FUTURITY OF SOCIAL SCIENCE

DOI: https://doi.org/10.57125/FS.2026.03.20.10

How to cite: Milcu, M. (2026). Artificial intelligence in simultaneous interpreting: capabilities, limitations, and human–machine collaboration. *Futurity of Social Sciences, 4*(1), 169–187. https://doi.org/10.57125/FS.2026.03.20.10

Artificial Intelligence in Simultaneous Interpreting: Capabilities, Limitations, and Human–Machine Collaboration

Marilena Milcu

Associate Professor, Lucian Blaga University of Sibiu, Romania,
https://orcid.org/0000-0002-4221-8855

Corresponding author: *marilena_milcu@yahoo.com.*

Received: December 2, 2025 | **Accepted:** February 23, 2026 | **Published:** March 15, 2026

Abstract: The integration of artificial intelligence (AI) into simultaneous interpreting (SI) constitutes one of the most consequential developments in language mediation studies of the past decade. Recent advances in neural machine translation (NMT), automatic speech recognition (ASR), and large language models (LLMs) have produced systems capable of rendering speech across language boundaries in near-real time, yet the fundamental cognitive, pragmatic, and cultural demands of professional conference interpreting continue to expose critical architectural limitations in current AI designs. This article provides a comprehensive interdisciplinary analysis of the state of AI in SI, structured around three principal aims: (1) to map the cognitive architecture of human SI against the computational strategies of current AI systems; (2) to synthesise empirical evidence on AI interpreting performance across language pairs, registers, and delivery conditions; and (3) to evaluate human–AI collaboration frameworks — including computer-assisted interpreting (CAI) and

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

remote simultaneous interpreting (RSI) platforms — for their practical and ethical implications. We argue that the prevalent framing of AI as a replacement for human interpreters is both technically premature and epistemically inadequate. A more productive paradigm — augmentation — positions AI tools as extending interpreter competencies while preserving the pragmatic, affective, and cultural mediation capacities that remain beyond the reach of existing architectures.

Keywords: artificial intelligence; simultaneous interpreting, automatic speech recognition, neural machine translation, human–machine collaboration, conference interpreting, cognitive load, computer-assisted interpreting, remote simultaneous interpreting, natural language processing.

Introduction

Simultaneous interpreting (SI) — the real-time, continuous oral rendering of source-language speech into a target language with minimal temporal lag — represents one of the most cognitively demanding tasks in professional language practice. Unlike consecutive interpreting or written translation, SI requires the interpreter to listen, comprehend, syntactically restructure, and produce speech in the target language while the source speech stream continues uninterrupted, typically for periods of 20–30 minutes in professional conference settings (Gile, 1995, 2009; Pöchhacker, 2016). The resulting cognitive architecture — simultaneously engaging working memory, executive function, phonological encoding, lexical retrieval, syntactic processing, and pragmatic inference — has been characterised as a paradigmatic case of human expertise under extreme cognitive constraint (Moser-Mercer, 2005; Seeber, 2011).

Against this background, the rapid acceleration of artificial intelligence research — particularly in ASR, NMT, and LLMs — has generated sustained debate about the degree to which SI can be automated, augmented, or restructured by machine systems. The chronology of this debate tracks the broader history of machine translation: from cautious rule-based approaches (Bar-Hillel, 1960; Hutchins, 2007), through the transformative but limited advances of statistical phrase-based models (Brown et al., 1990; Koehn et al., 2007), to the transformer architecture (Vaswani et al., 2017) and the subsequent generation of large-scale neural models — Whisper (Radford et al., 2022), SeamlessM4T (Barrault et al., 2023), and GPT-4 (OpenAI, 2023) — whose performance in controlled conditions has prompted renewed and sometimes hyperbolic claims about AI's capacity to replicate or replace human interpreters.

Professional interpreting organisations — most notably the International Association of Conference Interpreters (AIIC) — have documented persistent quality gaps in AI-generated output across all high-stakes interpreting domains and have raised substantive concerns about accountability, informed consent, and labour market displacement (AIIC, 2022). Simultaneously, a growing body of research in computer-assisted interpreting (CAI) has demonstrated the value of AI tools that support rather than supplant human interpreters, particularly for terminology management, cognitive load monitoring, and preparation assistance (Fantinuoli, 2017, 2018; Sandrelli, 2021). The COVID-19 pandemic further accelerated this trajectory by normalising remote

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

simultaneous interpreting (RSI) platforms, creating digital environments in which AI integration is both technically feasible and commercially compelling (Braun, 2020).

This article situates itself at the intersection of interpreting studies, cognitive science, and natural language processing. Section 2 examines the cognitive architecture of human SI and its implications for machine replication. Section 3 traces the technological development of AI interpreting systems. Section 4 synthesises empirical evidence on system performance. Section 5 analyses human–AI collaboration models. Section 6 addresses ethical, legal, and labour market dimensions. Section 7 presents a structured research agenda, and Section 8 concludes.

Cognitive Architecture of Simultaneous Interpreting

Gile's Effort Models: A Computational Lens

Daniel Gile's Effort Models (1995, 2009) remain the most widely cited framework for characterising the cognitive processing demands of SI. Gile proposed that simultaneous interpreting requires the allocation of finite cognitive processing capacity among three concurrently active efforts: the Listening and Analysis Effort (LA), the Production Effort (P), and the Memory Effort (M), along with a Coordination component (C). The critical insight of the Tightrope Hypothesis is that the total capacity demanded by LA + P + M + C at any processing moment may approach or exceed the interpreter's total available capacity, producing "saturation" — manifested as omissions, errors, approximations, or output degradation (Gile, 2009).

