Predicting Consent: Investigating Ethical Methods for Mining Social Media Data in Health Research

Social media services are an attractive data source for researchers, particularly within the field of health. Such work has found that social data can act as a predictor for health conditions including depression, suicidal ideation, mood changes, flu outbreaks, and problem drinking. Yet these findings highlight the sensitivity of what can be inferred from social data. This is particularly problematic if the users of such sites are unaware of their data being used and unable to withdraw from the process. This is increasingly common in health research using social media data and has significant ethical implications.

There are a few ways that participant consent can be achieved, but these often have their own limitations. The most common approach is a broad, one-off request at the outset. This is convenient for both the researcher and the participant, but it fails to give the participant detailed control over what data should be accessible and for what reasons. This has led to some academics advocating for Dynamic Consent, a system giving the participant detailed control over their data. While this encourages more transparent research, the downside is the significant cognitive burden associated with answering repeated requests for data.

Our research explores whether machine learning can be used to predict when users would consent to certain data being accessed by different hypothetical audiences. We are currently conducting a study involving up to 500 UK-based adults who are members of medical support groups on Facebook. These participants are repeatedly asked whether they would share different social items from their Facebook profile (e.g. a status update, photo, check-in location, photo album, or a page they have 'liked') with a hypothetical audience (e.g. researchers, clinicians, fellow members of the online support group, or open to the public). We collect metadata about the request, such as the data type, the hypothetical audience, attributes of the social item, and self-reported demographic information about the participant.

This study is producing a corpus of consent decisions, alongside contextual information about the request. This corpus will be used to train classifiers using machine learning techniques. These models will serve as a benchmark for future research, demonstrating whether it is possible to predict when people give consent for their social data to be used in health research using contextual data alone. The results of this research will inform the development of new and usable methods of consent, allowing researchers to continue making important findings while providing participants with easy and detailed control over their social data.

This work highlights several ethical questions about using AI to predict consent decisions. For example, what are the implications when such a model gets a prediction wrong? How should these algorithms balance transparency and accuracy if a trade-off exists? And under what circumstances should an algorithm be allowed to consent on behalf of a person? We welcome the opportunity to discuss our study and the ethical implications of our research with experts from the digital health community.

The authors have declared that no competing interests exist.