ETHICS IN TECH PRACTICE:

Case Study: AI Social Media Content Moderation

MARKKULA CENTER FOR APPLIED ETHICS
at Santa Clara University

AI Social Media Content Moderation

By Brian Patrick Green

A social media company is having trouble with political actors manipulating the flow of information on its service. Specifically, certain governments are producing tens or hundreds of thousands of fake accounts to promote government propaganda, thereby attempting to swamp any news which does not fit the government's perspective.

The social media platform is attempting to respond by using machine learning to determine which accounts are fake, based on their activity patterns, but the adversary governments are responding with their own machine learning to better hide those patterns and impersonate real users. For every batch of fake accounts deactivated, just as many seem to pop up again.

Furthermore, the machine learning algorithms are imperfect, and balancing false-positives with false negatives can lead to deactivating real people's accounts, leading to anger, frustration, and bad publicity for the company. On the other hand, scaling back to avoid false positives leads to more fake accounts slipping through.

What should the company do? What ethical questions should they consider? How might the questions below inspire perspectives on this problem?

Discussion questions:

1. What unique ethical concerns does this effort entail ?
2. Who are the stakeholders involved? Who should be consulted about the project's goals and development?
3. What additional facts might be required? What practical steps might you need to take in order to access the information/perspectives needed to manage the ethical landscape of this project?
4. What are some of the ethical issues that any designers/developers involved in this a project need to address?
5. How might this effort be evaluated through the various ethical 'lenses' described in the "Conceptual Frameworks" document?
6. In this project, what moral values are potentially conflicting with each other? Is there any way for the disagreeing sides to reconcile or does success for one necessarily mean failure for the other?
7. As a project team, how might you go about sorting through these ethical issues and addressing them? Which of the ethical issues you have identified would you prioritize, and why?
8. Who would be the appropriate persons on a team to take those steps? At what level, and by what methods, should decisions be made about how to manage the ethical issues raised by this project?