When applied as a heuristic for analysing AI interpreting architectures, the Effort Models offer a revealing but imperfect analogy. ASR subsystems perform the functional equivalent of the Listening and Analysis Effort; NMT components execute the Production Effort; and attention mechanisms, memory buffers, and contextual encoders serve functions loosely analogous to the Memory Effort. A key asymmetry, however, is that current AI architectures are not subject to the same bottleneck dynamics that create saturation in human interpreters: parallel computational pipelines eliminate the single-channel capacity constraint. The implication is not that AI faces no constraints, but that its limitations are qualitatively different — rooted in training data coverage, architectural inductive biases, and pragmatic depth rather than processing capacity per se (Fantinuoli, 2018).

Interpretive Theory: The Meaning-Extraction Problem

The Interpretive Theory of Translation — the Théorie du Sens developed by Seleskovitch (1975) and elaborated by Lederer (1994) — advances a claim with profound implications for AI interpreting: the interpreter's fundamental task is not linguistic transcoding but the extraction of non-verbal meaning (vouloir dire) from the source utterance, followed by its re-verbalisation in the target language. The unit of interpreting is not the word or phrase but the intended meaning. This entity is irreducibly contextual, extralinguistic, and dependent on the interpreter's world knowledge, situational awareness, and theory of the speaker's mind.

This theoretical claim directly challenges the adequacy of current neural language models as interpreting systems. Notwithstanding their impressive surface-level fluency, transformer-based

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

NMT systems are fundamentally distributional pattern-matchers: they lack world knowledge, intentional states, or genuine comprehension (Bender et al., 2021; Marcus & Davis, 2019). Their facility with language reflects statistical regularities in training corpora rather than meaning in the phenomenological sense invoked by the Paris School. The practical consequence — visible in AI output analyses — is systematic fragility in the presence of pragmatic indirection, culturally specific allusion, diplomatic understatement, technical domain specificity, and speaker-idiosyncratic register (Koehn & Knowles, 2017; Papi et al., 2023).

Neuroscience of Professional Interpreting

Neuroimaging research has progressively illuminated the neural underpinnings of professional SI, revealing a distributed cortical and subcortical network that is substantially reorganised relative to that of bilingual non-interpreters. Key regions consistently implicated include Broca's area (BA44/45) and Wernicke's area for linguistic production and comprehension; the supplementary motor area (SMA) for speech planning; the anterior cingulate cortex (ACC) for conflict monitoring and error detection; the caudate nucleus for language switching; and prefrontal regions associated with working memory maintenance and executive control (Hervais-Adelman et al., 2015; Rinne et al., 2000). Structural imaging studies have documented grey matter adaptations in bilateral Heschl's gyri and inferior frontal regions in professional interpreters, consistent with neuroplastic reorganisation through intensive training (Elmer et al., 2014; Becker et al., 2016).

These findings have two implications for AI research. First, professional SI competence is not reducible to a collection of separable linguistic sub-skills. However, it reflects a deeply integrated, embodied neural architecture — developed over years of deliberate practice — with no direct computational analogue. Second, the interpreter's physical and affective state is performance-relevant: fatigue-induced degradation (Moser-Mercer et al., 1998), anxiety effects on working memory (Köpke & Nespoulous, 2006), and stress responses to high-stakes content all constitute dimensions that AI systems neither experience nor model.

Working Memory, Anticipation, and Syntactic Asymmetry

Working memory capacity — specifically the phonological loop and central executive components of Baddeley and Hitch's (1974) multicomponent model — is a widely documented predictor of both interpreter trainee performance and professional SI quality (Christoffels & De Groot, 2005; Signorelli et al., 2012). The phonological loop maintains a temporary acoustic trace of the most recent source-language input while the central executive coordinates the parallel demands of comprehension and production; degradation in either component under saturation produces the characteristic error patterns described in Gile's (1995) Tightrope Hypothesis.

A related challenge largely unresolved in current AI architectures is the anticipation problem. In subject-object-verb (SOV) or verb-final languages (German, Japanese, Turkish, Korean), the main verb — which anchors the semantic frame of the entire clause — appears at the end of the source utterance. Human interpreters develop language-pair-specific anticipation strategies combining syntactic expectation, pragmatic inference, and domain knowledge to begin target-language

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

production before the critical source material is available — a cognitive feat with no straightforward implementation in streaming neural architectures (Seeber, 2011; Van Besien, 1999).

Technological Development of AI Interpreting Systems

Rule-Based and Statistical Precursors

The history of machine translation provides the essential developmental context for understanding current AI interpreting capabilities. Rule-based MT (RBMT) systems, dominant from the 1950s through the 1980s, encoded explicit linguistic knowledge — lexicons, morphological rules, syntactic grammars — to perform interlingual transfer. The ALPAC Report (1966) delivered a damaging assessment of RBMT's capacity to achieve acceptable quality at reasonable cost, effectively halting significant investment for over a decade. Statistical machine translation (SMT), formalised by Brown et al. (1990) at IBM, modelled translation as the maximisation of a posterior probability from a translation model and a target-language model. Phrase-based SMT underpinned systems such as early Google Translate (Koehn et al., 2007), achieving substantially better performance on benchmark evaluations but remaining unable to handle long-distance syntactic dependencies, morphological richness, and discourse-level coherence — all of which are critically relevant to continuous spoken language SI.

