Ethics Case Study

AI Based Ethics – Rachel's Dilemma

Acknowledgement

This sample case study is derived from actual case that is documented in the 2019 CSIRO Discussion Paper "Artificial Intelligence: Australia's Ethics Framework".

Context

Rachel, an analyst at a big-data consultancy firm, is conducting research on social media activity. She gathers posts from the past two years across 10,000 users, ensuring their identities remain hidden in any published results. To process over 200,000 posts, she employs a machine learning (ML) system.

Dilemma

In a separate initiative through a data sharing scheme, Rachel acquires 100 million anonymised posts from 5 million users over two years. She later discovers that the usernames in reposts weren't removed, which compromises the anonymity. Her data science tools and ML system can identify many usernames in this supposedly anonymised data. This realisation shows that the additional data greatly enhances the value of her research.

Options

Rachel has several options to choose from. She can:

1. Scrub the data to remove usernames and then run her analysis on properly anonymised data.
2. Immediately discard the identifiable data, advise the supplying agency, and request the data again.
3. Research the supplying agency's data management policies, and if they're in breach of their policy, show them the identifiable data, request that it be replaced with anonymised data, and point out the potential for reputational damage arising from their data management practices.
4. Run the analysis just to see what it would produce, but then discard the identifiable data.
5. Run the analysis and produce and publish the results (not to mention the data breach).

Considerations

To uphold ACS values, she needs to be mindful of the following from the Code of Professional Ethics:

2.1a – Be honest, open and truthful: Rachel should not pretend she doesn't know there has been a breach; even if not challenged as to whether she knew, she should be proactive in being open about detecting the breach.

2.1c – Not remain silent: This reinforces the last part of analysing 2.1a above.

2.2a and 2.2b – Be accountable and practice integrity: If challenged at any time in the future about whether she knew of the breach, she should be able to say with integrity what she knew and did.

2.2d – Respect privacy and confidentiality of personal data: Clearly, she did not intend to see the IDs of those supposedly anonymised posts, so she should ensure that the data she has received is properly destroyed and check all links in the chain to the originator to ensure those data are not still exposed. Once she has noticed the breach, she should immediately stop examining the data.

Ethical Decision

The best approach for Rachel would be option 2 (immediately discard the identifiable data) or option 3 (research the agency's data management policies and contact them if required). Doing so ensures she acts with integrity and honesty.

Here's why the other options are inadequate:

Option 1 – This may be acceptable, but she does have a responsibility to the supplier of those data to ensure that they know and can take whatever corrective action may be required, including removing any intermediate copies of those non-anonymised data.

Option 5 – Unethical and not permissible.

Option 4 – Also, unethical and not permissible, no matter how tempting and discreet she may be about it.