



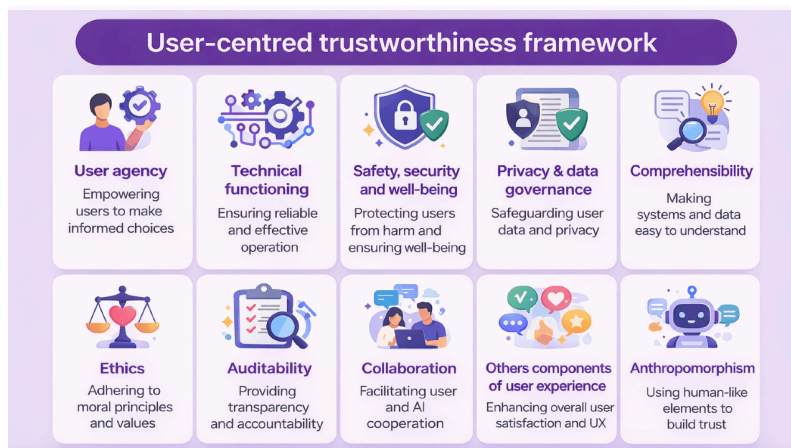
User-centred trustworthiness framework

How to evaluate the trust of the user in a system?

It is not enough to simply ask, “Do you trust the system?”

In the first years of the PEER project, CATIE worked to build a structured, multidimensional approach: a user-centered trustworthiness framework.

CATIE proposes a first version to operationalize perceived trustworthiness through ten core constructs retained for their relevance and collective contribution to perceived trustworthiness. The framework differs from design-centered models by focusing on capturing human judgments over system attributes.



1. **User agency** refers to the perceived user autonomy offered by the AI system (independent acts without over-relying on the system) and feeling of control (e.g. does the AI assistant follow my directions?) and how the implemented governance mechanism is satisfying for the user.
2. **Technical functioning** refers to the overall subjective performance of the system (i.e. does the AI assistant perform its job well?) which encompasses perceived accuracy, correctness, and reliability when facing new situations and new data, efficiency as in working quickly enough, and consistency over similar situ actions.
3. **Safety, security and well-being** encompasses the perception of how errors can be and are managed (i.e. detected, anticipated, avoided, reduced and corrected), as well as how the AI assistant is perceived as vulnerable to attacks, capable to warn people of potential risks, how users perceive that the AI assistant is concerned about their welfare and would not knowingly do things to hurt them or others.
4. **Privacy and data governance** refers to the user's view on data protection, data collection, access to data structure, or the appropriateness of the training data used.
5. **Comprehensibility** is the overall perceived clarity (i.e. whether the user can form a mental model and predict future system behavior) and learnability of

form a mental model and predict future system behavior) and learnability of the AI, because of the various design processes of explainability, interpretability and traceability.

6. **Ethics** refers to how the user perceives that the AI assistant considers diversity, non-discrimination, cultures, languages, accessibility, social impact on jobs and skills, or environmental friendliness.
7. **Auditability** refers to the extent to which users believe that the inner processes, decisions, data, algorithms and outcomes of the AI system can be traced and (independently) verified for an inspection or external audit.
8. **Collaboration** refers to the perceived dynamics of teaming between the user and the AI system. It encompasses how the AI gives feedback and (openly) communicates, how the user feels supported by the AI, how the AI considers user's feedback, the extent to which it allows for personalization or customization.
9. **Anthropomorphism** refers to the attribution of humanlike characteristics or mental states to the AI assistant (e.g. intentions, emotions, free will, mind or consciousness).
10. **Other components of user experience** broaden the framework to include all UX aspects, i.e. both the instrumental qualities (e.g., perceived usefulness, perspicuity) and non-instrumental components (e.g., novelty or stimulation).

 [Discover more about PEER](#) 

[PEER official website](#)

[PEER YouTube](#)

You are receiving this email because you signed up for our newsletter.

FOLLOW us on social media!



If you would like to unsubscribe, please click [here](#).