Neural Machine Translation: The Transformer Revolution

Neural machine translation using recurrent encoder-decoder architectures with soft attention mechanisms (Bahdanau et al., 2015; Cho et al., 2014) constituted a qualitative advance over SMT, enabling the modelling of context-dependent lexical choices, morphological agreement, and longer-range dependencies. The transformer architecture (Vaswani et al., 2017) — based entirely on multi-head self-attention without recurrence — established the foundation for all subsequent state-of-the-art language models. By enabling fully parallelised computation and modelling global dependency structures without positional decay, the transformer dramatically improved both training efficiency and translation quality. Pre-trained language model paradigms (Devlin et al., 2019; Radford et al., 2018) demonstrated that massive unsupervised pre-training on diverse textual data, followed by task-specific fine-tuning, could achieve strong NLP performance with minimal domain-specific data.

For simultaneous speech translation, the central technical challenge is the quality-latency trade-off. Simultaneous NMT systems — including SimulTrans (Ma et al., 2019), CIF-based approaches (Dong & Xu, 2020), and prefix-to-prefix training methods — address this through read/write policies that determine, at each decoding step, whether to read more source input or generate the next target token. Earlier commitment reduces latency but increases the probability of committing to an incorrect rendering; this fundamental trade-off remains unresolved in the literature (Arivazhagan et al., 2019; Elbayad et al., 2020).

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

Automatic Speech Recognition: The First Bottleneck

ASR constitutes the primary input stage of any AI interpreting pipeline, and its error rate directly constrains the performance of downstream translation components. Contemporary end-to-end ASR systems achieve near-human word error rates (WERs below 4%) on high-resource clean-speech benchmarks. However, performance degrades substantially in the conditions characteristic of professional interpreting: non-native speaker accents (WER increases of 20–60% relative to native speakers), spontaneous speech with filled pauses and repairs, overlapping speech, domain-specific technical terminology, and varied microphone quality in RSI setups (Radford et al., 2022; Tatman & Kasten, 2017).

OpenAI's Whisper model (Radford et al., 2022), trained on 680,000 hours of multilingual audio data, represents the current state of the art in robust multilingual ASR. Whisper demonstrates near-human WER on English benchmarks and competitive performance across 96 languages. Its robustness to acoustic noise and accented speech is particularly notable. However, its architecture was designed for offline transcription with access to full utterances; adaptation to streaming simultaneous recognition — where transcription must proceed incrementally without full context — introduces latency and accuracy trade-offs that are the subject of active research (Macháček et al., 2023).

End-to-End Speech-to-Speech Translation

The most technically ambitious AI interpreting paradigm is end-to-end speech-to-speech translation (S2ST), in which a unified neural architecture maps source-language audio directly to target-language speech output, bypassing intermediate text representations. Early S2ST work with Translatotron (Jia et al., 2019) demonstrated proof-of-concept feasibility and the potential of preserving speaker voice characteristics — a feature with important implications for naturalness and speaker identity preservation. Meta AI Research's SeamlessM4T (Barrault et al., 2023) represents the current frontier of large-scale multilingual S2ST, demonstrating competitive performance across nearly 100 languages within a single unified model architecture. Nevertheless, controlled quality evaluations against professional human interpreters in high-stakes conference settings reveal substantial deficits in pragmatic appropriateness, terminological precision, and prosodic coherence (Papi et al., 2023; Bentivogli et al., 2021).

Empirical Evidence on AI Interpreting Performance

Quality Assessment Methodologies

The evaluation of AI simultaneous interpreting quality presents a layered set of methodological challenges. Standard MT metrics such as BLEU (Papineni et al., 2002) — which measures n-gram overlap between system output and human reference translations — are poorly suited to interpreting quality assessment: they assume access to appropriate reference translations, are insensitive to fluency and naturalness, do not account for the temporal dimension of simultaneous delivery, and conflate different types of translational choices. More recent neural evaluation metrics — BLEURT

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

(Sellam et al., 2020) and COMET (Rei et al., 2020) — are substantially more sensitive to semantic equivalence and stylistic appropriateness, but have been validated primarily on written MT output and require adaptation for spoken simultaneous translation (Bentivogli et al., 2021).

In interpreting studies, quality evaluation has traditionally employed multidimensional frameworks that assess accuracy, completeness, fluency, and pragmatic appropriateness. Instruments such as the Interpreting Quality Assessment (IQA) protocol (Pagura et al., 1994) and the Assessment Scale for Interpreting Quality (Viezzi, 1996) have been developed for professional evaluation contexts but have not been systematically applied to AI outputs. Hybrid evaluation methodologies -that combine automatic metrics with expert human assessment and naive listener comprehension tests are increasingly advocated as the gold standard for AI SI evaluation (Papi et al., 2023). However, no consensus protocol currently exists across the relevant research communities.

Cross-System Performance Analysis

Table 1 presents a synthesised comparative analysis of reported performance metrics for representative AI simultaneous interpreting systems across selected language pairs and evaluation conditions. Cross-system comparability is limited by methodological heterogeneity in evaluation protocols, reference sets, and domain selection.

Table 1

Comparative Performance of AI Simultaneous Interpreting Systems

System / Study	Lang. Pair	Domain	Metric	Score	Latency	Architecture
SeamlessM4T v2 (Barrault et al., 2023)	EN→ES	General speech	COMET	0.812	~2.8s	End-to-end S2ST
SeamlessM4T v2 (Barrault et al., 2023)	EN→FR	General speech	COMET	0.798	~2.8s	End-to-end S2ST
Whisper + NLLB (cascaded)	EN→DE	TED Talks	BLEU	28.4	~3.5s	Cascaded ASR+MT
SimulTrans (Ma et al., 2019)	EN→DE	News/Talks	BLEU	18.6	3.0s AL	Simult. NMT
Translatotron 2 (Jia et al., 2022)	EN→ES	Read speech	BLEU	22.1	—	End-to-end S2ST
BabelNet AI (Fantinuoli, 2021)	EN→IT	EU Parliament	BLEU	27.3	4.1s	CAI assisted
Human Professional SI (reference)	EN→DE	Conference	BLEU (est.)	~38–45	Variable	Human baseline

Note. AL = Average Lagging. COMET scores range from 0 to 1. Human BLEU estimates are approximate and context-dependent.

