International Journal of Language, Translation and Intercultural Communication

Vol 11 (2026)

The Sustainability of Interpreting as a Profession in the Era of Artificial Intelligence

The Loop We Didn't Choose

Stefanos Vlachopoulos

doi: 10.12681/ijltic.44870

To cite this article:

Vlachopoulos, S. (2026). The Loop We Didn't Choose: Human-AI Configurations and the Sustainability of Interpreting . *International Journal of Language, Translation and Intercultural Communication*, *11*, i-xii. https://doi.org/10.12681/ijltic.44870

The Loop We Didn't Choose:
Human-AI Configurations and the Sustainability of Interpreting

Stefanos Vlachopoulos

Ionian University

vlachopoulos@ionio.gr

Abstract

This special issue examines whether the interpreting profession can remain sustainable amid rapid advances in artificial intelligence. The papers in this issue provide empirical evidence, theoretical frameworks, and practical insights from diverse geographical contexts - Poland, Greece, Cyprus, Italy, Malta, Austria, Germany, and Japan - and across multiple interpreting modalities, including conference interpreting, community interpreting, speech-to-text interpreting, and broadcast interpreting. Together, they paint a complex picture of a profession at a critical crossroads. This introduction synthesizes the contributions through the conceptual lens of human-machine configurations - distinguishing between contexts where humans remain in the loop, where they operate on the loop in supervisory roles, and where they are being placed out of the loop illuminating not only the current state of AI interpreting but also the ethical conditions under which professional sustainability might be achieved.

Keywords*: interpreting sustainability, artificial intelligence, human-in-the-loop, human-on-the-loop, human-out-of-the-loop, AI ethics, professional identity, interpreter training*

1 Situating the Challenge: Humans and the loop

In an increasingly globalized world where communication across linguistic and cultural boundaries is more vital than ever, the interpreting profession stands at a critical crossroads. The call for papers for this special issue, issued in May 2025, posed a deceptively simple question: What constitutes a sustainable future for interpreting in economic, environmental, social, cultural, and academic terms, against the backdrop of unprecedented artificial intelligence innovation? The papers coming together in this volume represent scholarly responses to this question, offering empirical evidence, theoretical insights, and practical recommendations from researchers working across Europe and Asia.

The timing of this publication is not coincidental. The years 2024-2026 have witnessed an acceleration in AI capabilities that few in the profession could have anticipated even a decade ago. Large Language Models (LLMs) have evolved from text-based curiosities to multimodal systems capable of processing and generating spoken language in real time. Speech-to-Speech (S2S) translation platforms have moved from laboratory prototypes to commercially available services, aggressively marketed as cost-effective alternatives to human interpreters. Industry reports cited

across multiple papers in this issue paint a stark picture: interpreters and translators rank first among occupations with the highest AI applicability scores, with 98% of their work activities overlapping with tasks that AI systems can purportedly perform (Tomlinson et al., 2025).

Yet, as the papers in this issue collectively demonstrate, the gap between technological capability and professional reality remains substantial. The industry narrative of imminent *human parity* collides repeatedly with empirical evidence from controlled studies, real-world observations, and the lived experiences of practicing interpreters.

The papers published in this volume advance an analytical framework built around three indispensable configurations of human - machine relations. Rather than functioning as optional interpretive tools, these configurations emerge as the essential lenses through which the future of the profession could be understood: the *human-in-the-loop* configuration, a human-guided configuration, required where interpreters must actively shape, refine, and validate system output; *a human-on-the-loop* configuration, a human-supervisory configuration, in which practitioners oversee and manage AI systems, intervening strategically rather than continuously; and finally, a fully automated configuration, identified by the research in this volume as especially critical and high-risk, in which communicative decisions proceed without human expertise, the *human-out-of-the-loop* configuration (Natarajan et al. 2025).

1.1 On Numbers, Machines and Humans

Recent market surveys consistently highlight a growing interdependence between human linguists and AI systems. *Slator* (2025) shows that while AI-driven language technologies are expanding rapidly, effective workflows increasingly rely on *hybrid, expert-in-the-loop models* in which human judgement safeguards both quality and ethical oversight. The *2025 Nimdzi 100* similarly reports that although AI is reshaping translation and interpreting, it primarily amplifies rather than replaces human expertise - especially in domains demanding precision, contextual awareness, and cultural competence. *Ad Astra (2025)* further confirms that AI has become standard in everyday workflows but remains unsuitable for *regulated or high-stakes communication* without sustained human supervision. Taken together, these surveys depict a future in which AI augments - rather than displaces - the professional interpreter.

