Research Methods and Professional Practice January 2022

« Collaborative Learning Discussion 1

Vaibhav Chawla

Initial Post

7 days ago

2 replies

Last now

Case Study: Malicious Inputs to Content Filters

The ever-increasing number of websites offering services, information, and content in the name of entertainment has led to online chaos. This has inevitably put the parents, schools, institutions, organisations in a position to filter the detrimental online content for the end-users.

This case study if one of early revelations of integrating an unmanned autonomous solution to into the fabric of social conduct, which as understood can indirectly manipulate and skew the learning sources.

Blocker Plus is one of the many attempts to filter and restrict access to such content online. (ACM, N.D.) Blocker Plus is an automated solution that relies on artificial intelligence and machine learning algorithms to update, maintain, filter, and debar inappropriate content. These algorithms are trained repetitively (epochs) on huge datasets to achieve adequate accuracy. In the process of training, these algorithms can become vulnerable if they develop an implicit bias. The bias is highly probable if the following aspects are not monitored regularly:

- Versatility of datasets used in training.
- Feedback loop.
- Reinforcement Learning.
- User inputs.

The current case study draws our attention to the importance of the last element in the list 'user inputs', which can be used as optimises to guide these language processing algorithms. The inventors made a design choice while developing this self-governing, self-scaling platform by conferring unfiltered trust in the legitimacy of the users and their inputs. This design choice can be held in violation of the BCS Code of Conduct article 2a(BCS,2021), due to the lack of consideration of security aspects which made the final product gullible and unreliable.

One could argue, individuals who are subjects/victims these solutions are at risk of developing unilateral and polarized learning due to the biases developed by the system. The system if not introduced in a phased, controlled and methodically scrutinised fashion might incapacitate the children to develop a well-informed, unprejudiced outlook for the wide spoken sensitive topics, inflicting far more damage to the children and hence to the future society than the original problem it was intended to solve. The aerial view from the case study reinforces one fundamental principle: These technologies and their surrounding algorithms are just tools *"Garbage in, Garbage out !"*; and demand periodic inspection. Introduction of an open ended feedback cycle for report abuse with no filtering and scrutiny will only digress the final product from the intended use case. The designers should not be allowed dust this case off by resorting to "People's power in people's hands" simply; instead should required to follow laws that enforce logging to avoid repudiation and to facilitate timely analysis of the collected user inputs to gauge behaviour/trend before and after introducing them to the system.

References

ACM (N.D.) Case: Malicious Inputs to Content Filters. Available from: https://ethics.acm.org/code-of-ethics/using-the-code/case-malicious-inputs-to-content-filters/ [Accessed 25 January 2022].

BCS (2021) Code of Conduct for BCS Members. Available from: https://www.bcs.org/media/2211/bcs-code-of-conduct.pdf [Accessed 25 January 2022].

Reply

2 replies

1  Post by **Freya Basey**

6 days ago

Peer Response

Further to Vaibhav's points, unethical implementations of artificial intelligence algorithms can stem from a lack of understanding as well as an innate bias to trust automated decisions (Fry, 2018). For example, after an investigation into an algorithm that was employed to determine disability benefits in the state of Idaho, it was shown that not only was the logic almost random in some cases but the historical data it used was erroneous (Stanley, 2017). Medicaid did not question the outputs of the algorithm and this led to serious impacts on claimants who had their benefits cut. This example shows the ethical importance of scrutinising both the logic and the data inputs used to inform automated decisions. To compound the issue of understanding algorithm logic, there may be certain situations where the logic cannot be followed by humans due to sheer complexity and this could lead to further ethical questions (Fry, 2018).

One of the four pillars of the BCS code of conduct centres around professional competence (BCS, 2021). In this case study, the lack of competence of the engineers and decision-makers involved with Blocker Plus has been exposed in their approach to not only the implementation of machine learning without scrutinising the data inputs, but also the decision to continue to use the same logic and data approach even after exposing the underlying issue (ACM, N.D.). The hope that the algorithm will use new data inputs to in effect correct itself is fundamentally flawed and does not acknowledge the bias that may already have been created.

In conclusion, these case studies expose the importance of ensuring competence in the use of advanced technologies before employing them. This competence should include an understanding of both the limitations and potential ethi-

cal impacts of the technology being used.

References

ACM (N.D.) Case: Malicious Inputs to Content Filters. Available from: **https://ethics.acm.org/code-of-ethics/using-the-code/case-malicious-inputs-to-content-filters/** [Accessed 7 February 2022].

BCS (2021) Code of Conduct for BCS Members. Available from: **https://www.bcs.org/media/2211/bcs-code-of-conduct.pdf** [Accessed 7 February 2022].

Fry, H. (2018) *Hello World: How to be Human in the Age of the Machine*. 1st ed. Transworld Digital.

Stanley, J. (2017) Pitfalls of Artificial Intelligence Decisionmaking Highlighted In Idaho ACLU Case. Available from: **https://www.aclu.org/blog/privacy-technology/pitfalls-artificial-intelligence-decisionmaking-highlighted-idaho-aclu-case** [Accessed 7 February 2022].

Reply

2  Post by **Jan Küfner**

now

peer response

As mentioned by Vaibhav it is not good practice to disregard cyber security concerns in such type of software, but I would even go further here. The company not only violated this item of the BCS Code of Conduct, but they also managed to damage the reputation of the profession, since omitting cyber security completely in a machine learning algorithm that is supposed to keep children safe, is scary to many people, which is likely to be covered in news. By that type of news coverage, it is inevitable that the public might start to review their meaning towards IT professionals in a negative way. (BCS 2022)

References:

BCS Code of Conduct (2022) | BCS. Available at: https://www.bcs.org/membership-and-registrations/become-a-member/bcs-code-of-conduct/. [Accessed January 29, 2022]

Reply Edit Delete

Maximum rating: -

Add your reply

Your subject

Type your post

Dateien auswählen Keine ausgewählt

Submit

Use advanced editor and additional options

OLDER DISCUSSION

Initial Post

NEWER DISCUSSION

Initial Post