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

The cross-system comparison reveals a consistent pattern: AI systems achieve performance approaching human reference levels in high-resource, low-register-specificity language pairs on general-purpose spoken corpora, while exhibiting substantially larger gaps in domain-specific technical content, lower-resource language pairs, and spontaneous, unscripted delivery styles. Performance on English–German is particularly informative given German's SOV structure, which imposes high anticipatory demands.

Qualitative Failure Analysis: Prosody, Pragmatics, and Cultural Mediation

Beyond aggregate quantitative metrics, qualitative analyses of AI interpreting output reveal systematic and theoretically significant failure patterns. Prosodic rendering — the preservation or appropriate target-language adaptation of the source speaker's intonation, emphasis, rhythm, and emotional register — represents a critical weakness of cascaded ASR-NMT-TTS pipelines, in which affective information encoded in the source signal is largely discarded during text-mediation (Rabinovich et al., 2017). While end-to-end S2ST models show improved prosodic transfer, evaluations of naturalness and emotional appropriateness consistently rate their output substantially below that of professional human interpreters (Barrault et al., 2023; Jia et al., 2022). In diplomatic, legal, and medical settings, prosodic flattening or misalignment can fundamentally misrepresent the speaker's communicative intent.

Pragmatic adequacy — encompassing the interpretation of indirect speech acts, implicature, irony, register management, and the integration of non-linguistic contextual information — perhaps represents the deepest gap between AI and human interpretation. Contemporary LLMs exhibit impressive surface-level pragmatic facility on constrained benchmarks, but consistently fail in cases requiring genuine Theory of Mind, real-world situational reasoning, or integration of non-linguistic contextual cues (Bender et al., 2021; Marcus & Davis, 2019). In international conference settings, where political subtext, diplomatic register, and culturally specific rhetorical conventions are pervasive, these failures carry significant practical consequences.

Language Pair and Domain Asymmetries

Performance asymmetries across language pairs and subject domains are among the most extensively documented characteristics of current AI interpreting systems. All evaluated architectures show dramatically higher performance for high-resource language pairs – those well-represented in training data – than for lower-resource pairs. English–Romanian, English–Swahili, and virtually all Indigenous language pairs frequently fall below usefulness thresholds (Nekoto et al., 2020; Bender et al., 2021). The commercial logic of AI development amplifies rather than corrects this disparity, entrenching a technological stratification with troubling implications for access to quality interpreting services in lower-resource linguistic communities.

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

Human–AI Collaboration Frameworks

The Augmentation Paradigm

The most theoretically coherent and practically productive framework for AI in SI is not replacement but augmentation: the strategic deployment of AI tools to extend and enhance the capabilities of human interpreters while preserving the irreplaceable human dimensions of language mediation. This paradigm has deep roots in the human-computer interaction literature, where Engelbart's (1962) foundational vision of "intelligence augmentation" positioned computational technology as one that amplifies rather than substitutes for human cognitive capacities. Applied to SI, augmentation encompasses terminology and glossary assistance tools, cognitive load monitoring, AI-powered preparation tools, and ASR-based captioning support – each targeting a specific component of the cognitive effort architecture (Fantinuoli, 2018; Corpas Pastor & Gaber, 2020).

Computer-Assisted Interpreting Tools: Taxonomy and Evidence

Table 2 presents a structured taxonomy of currently available and experimentally evaluated CAI tool categories, organised by targeted cognitive function, primary implementation in existing platforms, and the strength of available evidence for cognitive load effects.

Table 2

Taxonomy of Computer-Assisted Interpreting (CAI) Tools

Tool Category	Target Cognitive Function	Platform Examples	Evidence Strength
Terminology assistance / glossary	Lexical retrieval (LA + P effort)	InterpretBank, Intragloss, Kudo	Moderate (Fantinuoli, 2018)
Real-time ASR / captioning	Listening support (LA effort)	Interprefy AI, Whisper, Otter.ai	Weak-moderate (Sandrelli, 2021)
Speech rate monitor	Cognitive load management	Speechmatics dashboard	Preliminary (Seeber, 2011)
Document preparation / summarisation	Domain knowledge acquisition	ChatGPT, Claude AI	Emerging (Corpas Pastor, 2020)
AI post-editing support	Output revision (P effort)	Wordly, KUDO AI	Limited (Papi et al., 2023)
Parallel AI output channel	Fallback / quality monitoring	Wordly full AI mode	Contested (AIIC, 2022)

Note. LA = Listening and Analysis; P = Production (Gile, 1995). Evidence strength reflects the current state of controlled empirical evaluations.

Fantinuoli's (2018) controlled pilot study of terminology assistance tools is among the most methodologically rigorous contributions to CAI evaluation. Using a within-subjects experimental design with 12 professional interpreters, the study demonstrated that glossary assistance significantly improved terminology accuracy in specialised technical domains without statistically

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

significant increases in measured cognitive load. However, the small sample size and laboratory conditions constrain generalisability. A key finding across available CAI studies is that the cognitive load implications of AI tool use are not unidirectional: while tools targeting lexical retrieval can reduce saturation-related errors, poorly designed interfaces and unreliable ASR output can introduce new cognitive demands that partially offset the intended benefits (Corpas Pastor & Gaber, 2020).

