

The Impact of Human-AI Collaborative Feedback on Translation Quality and Decision-Making Processes among Undergraduate

Abstract

The integration of Artificial Intelligence (AI) in higher education is transforming translation pedagogy, shifting the role of human translators toward AI-assisted post-editing and evaluative mediation. While AI-driven tools facilitate dynamic and interactive learning environments, concerns persist regarding AI's cultural blind spots, stylistic flattening, and students' overreliance on AI at the expense of critical thinking. Despite the rapid adoption of these technologies, a systematic analysis of human-AI collaborative feedback in translation education remains a critical research gap.

The primary objective of this study is to investigate how a "Human-AI Collaborative Feedback" model influences undergraduate students' translation performance and cognitive decision-making processes. The study is situated within authentic educational contexts to address the current lack of empirical evaluations in AI-assisted translation learning. It further examines whether instructor-mediated AI feedback can reduce students' over-reliance on AI-generated literal translations and compensate for AI's limited cultural depth.

In addition, the study explores the development of "feedback literacy," an essential process-oriented competency. Through human-AI collaborative feedback, students are encouraged to move beyond passive acceptance of feedback and become active negotiators who can critically evaluate, compare, and reconcile diverse feedback sources.

This study adopted a mixed-methods approach to examine both linguistic outcomes and cognitive patterns. The participants consisted of 65 third-year students majoring in Applied Foreign Languages at a national university of science and technology. The course instructional design emphasizes a paradigm shift from literal "language switching" to "pragmatic translation" in authentic communicative contexts. Students were trained to identify and apply nuances in tone, role, stance, and register, which are dimensions where AI often demonstrates significant limitations.

The quantitative phase included a three-stage feedback process. The analysis compared the translation quality of three versions: the students' **initial drafts** produced under exam conditions; **AI-only revisions** based on unmediated ChatGPT-generated feedback; and **human-AI collaborative revisions** based on instructor-validated AI

feedback. Two certified translators scored the pre- and post-revisions using the Multidimensional Quality Metrics (MQM) framework to evaluate dimensions such as accuracy, terminology, language conventions, cohesion, cultural adaptation, register, and tone.

In the qualitative analysis, purposive sampling was used to select 15 representative students across performance levels for a thematic analysis of their decision-making recordings collected during the revision phase. Four core themes were identified: contextual awareness, noticing AI limitations, metacognitive decision-making, and awareness of implied meaning.

The statistical results indicated that while revisions based on AI feedback improved compared to the initial drafts, the difference was not statistically significant. However, revisions based on instructor-validated AI feedback significantly outperformed both the initial drafts and the AI-only feedback versions. The largest improvements occurred in cultural adaptation and tone; these findings confirm that human intervention in the feedback process is essential for higher-order pragmatic and cultural nuances.

Thematic analysis identified four core themes categorizing students' cognitive engagement. High-performing students frequently exhibited **contextual awareness** via instructor-validated AI feedback to correct the “contextual myopia” often found in unmediated AI output. In addition, participants recognized AI’s cultural blind spots and its tendency toward stylistic flattening. Recordings also revealed the metacognitive decision-making process through a two-stage revision pattern: students corrected surface-level errors first, and engaged in deeper deliberation when AI-generated feedback and instructor feedback diverged. Regarding awareness of implied meaning, students demonstrated an increased ability to discern non-literal, implied meanings that AI often misinterpreted.

These findings suggest that AI alone may not adequately support higher-order pragmatic and cultural mediation in translation-learning contexts due to its limited cultural depth and occasional accuracy errors. Instructor-mediated “human-AI collaborative feedback” fills these gaps, fosters critical engagement, and encourages students to engage in deeper pragmatic reasoning when faced with feedback divergence. This human-in-the-loop framework establishes a dialogic learning mechanism that empowers students to evaluate and negotiate diverse perspectives, ultimately enhancing their professional translation competence and feedback literacy.

Keywords: human-AI collaborative feedback, translation pedagogy, pragmatic translation, AI limitations, feedback literacy