PSYCHOLOGICAL TRUST IN ARTIFICIAL INTELLIGENCE TRANSLATIONS: A THEORETICAL APPROACH ABSTRACT

Izzatilloyeva Saida
Student of Navoi State University
Email: ew40600@gmail.com
Tel: +998918523731
Orcid ID: 0009-0006-0138-0748

Abstract: *This study investigates the determinants of human trust in Artificial Intelligence (AI) and Machine Translation (MT). While AI translation quality has improved significantly, the psychological acceptance of these outputs remains a complex challenge. This paper analyzes the mechanisms of "Algorithm Aversion" and the role of cognitive and affective trust in user interaction with translation technologies.*

Keywords: *Artificial Intelligence (AI) Translation, Psychological Trust, Algorithm Aversion, Neural Machine Translation (NMT), Cognitive Trust, Interpretability, Explainable AI, Human-Computer Interaction (HCI), Technology Acceptance Model (TAM), Uncertainty Signaling.*

Introduction

The integration of Neural Machine Translation (NMT) into daily communication has transformed linguistic mediation. However, a significant gap remains between technical accuracy and user confidence. Psychological trust in AI is defined as the user’s willingness to rely on an automated system in situations of uncertainty. A unique challenge in this field is that users often exhibit higher sensitivity to AI errors compared to human errors—a phenomenon that necessitates a deeper understanding of "trust filters" in the human mind.

Objective: To categorize the theoretical pillars of trust in AI translation and explain the psychological barriers to its adoption.

Methods

This research utilizes a thematic synthesis of existing Human-Computer Interaction (HCI) theories. We evaluated the Technology Acceptance Model (TAM) and Mayer’s Model of Trust to identify how users process AI-generated linguistic data. The focus was placed on user perception, error tolerance, and the impact of system transparency on reliability.

Results

The research identifies three primary factors that govern psychological trust in AI translations:

Cognitive Trust (Reliability): This is based on the rational evaluation of the output. If the translation is syntactically correct and contextually appropriate, the user attributes "competence" to the system.

Interpretability (Explainable AI): Trust levels increase significantly when the AI provides a rationale for its linguistic choices (e.g., providing alternative meanings for a polysemous word). Transparency reduces the "black box" effect. Trust in AI is complicated by the Transparency-Trust Paradox, which suggests that too much information can be as detrimental as too little. Excessive detail regarding AI decision-making often results in cognitive overload, leading to a decline in user confidence. To maximize reliability, AI interfaces should implement 'calibrated transparency', offering just enough contextual rationale to support a translation while avoiding the inclusion of unnecessary data."

Perceptual Quality (Interface Impact): The professional design and brand reputation of the translation platform act as psychological cues. A polished interface can create a "halo effect," leading users to perceive the underlying translation as more accurate.

Discussion

A critical barrier to trust is Algorithm Aversion. Studies show that after witnessing a single mistake, users lose confidence in an algorithm more rapidly than they do in a human peer. This is because humans expect AI to be "perfectly consistent."

To bridge this gap, AI systems must move toward a collaborative model. Instead of providing a single static output, trust is better maintained when the system engages in "uncertainty signaling"—flagging low-confidence translations and offering the user a choice. This shifts the user's role from a passive recipient to an active collaborator, which psychologically reinforces trust.

To mitigate algorithm aversion, the integration of a Human-in-the-Loop (HITL) framework is paramount. This paradigm posits that trust is not merely a static output of autonomous system performance but rather a dynamic, co-constituted process. When a Neural Machine Translation (NMT) system proactively identifies low-confidence linguistic segments and solicits user intervention through post-editing tasks, it effectively restores the user’s sense of agency. From a psychological perspective, this mechanism facilitates a cognitive shift in the user’s role—transitioning from a reactive 'error arbiter' to a proactive 'collaborative expert'—thereby fostering sustained, long-term systemic reliance."

Conclusion

Psychological trust in AI translation is a multidimensional construct that extends beyond mere linguistic precision. It is a balance of technical performance, transparency, and human-centric design. Future developments in translation technology should prioritize not just "better" translations, but more "explainable" and "human-compatible" interactions to ensure long-term user reliance.

References:

Mayer, R. C., et al. (1995). An Integrative Model of Organizational Trust.

Dietvorst, B. J., et al. (2015). Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err.

Venkatesh, V. (2003). User Acceptance of Information Technology: Toward a Unified View.

Schmidt, P., & Biessmann, F. (2019). Quantifying Interpretability and Trust in Machine Learning.

Kizilcec, R. F. (2016). Effects of Transparency on Trust in an Algorithmic Interface. CHI Conference on Human Factors in Computing Systems.