Malicious Inputs to Content Filters

I have chosen the case study of Blocker Plus, a US-based content filter which helps prevent underage internet users from accessing certain materials, because I have strong, sometimes conflicting, personal views on the topic of censorship and freedom of information.

Blocker Plus was used to keep publicly accessible computers in schools and libraries safe for children. It maintained a central repository listing material the U.S. Children's Internet Protection Act deemed illegal and unsuitable for children.

The problem with this approach was that it required constant costly maintenance to keep the blacklisted materials up-to-date. As an alternative, they allowed home users to add materials to the list, and designed a machine-learning algorithm to automatically update the blacklist.

Some consumers realised that this could be manipulated and organised groups to add materials that went against their personal beliefs, including topics like vaccination and sexual preference, effectively censoring other users from accessing materials not actual under the jurisdiction of the CIPA.

The company uncovered this, but chose to keep quiet about it, hoping the algorithm would work itself out.

Action	**ACM**	
Didn't fully analyse or predict the risks of using the ML algorithm for censorship	**2.5** Give comprehensive and thorough evaluations of computer systems and their impacts, including analysis of possible risks.	**3.1** carry out your professional responsibilities with due care and diligence in accordance with the relevant authority's requirements while exercising your professional judgement at all times;
Didn't prevent its misuse	**2.9** Design and implement systems that are robustly and usably secure.	
Blocked useful legal information about vaccines from being accessed	**1.2** Avoid harm.	

Discriminated against the gay and lesbian communities	**1.4** Be fair and take action not to discriminate.	**1.3** conduct your professional activities without discrimination on the grounds of sex, sexual orientation, marital status, nationality, colour, race, ethnic origin, religion, age or disability, or of any other condition or requirement; **2.6** avoid injuring others, their property, reputation, or employment by false or malicious or negligent action or inaction;
Hid the truth from the users	**1.3** Be honest and trustworthy. **2.4** 4 Accept and provide appropriate professional review. **2.7** 7 Foster public awareness and understanding of computing, related technologies, and their consequences.	**3.5** NOT misrepresent or withhold information on the performance of products, systems or services (unless lawfully bound by a duty of confidentiality not to disclose such information), or take advantage of the lack of relevant knowledge or inexperience of others. **2.5** respect and value alternative viewpoints and seek, accept and offer honest criticisms of work; **4.1** accept your personal duty to uphold the reputation of the profession and not take any action which could bring the profession into disrepute;
Software misused in educational context	**3.7** Recognize and take special care of systems that become integrated into the infrastructure of society.	**3.1** carry out your professional responsibilities with due care and diligence in accordance with the relevant authority's requirements while exercising your professional judgement at all times;

REFERENCES:

Association for Computing Machinery, 2018. ACM Code of Ethics and Professional Conduct. Available from: https://www.acm.org/code-of-ethics [Accessed 13 March 2022].

BCS The Chartered Institute for IT. 2021 The Code of Conduct. Available from: https://www.bcs.org/media/2211/bcs-code-of-conduct.pdf [Accessed 13 March 2022].