Peer to Peer Support

Using machine learning to identify individuals who are at risk of harm

The Alan Turing Institute

Turing Commons

Overview

A for-profit company has developed a peer-to-peer support service for individuals wishing to share experiences and seek advice from others with lived experiences of mental health issues.

This system is marketed to institutions such as schools, universities, and workplaces as a tool which can improve well-being as well as productivity.

The appeal of these peer-to-peer support platforms is significantly dependent on the fact users feel comfortable sharing personal details thanks to their anonymity. Therefore, all users are (by default) anonymous on the site and, in the vast majority of cases, can message one another free from outside intervention.

However, when certain combinations of 'trigger words' are entered into the system, indicating a user is at risk of causing harm to themselves or others, the incident is escalated. In these cases, the institutional mental health lead is notified, and anonymity is broken. The individual is then given personal assistance and intervention.

Consent for this platform does indicate that anonymity may be breached in extreme scenarios, but it does not detail the specific triggers of case escalation. Nor does it detail the means by which algorithmic technqiues are used to identity "at risk" users.

Key Consideration

Universities have multiple motivations for adopting these systems.

On the one hand, they are genuinely interested in improving member wellbeing. However, these systems also claim to help them improve productivity and at times can improve institutional reputation.

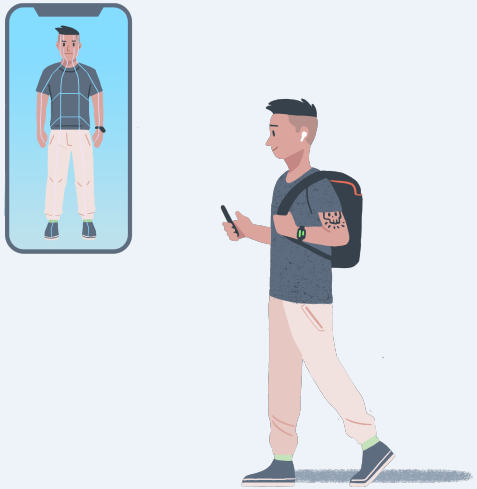
Deliberative prompts

1. **How do you think the developers of this service should navigate the trade-off between privacy-preserving anonymity and safety-conscious escalation of particular cases?**
2. **Would it be preferable for there to be no monitoring and just peer-to-peer support?**
3. **Do the power relations present in institutional settings impact your views on whether breaking anonymity is justifiable? Would it be preferable for the individual's identity to be disclosed to a third-party psychiatrist rather than the institution?**

Datasheet

Category	Details
Available Data	• Transcripts of the conversations between users and the counsellor • Metadata extracted from the app, including timestamps for the messages and IP address of the user • Additional comments recorded by the counsellor during the conversation • Post-conversation survey results from those users who completed the questionnaire

Analysis Techniques	• Natural language processing (NLP) of conversation transcripts to infer the following:
	  • Risk assessment of the user (e.g., risk of immediate harm to themselves or others)
	  • Possible demographic variables, including sexual orientation and gender identity, age, religious beliefs, and race



Groups, Organisations and Affected Individuals

1. **Users of support forum**
2. **For-profit company**
3. **University administrators**