

Questions for Evaluation

- Do you think this claim could be useful for demonstrating how an ethical value or principle has been upheld by the developers? If so, which value or principle?
- With reference to the previous value or principle, do you find this claim reassuring? If so, why? If not, why?
- What evidence, if any, would you expect to see to support or validate the claim?

Questions for Evaluation

- Do you think this claim could be useful for demonstrating how an ethical value or principle has been upheld by the developers? If so, which value or principle?
- With reference to the previous value or principle, do you find this claim reassuring? If so, why? If not, why?
- What evidence, if any, would you expect to see to support or validate the claim?

"Users agree to our terms of service during the sign-up process. When deciding whether to agree, users have access to our privacy policy. The privacy policy details how anonymity will be broken and shared with their organisation if a high chance of harm (to either the user or another) is identified by our algorithm and verified by a trained human healthcare professional."

Claim 1

"Use of the app has been linked to increased productivity and well-being."

Claim 2

"Monitoring of the platform ensures that all users are protected from abuse, bullying, or harm by others."

Claim 3

"When testing our algorithm, we verified that similar levels of accuracy were achieved across protected groups. Therefore, our algorithm is able to deal with diverse uses of language and uncommon phrases."

Claim 4

List of protected characteristics:

- age
- disability
- gender reassignment
- marriage and civil partnership
- pregnancy and maternity
- race
- religion or belief
- sex
- sexual orientation

"If a user is de-anonymised and their organisation is contacted, then a notification is sent to them with an explanation about why they were de-anonymised. This explanation consists of the forum post that was flagged by our algorithm and verified by the healthcare professional."

Claim 5

"Our system does not reduce the decision-making autonomy or negatively impact the professional judgement of a psychiatrist."

Claim 1

"Psychiatrists receive training on how to correctly use the system so that it does not negatively impact their communicative skills or ability to empathise on an interpersonal level with the patient."

Claim 2

"Recommendations made by the system are linked to specific properties of the patient's speech, with additional information about why a phrase or speech pattern is relevant for the assessment or specific diagnosis."

Claim 3

"The AI system uses information about the speech patterns of the patient, such as tone, pace, stuttering to make additional predictions about the patient's mental health."

Claim 4

"Anonymised data is collected from interactions and is used to identify opportunities to improve the system. This data is also used to develop additional products by our company, which are used outside of mental healthcare."

Claim 5