Building directly on this, the same surveys underscore that ethical considerations are central to managing human - AI collaboration in language services. Across reports, responsible AI adoption is framed as contingent on *trust, transparency, and human accountability*, with sustainable workflows requiring governance structures that address data use, privacy, and model reliability. *Slator* (2025) stresses that hybrid systems are necessary not only to protect quality but also to uphold ethical responsibilities in sensitive multilingual contexts. Likewise, the *Nimdzi 100* identifies ethical risk - particularly in legal, medical, and institutional settings - as a key reason AI must remain subordinate to human contextual judgement. *Ad Astra* (2025) reaffirms that unresolved concerns around accuracy, compliance, and user safety render AI

unsuitable for unsupervised deployment in environments where communicative errors carry significant consequences. Collectively, these surveys confirm that ethical accountability, cultural sensitivity, and professional judgement cannot be automated.

These concerns connect directly to broader debates on human - AI interaction and our proposed set of configurations for classifying the findings in the papers published in this issue.

As Hau (2025) argues, the familiar *human-in-the-loop* metaphor is not neutral: it subtly positions the machine as primary and the human as a corrective afterthought, thereby creating a hierarchy of agency. If the loop framework is adopted uncritically, human interpreters risk being reduced to mere *safety mechanisms* for algorithmic systems - what Htin (2026) terms the *moral crumple zone*, in which humans absorb the moral, legal, and reputational consequences of system failures despite having little real authority or control over the system's behaviour.

The ethical challenge, then, is not only to determine which technical configuration works best, but to question whether the loop metaphor itself aligns with the values the profession seeks to uphold. As the European Language Council's reflection paper *AI for Translation and Interpreting: A Roadmap for Users and Policy Makers* (Peeters et al., 2025) asserts, "responsible AI use - anchored in human values, ethics, equity and the pursuit of quality of communication - is the only sustainable path forward."

This introduction adopts that stance by treating the above human - AI configurations not merely as technical options but as sites of ethical and professional contestation that must be evaluated through the lens of interpreter agency, communicative responsibility, and public trust.

2 The Human-in-the-Loop Configuration

The first and most extensively documented configuration across these papers is the *human-in-the-loop* model, in which human interpreters remain central to the production of high-quality interpreting output, whether working independently or in collaboration with AI tools. The evidence from studies published in this issue demonstrates that human intervention is not merely desirable but essential.

2.1 AI produces Words. Humans produce Meaning

Korybski et al., in their study of the Polish-English language pair, simulated a conference setting, complete with a live audience, a bilingual speaker delivering a 40-minute lecture, and a bi-directional Q&A session. Two groups of human interpreters - EU-accredited professionals and MA-level students - worked alongside two leading commercial MI platforms, all interpreting the same source material simultaneously. The findings are unequivocal: error counts revealed a clear and consistent performance hierarchy, with accredited interpreters producing the fewest errors, followed by students, with both MI systems trailing significantly. More importantly, the qualitative analysis identified systematic weaknesses in MI output that simple error counts cannot

capture: ASR-related misrecognitions, overly literal translations, redundant punctuation voiced by synthetic speech, random language switches, gender bias, and failures linked to limited contextual memory. The authors conclude that "despite technological progress, MI performance in this setting remained far from human-like as of mid-2025."

Matsushita's contribution approaches the same question from a different angle, focusing not on error counts but on communicative effectiveness. Working with 56 professional journalists in Japan, she compared comprehension outcomes between those listening to human interpretation and those relying on MI during an actual press conference on climate policy. The results are striking: the human group achieved a mean comprehension score of 4.5/10 compared to 3.7/10 for the AI group, with the latter exhibiting a 17.9% "Don't Know" rate. Qualitative feedback highlighted that AI's lack of prosodic salience—its flat, mechanical delivery—increased cognitive load, hindering listeners' ability to synthesize information deeply.

