https://doi.org/10.5007/2175-8026.2026.e112463

THE RELIABILITY TRAP IN AI-DRIVEN TRANSLATION: CONTEXT-BLINDNESS AND PRAGMATIC HAZARDS IN AUGMENTED REALITY

Michał Organ[1*]
[1]University of Rzeszów, Rzeszów, Poland

Abstract

This study investigates the reliability trap in Google Lens, an artificial intelligence-driven augmented reality translation tool. Viewing translation as human-computer interaction, this qualitative field study evaluates artificial intelligence algorithms in the Polish-to-English direction, for so-called incidental texts (text encountered in public spaces, such as signs, notices, and instructions). Analysis indicates that while neural networks deliver high grammatical fluency, their spatial context-blindness to physical geosemiotics causes severe pragmatic pitfalls affecting users, including safety, as well as legal and socio-economic risks. To address this, the paper advocates that traditional equivalence-based quality assessment must give way to a risk management paradigm. Ultimately, operating safely within the augmented linguistic landscape would entail fostering machine translation literacy and risk-aware artificial intelligence interface design.

Keywords: Artificial intelligence; Augmented reality translation; Machine translation; Machine translation literacy; Risk management.

*Michał Organ, PhD, is an assistant professor at the Chair of Translation Studies in the Department of English Studies at the University of Rzeszów (Rzeszów, Poland). His main research interests include translation technologies, Computer-Assisted Translation (CAT) tools, machine translation, and the application of artificial intelligence—particularly augmented reality—in translation and intercultural communication. Email: morgan@ur.edu.pl. ORCID: https://orcid.org/0000-0001-8072-7936.

Esta obra tem licença *Creative Commons*
http://creativecommons.org/licenses/by/4.0

1. Introduction

The digital age redefines communication, reorienting translation from a purely linguistic-cognitive endeavour into a form of human-computer interaction (HCI) marked by interdependence between human cognition and machine processing (O'Brien, 2012, p. 103). Technology is no longer a mere lifeless tool; rather, it constitutes a dynamic extension of human ecosystems that synergises transversal subjectivity and renders the physical world endlessly translatable (Cronin, 2020, p. 526). In this context, augmented reality (AR) can be conceptualized as an interactive technology that enriches the user's real-time perception of the physical world by superimposing computer-generated digital elements onto their actual environment (Azuma, 1997, pp. 355–356; Carmigniani et al., 2011, p. 342). Accordingly, mobile AR tools such as Google Lens (GL) serve as robust cognitive prostheses and ubiquitous access points, allowing users to nearly instantaneously decode their environments (Maurya et al., 2023, p. 113).

Effective and responsible engagement with this hybrid space requires machine translation (MT) literacy, particularly an understanding of how MT systems generate texts. However, insufficient MT literacy remains a widespread problem among both general users and academic researchers (Bowker & Buitrago Ciro, 2019, pp. 1, 88). This vulnerability is critically amplified by media narratives portraying artificial intelligence (AI) as an infallible, quasi-magical entity; such depictions induce questionable expectations of human-level technological performance, systematically predisposing users to place blind and uncritical trust in AR overlays (Vieira, 2020a, pp. 108–109).

This uncritical trust stands in stark contradiction to the pragmatic reality of the physical environment. One area in which the relationship between language and its physical and social context becomes particularly evident is public signage. Linguistic landscape (LL) studies approach public signs as multidimensional texts that actively and intentionally shape the sociolinguistic context (Cenoz & Gorter, 2006, p. 68; Shohamy & Gorter, 2009, p. 1). Furthermore, geosemiotics elucidates that a text's pragmatic force and regulatory meaning are intrinsically rooted in its physical emplacement and indexicality (e.g., the specific indoor or outdoor placement of a prohibition sign) (Blommaert, 2013, pp. 39–40, 43; Scollon & Scollon, 2003, pp. 29–31, 142). Consequently, a critical technological failure arises as automated AI vision systems detect, extract, and render isolated text strings, demonstrating a direct clash with multifaceted urban environments where communicative value and safety rely on excluded physical background.

2. The Pragmatic Nature of Incidental Texts

In this study, the term incidental text refers to a written message that users encounter unexpectedly while navigating a physical environment, rather than a text they have deliberately sought out. Such messages may include public warnings, notices, instructions, or informational leaflets. They are often brief and

serve an immediate pragmatic purpose, such as warning, informing, directing, or instructing the reader. For example, a foreign visitor using a mobile AR translation application may encounter a sign reading "Danger! Wet floor" and rely on its automatically generated translation to avoid a potentially hazardous situation. In this context, the sign functions not merely as a source of linguistic information but as a communicative act intended to influence the reader's behaviour. Its interpretation therefore depends on the relationship between the wording, the physical setting, and the intended communicative function. This functional perspective is consistent with earlier approaches to translation quality that emphasise the pragmatic functions of texts (House, 1977; 2015).