Remote Simultaneous Interpreting and AI Integration

The normalisation of remote simultaneous interpreting (RSI) platforms — accelerated dramatically by the COVID-19 pandemic — has created a digital infrastructure in which AI integration is both technically feasible and commercially compelling at scale. Major RSI platforms, including Interprefy, Kudo, Zoom (with third-party language routing), and Interactio have incorporated varying degrees of AI-powered functionality: automatic transcription, AI-generated glossaries, post-hoc MT as an accessibility backup, and — in the case of Wordly — fully automated simultaneous interpretation without human interpreters (Braun, 2020; Sandrelli, 2021). User satisfaction research suggests that meeting participants rate fully automated output as adequate for general informational content but substantially below the quality of professional human interpreters for presentations involving technical precision, rhetorical complexity, or politically sensitive content (Wordly User Research, 2023).

Cognitive Load Implications of AI-Assisted Interpreting

Cognitive load theory (Sweller, 1988; Paas et al., 2003) provides a useful additional lens for evaluating CAI tool design. Effective AI tools should reduce extraneous cognitive load while preserving the germane load associated with constructing domain representations and exercising professional judgement. Physiological measures of cognitive load — pupillometry, electroencephalography (EEG) spectral power changes, and galvanic skin response — have been applied to SI research with promising results (Seeber & Kerzel, 2012; Koshkin et al., 2018) but have not yet been systematically deployed in evaluations of AI-assisted interpreting. This methodological gap represents a significant limitation that requires a research programme combining output quality assessment, measurement of physiological cognitive load, and interpreter self-report under ecologically valid conditions.

Ethical, Legal, and Professional Dimensions

Quality Standards, Accountability, and Informed Consent

The deployment of AI interpreting systems in high-stakes communicative contexts — criminal proceedings, asylum hearings, medical consultations, human rights tribunals, and diplomatic negotiations — raises acute ethical questions about quality assurance, accountability, and informed consent. International instruments protect the right to a competent interpreter in legal proceedings including Article 6(3)(e) of the European Convention on Human Rights and Directive 2010/64/EU. These instruments set quality thresholds that current AI systems demonstrably fail to meet in the

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

spontaneous, domain-specific, pragmatically complex conditions of actual legal proceedings (Mikkelson, 2017). The accountability vacuum surrounding AI interpreting errors is particularly problematic: when a credentialed human interpreter commits an error, professional accountability mechanisms provide structured responses, whereas for AI systems, liability is typically disclaimed through terms-of-service agreements (Bender et al., 2021; FIT Europe, 2021).

A principle of minimal adequacy in professional ethics requires that participants in a communicative event be informed of the qualifications and limitations of the interpretation they are receiving. In AI-mediated or AI-assisted interpreting, disclosure of the technology's involvement, its known limitations, and the degree of human oversight present constitutes an ethical obligation that is not yet formalised in professional or regulatory frameworks in most jurisdictions. AIIC's (2022) position paper calls for mandatory disclosure when AI-generated interpretation is provided; However, this recommendation lacks enforcement mechanisms and has not been incorporated into procurement regulations for multilateral organisations, including the United Nations and the European institutions.

Labour Market Implications and Professional Identity

The labour market implications of AI in SI are likely to be uneven, domain-stratified, and temporally extended rather than sudden. Conference interpreting in high-resource language pairs faces the greatest near-term competitive pressure, particularly for lower-complexity content in structured, controlled delivery conditions. Community interpreting, legal interpreting, and healthcare interpreting in lower-resource language pairs are likely to remain predominantly human-delivered in the medium term due to higher pragmatic complexity, more vulnerable participant populations, and stronger legal quality requirements (Kelly, 2022; Rinsche & Portera-Zanotti, 2009). Historical analogies from adjacent professional fields are instructive: the introduction of CAT tools transformed the translation profession without eliminating it, restructuring tasks and compensation models rather than substituting practitioners wholesale (DePalma & Kelly, 2008). However, post-editing workflows have been associated with increased expectations for processing speed and reduced per-word compensation — analogous dynamics may emerge in AI-assisted interpreting (Moorkens et al., 2018).

Bias, Linguistic Diversity, and Equity

AI interpreting systems reproduce and potentially amplify the linguistic inequalities encoded in their training data. The concentration of high-quality training corpora in a small number of high-resource, economically dominant languages creates systematic performance disparities that are structural features rather than incidental technical limitations. Documented gender biases in NMT systems — including systematic masculinisation of professional role translations and inaccurate pronoun resolution for gender-neutral source languages — carry direct practical consequences in interpreting settings where speaker identity and social position are communicatively significant (Sun et al., 2021). Similarly, ASR performance disparities against non-native-speaker accents and dialectal variation mean that the populations most likely to depend on interpreting services for access to justice and healthcare receive the worst-performing AI interpretation — a structural injustice that

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

cannot be addressed through technical optimisation alone (Nekoto et al., 2020; Tatman & Kasten, 2017).

Future Research Agenda

The interdisciplinary research agenda implied by the foregoing analysis is structured around four priority areas.

Technical Research Priorities. The primary technical challenges are: (1) developing simultaneous NMT architectures with improved quality-latency trade-offs for typologically diverse language pairs, particularly SOV languages; (2) improving ASR robustness to spontaneous speech, non-native accents, and acoustic noise without degradation in streaming inference; (3) incorporating speaker-level and discourse-level contextual representations that better approximate the pragmatic depth of human interpretation; (4) developing principled methods for prosodic transfer in end-to-end S2ST; and (5) building domain-adaptive systems capable of rapid specialisation without catastrophic forgetting. Large-scale pretraining is unlikely to resolve the pragmatic depth limitations identified in Section 4; hybrid neuro-symbolic architectures that incorporate structured world knowledge warrant continued investigation (Marcus & Davis, 2019).