Ioannidis's analysis of English-Greek interpreting in the European Parliament adds a crucial dimension to this picture by revealing the fundamentally different operating logics of human and machine processing. His quantitative findings initially appear counterintuitive: the human interpreter produced 49 errors compared to only 28 for the AI system, yielding a higher error density. However, the qualitative analysis transforms our understanding of these numbers. The human interpreter's errors were dominated by omissions - strategic compressions that preserved fluency and discourse coherence at the cost of selective information loss. By contrast, the AI system produced no omissions but clustered its errors in categories with greater potential communicative risk: pragmatic appropriateness, mistranslations and semantic shifts, and terminology. His paper shows that even when AI output appears superficially fluent, it remains vulnerable to cascading errors that humans, drawing on contextual and world knowledge, would never produce.

The bottom line is that AI focuses on words and listeners understand less. Humans focus on meaning and listeners understand more, even with less information.

Fragkou’s contribution to this volume extends and refines this conclusion by transposing it from the domain of professional interpreting into that of interpreter training and assessment. Her corpus-based study of student renditions in an MA healthcare interpreting programme demonstrates that the AI–human divide identified by Korybski et al., Matsushita, and Ioannidis is replicated, in miniature and with characteristic clarity, when AI is called upon not to interpret but to evaluate. AI assessment tools prove consistently more reliable when measuring the linguistic surface of student output - accurately tracking omissions, additions, and distortions, and reliably evaluating grammatical correctness, lexical choices, and informational completeness. They perform far less dependably, however, when the object of assessment shifts to the prosodic and paralinguistic dimensions of the interpreted rendition: the student’s intonation contours, rhythmic patterning, pitch variation, control of pausing, pragmatic force, and the affective colouring of the delivery. These are precisely the features that Matsushita’s journalists found absent and cognitively costly in AI-interpreted output - the same features that human interpreters instinctively manage and that human assessors instinctively notice. Fragkou’s findings therefore

corroborate, from an educational measurement angle, what the performance studies demonstrate from the perspective of the listener: that what AI systems process most readily is the word-level content of interpreted discourse, while the meaning-bearing dimensions that give interpreting its communicative force remain substantially beyond their evaluative reach.

2.2 The Ethics of Keeping Humans in the Loop

Petrocheilou's survey reinforces these patterns: her analysis of responses from 49 highly experienced interpreters in Greece, Cyprus, and Italy reveals experiences and attitudes that closely mirror the experimental findings, combining clear apprehension about AI's shortcomings with cautious acknowledgement of its potential benefits. When asked about AI's main limitations, 77.6% of respondents identified the lack of contextual adaptability as the most significant concern, followed by the inability to account for overall context (69.4%), the high risk of miscommunication (65.3%), and monotony or lack of engagement (57.1%). The qualitative responses elaborate on these themes, noting that "a machine cannot be expected to genuinely understand a speaker or grasp their underlying intentions," whereas human interpreters "are able to empathize, discern what the speaker seeks to convey, and find an appropriate way to communicate that message."

Yet the survey also makes clear that interpreters are not fearful of technological innovation. When asked about AI's advantages, 79.6% highlighted its potential as a supplementary tool in low-impact scenarios, while 69.4% acknowledged its relatively low cost. Many respondents anticipated a restructuring of the profession rather than its disappearance, envisioning future roles in which interpreters act more as supervisors, proofreaders, or quality controllers of AI-generated output. This imagined hybrid future - where human expertise is redeployed rather than rendered obsolete -aligns with the second configuration of human–machine relations.

The ethical dimension highlighted in Petrocheilou's paper also resonates with Keller's (2025) argument that human-in-the-loop oversight is not optional but imperative for mitigating documented risks in AI-mediated interpreting. Both perspectives frame ethics not as a constraint on innovation but as its necessary foundation. Petrocheilou's survey further underscores the presence of accountability gaps, showing that ethical integration is essential for preserving the irreplaceable human capacities—connection, empathy, nuance—that remain absent from AI systems. This reframes interpreters not as replaceable labour but as essential ethical custodians, proactively adapting to AI's rise while safeguarding the human essence of communicative exchange.

