Subjects of Research. Available from:
<https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html> [Accessed on 01 July 2023.]