Cognitive and Empirical Research Priorities. These include: (1) physiological studies of cognitive load in AI-assisted interpreting using pupillometry and EEG under ecologically valid conditions; (2) longitudinal studies of skill development and atrophy effects in interpreters who use AI tools extensively; (3) comparative studies of end-user comprehension, trust calibration, and satisfaction with AI-generated versus human-provided interpretation; and (4) studies of error detection capacity under time pressure. A standardised SI-specific evaluation benchmark incorporating automated metrics, expert human assessment, and listener comprehension measures is a methodological priority (Papi et al., 2023).

Policy and Regulatory Research Priorities. These include: (1) comparative legal analysis of existing interpreter qualification and liability frameworks across jurisdictions; (2) development of minimum quality threshold standards for AI interpretation in specific use contexts; (3) design of disclosure and informed consent protocols appropriate to AI-mediated interpreting; and (4) economic modelling of AI's labour market effects on different interpreter market segments.

Training and Pedagogy Research Priorities. These include: (1) curriculum models for integrating AI literacy and CAI tool competence into interpreter training programmes; (2) assessment frameworks that can differentiate AI-augmented from unassisted interpreter performance; and (3) investigation of the pedagogical value of AI tools in trainee self-assessment and formative feedback contexts (Sandrelli, 2021).

Conclusion

This article has provided a theoretically grounded, empirically anchored, and critically informed analysis of AI's current role and future trajectory in simultaneous interpreting. The principal conclusions are as follows. First, current AI interpreting systems achieve performance approaching

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

professional human interpreting quality only in constrained conditions that are not representative of the majority of high-stakes professional SI contexts. Second, the gaps between AI and human interpreting are not merely quantitative deficits amenable to incremental technical improvement; they reflect qualitative architectural limitations in pragmatic depth, cultural embeddedness, and real-time contextual reasoning that constitute the core of professional interpreting competence.

Third, the augmentation paradigm — in which AI tools extend and support human interpreter capabilities rather than replacing them — is both technically well-supported and ethically preferable for the foreseeable horizon of AI development. The evidence base for cognitive load benefits of specific CAI tool categories is encouraging but methodologically limited; a sustained programme of ecologically valid empirical research is needed to establish the conditions under which AI assistance genuinely improves interpreter performance. Fourth, the ethical, legal, and equity dimensions — accountability gaps, linguistic diversity disparities, and the rights of vulnerable communicative participants — require urgent regulatory attention and cannot be resolved through technical optimisation alone.

The fundamental question confronting the field is not whether AI can simulate simultaneous interpreting — within the narrow set of conditions where it currently performs adequately, it demonstrably can — but what kind of interpreting society values and what communicative rights it is prepared to guarantee to all its members. Answering that question requires not only technical research but interdisciplinary dialogue, ethical deliberation, and inclusive policy-making in which the professional interpreting community, the communities served by interpreting, and the developers of AI systems all have a genuine voice.

Acknowledgements

None.

Conflict of Interest

The author declares no conflict of interest.

Funding

The author received no funding for this research.

References

AIIC. (2022). *AIIC position paper on artificial intelligence and machine interpreting*. International Association of Conference Interpreters. https://aiic.org/page/9107

Automatic Language Processing Advisory Committee. (1966). *Languages and machines: Computers in translation and linguistics*. National Academy of Sciences.

Arivazhagan, N., Cherry, C., Macherey, W., Chiu, C.-C., Yavuz, S., Pang, R., Li, W., & Wu, Y. (2019). Monotonic infinite lookback attention for simultaneous machine translation. In *Proceedings of*

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

the 57th Annual Meeting of the Association for Computational Linguistics (pp. 1313–1323). Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1126

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). Academic Press.

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*. https://doi.org/10.48550/arXiv.1409.0473

Bar-Hillel, Y. (1960). The present status of automatic translation of languages. *Advances in Computers*, *1*, 91–163.

Barrault, L., Bhandare, Y., Bhatt, I., Bhosale, S., Duquenne, P.-A., Heffernan, N., Hwang, J., & Le, H. (2023). *SeamlessM4T: Massively multilingual and multimodal machine translation*. arXiv. https://doi.org/10.48550/arXiv.2308.11596

Becker, M., Schubert, T., Strobach, T., Gallinat, J., & Kühn, S. (2016). Simultaneous interpreters vs. professional multilingual controls: Group differences in cognitive control as well as brain structure and function. *NeuroImage*, *134*, 250–260. https://doi.org/10.1016/j.neuroimage.2016.03.079

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)* (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Bentivogli, L., Bertero, M., Di Gangi, M., Gaido, M., Negri, M., & Turchi, M. (2021). Is it worth it? Comparing six evaluation sets for English to German simultaneous speech translation. In *Proceedings of the 18th International Conference on Spoken Language Translation (IWSLT 2021)*. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.iwslt-1.26

Braun, S. (2020). Towards a pedagogical framework for the integration of remote interpreting in interpreter education. *Journal of Specialised Translation*, *33*, 324–343.

Brown, P. F., Cocke, J., Della Pietra, S. A., Della Pietra, V. J., Jelinek, F., Lafferty, J., Mercer, R., & Roossin, P. (1990). A statistical approach to machine translation. *Computational Linguistics*, *16*(2), 79–85.

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1724–1734). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1179

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

Christoffels, I. K., & De Groot, A. M. B. (2005). Simultaneous interpreting: A cognitive perspective. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 454–479). Oxford University Press.