3 The Human-on-the-Loop Configuration

The *human-on-the-loop* configuration positions interpreters as supervisors and quality controllers of AI-generated output, intervening selectively rather than continuously. This model emerges across multiple domains, from conference interpreting to speech-

to-text accessibility services and is explicitly advocated in several papers in this collection.

3.1 Pedagogical Preparation for Supervisory Roles

The pedagogical papers in this volume address this configuration most directly. Colman's study of Maltese-English interpreter training examines the potential of GenAI tools (Gemini, Copilot, ChatGPT) for generating practice speeches – a task that positions trainers and students in a supervisory relationship with AI output. Her findings are sobering but instructive. Despite careful prompting based on DG SCIC's Speech Repository criteria, the generated speeches in both English (high-resource) and Maltese (low-resource) proved unsuitable for beginner consecutive practice. The speeches lacked clear argumentative structure, exhibited low factual density, and featured excessively simplistic terminology. Maltese outputs contained significant linguistic errors, reflecting the limitations of AI systems trained primarily on English-language data - a disparity that raises ethical concerns about linguistic equity in AI development.

The implication is clear: before AI-generated materials are to be used in training, they require extensive human post-editing and pedagogical framing. Trainers must remain firmly *on the loop*, evaluating, selecting, and adapting AI output rather than accepting it as ready-to-use.

Wiedenmayer takes a more optimistic but equally critical stance. She argues that AI-generated speeches, when embedded within a coherent pedagogical framework aligned with European Masters in Conference Interpreting (EMCI) principles, can enhance training effectiveness by enabling precise control over discourse parameters. The paper provides detailed prompt parameters for calibrating information density, terminology, syntax, cohesion, and cognitive load across training stages -from introductory to professional simulation level. Crucially, she emphasizes that critical thinking remains of central importance and that students must be encouraged to approach AI as part of "a dialogic process: formulating their own ideas, producing initial written work, submitting it to the model for feedback, critically evaluating the generated comments, and subsequently revising their texts.

This vision positions both trainers and students as actors *on-the-loop*, engaging with AI output reflectively and strategically rather than accepting it passively. The goal is not to replace human judgement but to augment it - to use AI as a partner in the development of interpreting competence. This aligns with what the *pedagogy first* principle (Hau 2025) where AI serves as a co-agent and contingent actor, invited selectively into the learning process where it proves occasionally helpful. Pedagogy remains the central force, ensuring AI stays secondary to human-centered teaching methods.

Fragkou's paper takes the pedagogical dimension of the on-the-loop configuration in a further and distinctive direction: rather than asking how AI can assist in generating training materials, she asks how AI can assist in evaluating students' interpreted output - and what, in that evaluative task, must remain firmly in human hands. Working from a corpus of thirty recordings produced by six first-year MA students across three

interpreting modes (sight translation, consecutive interpreting, and bidirectional medical dialogue), she conducts a longitudinal comparative analysis of human and AI assessment over a fourteen-month period. The findings establish a division of evaluative labour that both supports and constrains the pedagogical use of AI tools. On the loop, AI assessment can perform valuable preliminary screening work, flagging errors of content transfer, incomplete renditions, or terminological lapses with a reliability that human assessors acknowledge. But it cannot replace the instructor's ear. Prosodic and paralinguistic features - intonation, rhythm, stress, pausing, speaker affect, pragmatic force - require sustained human attention and remain beyond the evaluative capacity of current AI systems.

The pedagogical implication is clear: in interpreter training, as in Colman's and Wiedenmayer's accounts of AI-generated practice materials, the instructor must remain firmly on the loop, using AI as an analytical aid while retaining exclusive authority over the holistic evaluation of communicative performance. Fragkou also raises substantive methodological and ethical concerns specific to the assessment context: the anonymisation of student audio data, the status of voice recordings as potentially identifiable biometric information, and the sources of bias embedded in both human and AI evaluation instruments. These concerns reinforce the conclusion that AI-assisted assessment, like AI-assisted content generation, cannot be left to run unsupervised - it must be governed, monitored, and critically interrogated by competent human practitioners.

3.2 Supervisory Roles in Professional Practice

The professional papers gesture toward *on-the-loop* configurations with varying degrees of optimism. Petrocheilou's respondents anticipated "a restructuring of the profession, with interpreters serving less as primary providers of communication and more as supervisors, proofreaders, or quality controllers of AI-generated output." The WHO's decision to restrict AI use to internal meetings "under strict human supervision" exemplifies this model in institutional practice. Similarly, the American Translators Association and the National Association of Judiciary Interpreters and Translators recommend that "all AI-generated translations be reviewed for accuracy and completeness by human translators" - a clear on-the-loop prescription.

Platter and Eichmeyer-Hell's analysis of speech-to-text interpreting (STTI) in Germany and Austria reveals both the potential and the fragility of *on-the-loop* configurations. In their community interpreting case study, a hard-of-hearing user requiring STTI for parent evenings at her child's school found herself caught between authorities proposing different levels of human involvement. The local educational community proposed using a "free" AI-transcription smartboard, effectively placing the user in a *human-out-of-the-loop* configuration. The consulting interpreter's extensive advocacy - four hours of calls and emails to clarify responsibilities and argue for human STTI - represented an *on-the-loop* intervention aimed at maintaining quality standards. In this case, the interpreter's supervisory role extended beyond quality control to encompass advocacy for appropriate service provision.

3.3 The Ethics of Supervisory Responsibility: beyond Interpreting Expertise

The *human-on-the-loop* configuration raises distinctive ethical questions about responsibility, attention, and the limits of human cognitive capacity. As Htin (2026) warns, "the human brain is evolutionarily wired for active engagement, not for prolonged periods of inactive supervision. When an AI system performs correctly most of the time, the operator disengages, a state known as out-of-the-loop unfamiliarity". This creates what legal scholars call a "vigilance decrement", a deterioration in the ability to detect anomalies during passive monitoring tasks (Htin 2026).

For interpreters positioned *on the loop*, this is not merely a theoretical concern but a practical and ethical hazard. If an interpreter is expected to supervise AI output while simultaneously managing other tasks, or if the AI performs well enough most of the time to induce complacency, the moment of system failure may be precisely the moment when human attention is least available.

This analysis resonates with Platter and Eichmeyer-Hell's documentation of broadcast interpreting failures. When AI-generated captions included conversations from wrong courts or rendered only 16% of content, no human was present to catch these errors. The *on the loop* configuration had collapsed into effective human absence not because no human was responsible, but because the supervisory mechanisms were inadequate.

This adds another layer of concern: simply adding a human won't guarantee that the outcomes are better, some processes are inherently complex, and the human may just not understand the process enough to be able to meaningfully intervene. For interpreting, this suggests that effective *on-the-loop* supervision requires human competence and that competence must be specifically developed for supervisory roles, not simply assumed on the basis of interpreting expertise.

4 The *Human-Out-of-the-Loop* Configuration and Its Ethical Perils

The most concerning configuration documented across these papers is the *human-out-of-the-loop* model, in which decisions are fully automated and humans are excluded from the communicative chain. This configuration threatens professional sustainability directly by eliminating the need for human interpreters, and indirectly by normalizing lower quality standards that devalue the profession's core competencies.

4.1 The Illusion of Human Parity and the Accountability Gap

The experimental papers complicate any simple narrative of human superiority by revealing contexts where AI appears to outperform humans. Matsushita's data show that the AI group actually outperformed the human group on two specific questions (Q1 and Q5), suggesting that AI's literal repetition of keywords can sometimes serve as a mnemonic anchor. Korybski et al. note that MI systems sometimes produce more complete renderings than humans, who strategically omit information under time

pressure. Ioannidis demonstrates that AI can produce no omissions at all, preserving the full informational load of the source text.

These findings create a paradox: AI can be simultaneously – and linearly - better by certain – superficial - metrics (fewer errors, more complete information) and inferior by the ones that count when it comes to effective human communication, the successful transfer of meaning, intent, and interpersonal positioning. Here human interpreters retain decisive advantages - as the experiments have shown.

The danger is that decision-makers, lacking the conceptual frameworks to distinguish these dimensions of quality, will rely on surface-level metrics as justification for *human-out-of-the-loop configurations*. Petrocheilou's respondents express this anxiety directly: "few clients place genuine value on high-quality human interpreting"; "audiences will eventually accept the absence of human interpreters"; "even if the quality is not of the highest standard, users would gradually adapt to new norms and cease to expect premium results."