Corpas Pastor, G., & Gaber, M. (2020). Ad hoc interpreting and the language of COVID-19: Terminological challenges for non-professional interpreters. *Revista de Lenguas para Fines Específicos*, *26*(2), 10–33.

DePalma, D. A., & Kelly, N. (2008). Project management for translators. In J. Drugan (Ed.), *Quality in professional translation* (pp. 123–147). Continuum.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Di Gangi, M. A., Gaido, M., Bentivogli, L., Negri, M., & Turchi, M. (2019). Enhancing transformer for end-to-end speech-to-text translation. In *Proceedings of Machine Translation Summit XVII* (Vol. 1, pp. 21–31).

Dong, L., & Xu, C. (2020). CIF: Continuous integrate-and-fire for end-to-end speech recognition. In *Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6079–6083). IEEE. https://doi.org/10.1109/ICASSP40776.2020.9054250

Elbayad, M., Besacier, L., & Verbeek, J. (2020). Efficient wait-k models for simultaneous machine translation. In *Proceedings of Interspeech 2020* (pp. 1461–1465). ISCA. https://doi.org/10.21437/Interspeech.2020-1480

Elmer, S., Hänggi, J., & Jäncke, L. (2014). Processing demands upon cognitive, linguistic, and articulatory functions promote grey matter plasticity in the adult multilingual brain. *Cortex*, *54*, 179–191. https://doi.org/10.1016/j.cortex.2013.11.001

Engelbart, D. C. (1962). *Augmenting human intellect: A conceptual framework* (Summary Report AFOSR-3223). Stanford Research Institute.

Fantinuoli, C. (2017). Computer-assisted interpreting: Challenges and future perspectives. In G. Corpas Pastor (Ed.), *Researching and using terminological resources in interpreting* (pp. 224–250). Cambridge Scholars Publishing.

Fantinuoli, C. (2018). Interpreting and technology: The rising of the machine? In C. Fantinuoli (Ed.), *Interpreting and technology* (pp. 1–12). Language Science Press.

Fantinuoli, C. (2021). Machine interpreting: On the boundary of science and fiction. In *Proceedings of the 1st Workshop on AI for Simultaneous Interpreting (AI4SI), LREC 2020*.

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

FIT Europe. (2021). *Technology and the future of translation: A position paper*. International Federation of Translators, European Regional Centre.

Gaiba, F. (1998). *The origins of simultaneous interpretation: The Nuremberg Trial*. University of Ottawa Press.

Gile, D. (1995). *Basic concepts and models for interpreter and translator training*. John Benjamins.

Gile, D. (2009). *Basic concepts and models for interpreter and translator training* (2nd ed.). John Benjamins.

González-Davies, M., & Scott-Tennent, C. (2005). A problem-solving and student-centred approach to the translation of cultural references. *Meta: Journal des traducteurs*, *50*(1), 160–179. https://doi.org/10.7202/010663ar

Hervais-Adelman, A., Moser-Mercer, B., & Golestani, N. (2015). Brain functional plasticity associated with the emergence of expertise in extreme language control. *NeuroImage*, *114*, 264–274. https://doi.org/10.1016/j.neuroimage.2015.03.048

Hutchins, W. J. (2007). Machine translation: A concise history. *Computer Aided Translation: Theory and Practice*, *13*, 29–70.

Jia, Y., Ramanovich, M., Remez, R., Pomeranz, L., Moreno, I., & Weiss, R. J. (2019). Direct speech-to-speech translation with a sequence-to-sequence model. In *Proceedings of Interspeech 2019* (pp. 1123–1127). ISCA. https://doi.org/10.21437/Interspeech.2019-2820

Jia, Y., Ramanovich, M., Wang, Q., Zen, H., Wu, Y., Zhang, Y., & Bapna, A. (2022). Translatotron 2: High-quality direct speech-to-speech translation with voice preservation. In *Proceedings of the 39th International Conference on Machine Learning (ICML 2022)*. https://doi.org/10.48550/arXiv.2107.08661

Kelly, N. (2022). *The economics of the AI interpreting market* (Nimdzi Research Report 2022). Nimdzi Insights.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., & Ney, H. (2007). Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics: Demo and Poster Sessions* (pp. 177–180). Association for Computational Linguistics. https://doi.org/10.3115/1557769.1557821

Koehn, P., & Knowles, R. (2017). Six challenges for neural machine translation. In *Proceedings of the 1st Workshop on Neural Machine Translation* (pp. 28–39). Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-3204

Köpke, B., & Nespoulous, J.-L. (2006). Working memory performance in expert and novice interpreters. *Interpreting*, *8*(1), 1–23. https://doi.org/10.1075/intp.8.1.03kop

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

Koshkin, R., Plaksin, A., & Ossadtchi, A. (2018). Measuring cognitive load in simultaneous interpreters using pupillometric data. *PLOS ONE*, *13*(2), Article e0190736. https://doi.org/10.1371/journal.pone.0190736

Lederer, M. (1994). *La traduction aujourd'hui: Le modèle interprétatif*. Hachette.

Ma, M., Pino, J., Cross, J., Dong, L., & Diab, M. (2020). Monotonic multihead attention. In *Proceedings of the 8th International Conference on Learning Representations (ICLR 2020)*. https://doi.org/10.48550/arXiv.1909.02480

Macháček, D., Dabre, R., & Bojar, O. (2023). *Turning Whisper into real-time transcription system*. arXiv. https://doi.org/10.48550/arXiv.2307.14743

Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Pantheon Books.

Mikkelson, H. (2017). *Introduction to court interpreting* (2nd ed.). Routledge.