The danger of surface-level metrics is not confined to the professional sphere. Fragkou's study reveals the same dynamic operating within interpreter education, where AI assessment tools can generate plausible-looking evaluations of student renditions that are, in important respects, systematically incomplete. Because current AI tools score linguistic and informational dimensions reliably, they can produce numerical grades or qualitative labels that appear authoritative - high scores for grammatically correct, informationally complete renditions - while remaining entirely blind to the prosodic and paralinguistic dimensions that define communicative quality in healthcare interpreting: the student's capacity to modulate affect, signal pragmatic intent, manage turn-taking, and deliver information with appropriate pitch, rhythm, and emotional colouring.

Replacing human assessment with AI-generated evaluations would not merely reduce assessment quality; it would actively mismeasure what it means to interpret well, rewarding technically complete but communicatively flat output and, over time, reshaping student expectations accordingly. The trajectory Petrocheilou's respondents fear in professional contexts - a gradual accommodation to lower standards - could thus be set in motion at the training stage itself, before interpreters ever enter the profession, if assessment tools are adopted without adequate critical scrutiny of what they can and cannot measure.

This raises profound ethical questions about accountability. As Htin's (2026) analysis of AI negligence argues, when no human is present to catch errors, to exercise judgement, to take responsibility, the communicative chain becomes not only lower in quality but fundamentally unreliable. The accountability gap identified by 46.9% of Petrocheilou's survey respondents becomes unbridgeable in *human-out-of-the-loop* configurations.

The WHO's experience, cited in Petrocheilou's contribution, illustrates this accountability gap in practice. Across 90 interpretations in six official languages, only one obtained a passing grade, with reputational risks ranging from one to nine per speech. The misgendering of Dr. Moeti as a man, the rendering of "Hamas" as

incomprehensible "Ifer," the transformation of "hepatitis" into "Ebola" - these are not mere errors but potential sources of diplomatic incidents, legal liability, and harm to vulnerable individuals. In *human-out-of-the-loop* configurations, there is no one to catch these errors, no one to exercise judgement, no one to take responsibility.

As to the legal and economic aspects, Platter and Eichmeyer-Hell note that accessibility regulations in Germany and Austria, while mandating "adequate quality" and "completeness," leave these terms undefined, creating a regulatory vacuum that cost-driven decisions can exploit. The absence of enforceable quality standards, combined with the absence of human oversight, creates conditions under which "accessibility" becomes a hollow promise and the real risk lies not in technological capability but in economic decisions that shift the burden of poor quality onto audiences.

The European Language Council's roadmap (Peeters et al., 2025) offers specific recommendations for addressing these challenges. It urges organizations to "consider the ethical and legal implications of using AI for T&I purposes and potential liabilities and costs that may arise from issues related to output quality, data protection, accountability, and unwanted biases". It also recommends that organizations use professional translators and interpreters where data protection, high quality and nuance matter - a clear rejection of *human-out-of-the-loop* configurations (in high-stakes settings).

5 Reclaiming the Lead: Toward Ethically Sustainable Configurations

Taken together, the contributions to this special issue reveal an interpreting profession confronted not with a single technological destiny but with a spectrum of evolving human-machine configurations. The future is not predetermined by technological capability; it will be negotiated through context-specific choices, ethical commitments, and the collective agency of practitioners, trainers, researchers, institutions, and users. Across all papers, one insight stands out clearly: sustainability in interpreting hinges not on resisting AI outright, but on shaping the conditions under which it is adopted.

The evidence presented here shows that *human-in-the-loop* configurations remain essential wherever meaning, intention, and interpersonal nuance carry high stakes. *Human-on-the-loop* models offer promise, but they demand pedagogical and professional reorientation to avoid the vigilance and accountability gaps that arise when human expertise is miscast as passive oversight. The most ethically fraught option - the *human-out-of-the-loop* configuration - may appear economically attractive, yet the documented failures across broadcast, institutional, and accessibility settings demonstrate that removing human judgement does not merely reduce quality; it destabilizes trust, fairness, and responsibility.

At the same time, the pedagogical and research contributions in this issue underscore that sustaining the profession requires more than warning against risk. It requires cultivating the evaluative, reflective, and supervisory competencies that allow interpreters to work with AI without being subsumed by it. It requires training programs that treat AI as a contingent, pedagogically justified tool - not a shortcut - and research

agendas that combine rapid benchmarking with qualitative depth and user-centered evaluation. Colman's and Wiedenmayer's findings on AI-generated practice materials, and Fragkou's findings on AI-assisted student assessment, collectively demonstrate that pedagogical applications of AI in interpreting education demand the same critical, on-the-loop engagement that interpreting in professional and institutional contexts require: not passive acceptance of AI-generated outputs but active, competent human oversight of what those outputs do - and, crucially, what they fail to do. Fragkou's longitudinal design further underscores the importance of building assessment research agendas that track AI capabilities over time, since the boundary between what AI can and cannot evaluate is not fixed but shifts as models evolve - a reason for ongoing vigilance rather than settled confidence in any current configuration. It also requires legal and organizational frameworks that ensure oversight is real rather than symbolic, protecting both users and practitioners from the "moral crumple zone" of poorly governed automation.

Ultimately, the question facing the field is not whether AI will transform interpreting, but who will guide that transformation, according to which values, and toward which ends. If stakeholders prioritize communicative integrity, ethical governance, and the cultivation of human expertise, interpreting can not only survive but evolve in ways that reaffirm its social purpose. If cost-driven, *out-of-the-loop* automation becomes normative, the profession risks both degradation and displacement.

Our relationship to the loop should not be imposed; it should be something we consciously choose, deliberately design, and rigorously defend.

6 References

Benedikt, L., Joshi, C., Nolan, L.A., Henstra-Hill, R., Shaw, L. & Hook, S. (2020) 'Human-in-the-Loop AI in government: A case study', in *Proceedings of the 25th International Conference on Intelligent User Interfaces*, pp. 488–497. ACM.

European Language Council (2025) *AI for Translation and Interpreting: A Roadmap for Users and Policy Makers*. Peeters, K., Daems, J., Plieseis, C., Rivas Ginel, M.I. & Şahin, M. (eds).

Hau, D. (2025) *Beyond the loop: Reclaiming pedagogy in an AI age*. UNESCO. Available at: https://www.unesco.org/en/articles/beyond-loop-reclaiming-pedagogy-ai-age (Accessed: 9 February 2026).

Hoffmann, C.H. & Bansal, D. (eds.) (2025) *AI Ethics in Practice: Navigating Academic Insight, Managerial Expertise, and Philosophical Inquiry*. Cham: Springer Nature Switzerland.

Holzinger, A. (2016) 'Interactive machine learning for health informatics: When do we need the human-in-the-loop?', *Brain Informatics*, 3, pp. 119–131.

Htin, N.M. (2026) 'Redefining the standard of human oversight for AI negligence', *Harvard Journal of Law & Technology*.

Keller, L. (2025) ‘AI and language interpreting’, in Hoffmann, C.H. & Bansal, D. (eds.) *AI Ethics in Practice: Navigating Academic Insight, Managerial Expertise, and Philosophical Inquiry*. Cham: Springer Nature Switzerland, pp. 57–70.

Natarajan, S., Mathur, S., Sidheekh, S., Stammer, W. & Kersting, K. (2025) ‘Human-in-the-loop or AI-in-the-loop? Automate or collaborate?’, *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(27), pp. 28594–28600.

Tomlinson, K., Jaffe, S., Wang, W., Counts, S. & Suri, S. (2025) *Working with AI: Measuring the Occupational Implications of Generative AI*. Available at: https://doi.org/10.48550/arXiv.2507.07935 (Accessed: 23 February 2025).

United Nations Educational, Scientific and Cultural Organization (UNESCO) (2025) *Beyond the Loop: Reclaiming Pedagogy in an AI Age*. Available at: https://www.unesco.org/en/articles/beyond-loop-reclaiming-pedagogy-ai-age (Accessed: 24 February 2025).

The Routledge Handbook of Interpreting, Technology and AI (2025). Taylor & Francis.