Moorkens, J., Lewis, D., Kenny, D., Way, A., & Doğru Ünal, A. (2018). Correlations of perceived post-editing effort with measurements of machine translation quality. *Machine Translation*, *32*(1), 93–114. https://doi.org/10.1007/s10590-018-9224-8

Moser-Mercer, B. (2005). Remote interpreting: Issues of multi-sensory integration in a multilingual task. *Meta: Journal des traducteurs*, *50*(2), 727–738. https://doi.org/10.7202/011016ar

Moser-Mercer, B., Künzli, A., & Korac, M. (1998). Prolonged turns in interpreting: Effects on quality, physiological and psychological stress. *Interpreting*, *3*(1), 47–64. https://doi.org/10.1075/intp.3.1.03mos

Nekoto, W., Marivate, V., Matsila, T., Fasubaa, T., Kolawole, T., Fagbohungbe, T., Akinola, S. O., & Muhammad, S. H. (2020). Participatory research for low-resourced machine translation: A case study in African languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020* (pp. 2144–2160). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.findings-emnlp.195

OpenAI. (2023). *GPT-4 technical report*. arXiv. https://doi.org/10.48550/arXiv.2303.08774

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*(1), 1–4. https://doi.org/10.1207/S15326985EP3801_1

Pagura, J., Arjona, E., & Stenzl, C. (1994). Methodology for quality assessment of simultaneous interpretation. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation* (pp. 317–325). John Benjamins.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for*

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

Computational Linguistics (ACL 2002) (pp. 311–318). Association for Computational Linguistics. https://doi.org/10.3115/1073083.1073135

Papi, S., Bentivogli, L., Gaido, M., Karakanta, A., Negri, M., & Turchi, M. (2023). *Over-generation cannot be rewarded: Length-adaptive average lagging for simultaneous speech translation*. arXiv. https://doi.org/10.48550/arXiv.2309.16715

Pöchhacker, F. (2001). Quality assessment in conference and community interpreting. *Meta: Journal des traducteurs*, *46*(2), 410–425. https://doi.org/10.7202/004519ar

Pöchhacker, F. (2016). *Introducing interpreting studies* (2nd ed.). Routledge.

Rabinovich, E., Tsvetkov, Y., & Wintner, S. (2017). Personalized machine translation: Preserving original author traits. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2017)* (pp. 1074–1084). Association for Computational Linguistics. https://doi.org/10.18653/v1/E17-1101

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). *Robust speech recognition via large-scale weak supervision*. arXiv. https://doi.org/10.48550/arXiv.2212.04356

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training*. OpenAI. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

Rei, R., Stewart, C., Farinha, A. C., & Lavie, A. (2020). COMET: A neural framework for MT evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 2685–2702). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.213

Rinne, J. O., Tommola, J., Laine, M., Krause, B. J., Schmidt, D., Kaasinen, V., Teräs, M., & Solin, O. (2000). The translating brain: Cerebral activation patterns during simultaneous interpreting. *Neuroscience Letters*, *294*(2), 85–88. https://doi.org/10.1016/S0304-3940(00)01540-8

Rinsche, A., & Portera-Zanotti, N. (2009). *The size of the language industry in the EU* (DGT Studies). European Commission.

Sandrelli, A. (2021). Technology in conference interpreting: From AIIC's early concerns to the post-COVID digital revolution. *Translation, Cognition & Behavior*, *4*(2), 169–193.

Seeber, K. G. (2011). Cognitive load in simultaneous interpreting: Existing theories — new models. *Interpreting*, *13*(2), 176–204. https://doi.org/10.1075/intp.13.2.03see

Seeber, K. G., & Kerzel, D. (2012). Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism*, *16*(2), 228–242. https://doi.org/10.1177/1367006911403202

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

Seleskovitch, D. (1975). *Langage, langues et mémoire: Étude de la prise de notes en interprétation consécutive*. Minard Lettres Modernes.

Sellam, T., Das, D., & Parikh, A. (2020). BLEURT: Learning robust metrics for text generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)* (pp. 7881–7892). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.704

Signorelli, T. M., Haarmann, H. J., & Obler, L. K. (2012). Working memory in simultaneous interpreters: Effects of task and age. *International Journal of Bilingualism*, *16*(2), 198–212. https://doi.org/10.1177/1367006911403200

Sun, T., Gaut, A., Tang, S., Huang, Y., ElSherief, M., Zhao, J., Mirber, D., & Wang, W. Y. (2021). Mitigating gender bias in natural language processing: Literature review. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL 2021)*. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.acl-long.416

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems 27 (NeurIPS 2014)*. Curran Associates.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Tatman, R., & Kasten, C. (2017). Effects of talker dialect, gender and race on accuracy of Bing Speech and YouTube automatic captions. In *Proceedings of Interspeech 2017* (pp. 934–938). ISCA. https://doi.org/10.21437/Interspeech.2017-1746

Van Besien, F. (1999). Anticipation in simultaneous interpretation. *Meta: Journal des traducteurs*, *44*(2), 250–259. https://doi.org/10.7202/003686ar

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*. Curran Associates. https://doi.org/10.48550/arXiv.1706.03762

Viezzi, M. (1996). *Aspetti della qualità in interpretazione*. SERT, Scuola Superiore di Lingue Moderne per Interpreti e Traduttori.

Wordly User Research. (2023). *User satisfaction with AI-powered simultaneous interpretation in enterprise settings: Preliminary findings*. Wordly Inc.

Zhang, Y., Han, W., Qin, J., Wang, Y., Bapna, A., Chen, Z., Chen, N., & Wu, Y. (2023). *Google USM: Scaling automatic speech recognition beyond 100 languages*. arXiv. https://doi.org/10.48550/arXiv.2303.01037

©Copyright 2026 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.