Although the term incidental text is used here to describe a particular situation of text reception, the communicative functions of such texts have long been recognised in translation studies. House (1977, p. 86; 2015, pp. 26, 63) emphasises the importance of textual function and situational context in translation quality assessment, while Desilla (2025, pp. 17–21) discusses the relevance of speech acts and other pragmatic phenomena to translation. The translation of public signage has also been investigated by Radtke and Yuan (2011), who examine problems associated with Chinese-English translations of signs in public spaces. The present study builds on these approaches by focusing on situations in which users encounter such texts unexpectedly and depend on mobile AR translation to understand and respond to them.

As such, the concept should not be confused with "incidental learning" in educational exercises, as theorised by Laufer and Hulstijn (2001, pp. 10–11), nor reduced to a mere photographic coincidence in computer vision, as postulated by Karatzas et al. (2015, p. 1158). Instead, incidental texts are characterised by communicative immediacy, time pressure, and epistemic opacity. They are texts that demand an instantaneous response, leave no time for careful deliberation, and necessitate deciphering their true meaning. Users purposefully employ AR technology to capture them and reveal what is at stake; however, lacking inherent verification capabilities, they are instantly compelled to trust MT output to mitigate cross-cultural pragmatic failure (Xiao et al., 2025).

Gile (2009, pp. 42, 159) discusses the limitations of cognitive resources in professional translation and interpreting, demonstrating how excessive cognitive demands can interfere with task performance. Although his analysis concerns trained language professionals, the underlying principle of limited cognitive capacity is also relevant to non-professional users of MT, who may face additional difficulties due to their limited proficiency in the source language. In an AR translation environment, these difficulties may become particularly acute when foreign users, such as tourists or migrant workers, must make immediate decisions without sufficient time, linguistic competence, or other resources to assess the reliability of MT output.

Immediate time pressure constrains cognitive processing, forcing users to abandon the analytical reflection characteristic of System 2 and, consequently,

relegating decision-making to the automatic, effortless, and involuntary System 1 (System 1 and System 2 within Daniel Kahneman's dual-process model of human cognition). To address the cognitive bottleneck, the human mind intuitively resorts to a mental shortcut – the 'What You See Is All There Is' (WYSIATI) heuristic. In such situations, the overloaded users establish a coherent narrative driven strictly by immediate visual feedback, while simultaneously disregarding the quality and quantity of the information (Kahneman, 2013, pp. 20–21, 86). This mechanism breeds overconfidence derived from fragmentary data and ultimately fosters technological overtrust.

3. The Reliability Trap

Cognitive depletion caused by the demands of spatial orientation in foreign environments forces users to abandon their own judgment and transfer perceptual authority to AR tools. This ceding of authority can be understood in terms of trust in automation, which J.D. Lee and See (2004, p. 51) define as an attitude involving the expectation that a technological system will help users achieve their goals despite uncertainty and vulnerability. Such trust may also be influenced by broader expectations associated with human communication, particularly the assumption that communicated information will be relevant (Wilson & Sperber, 2002, pp. 256–259). However, these expectations may become problematic when extended to machine-generated translations, whose linguistic fluency does not necessarily guarantee their accuracy or pragmatic appropriateness.

Under time pressure, users tend to minimise cognitive effort and rely on readily available information rather than engage in detailed analysis. This tendency, sometimes described as cognitive miserliness, may encourage automation bias, whereby users accept the output of automated systems without adequately questioning its accuracy (Mosier & Skitka, 2009, p. 205). In the context of AR translation, users may therefore treat machine-generated translations as reliable guides to their immediate surroundings, overlooking contextual information that could challenge their interpretation. Such reliance can reduce their active engagement with the environment and their ability to detect potential errors (Parasuraman & Manzey, 2010, pp. 382, 404). This may eventually result in an out-of-the-loop phenomenon, in which dependence on automation weakens users' awareness of the situation and makes it more difficult for them to recognise when the system provides misleading or inadequate information (Endsley, 2017, p. 8).

The architecture of modern AI systems affects AR translation as it exacerbates automation bias, directly leading to the reliability trap. Mobile AR tools integrate translated text into the user's view of the physical environment, matching its colours and perspective. This seamless presentation reduces the need for conscious scrutiny and makes the translated text appear more natural and reliable. It also exploits cognitive fluency, the psychological tendency to perceive information that is easy to process as more likely to be true (Alter & Oppenheimer, 2009, p. 219). The lack of friction generates cognitive ease and

drives System 1 to accept the AR overlay without rigorous scrutiny by employing an 'easy-means-true' heuristic. As a consequence, users completely suspend their analytical reflection, delegate their critical assessment to a device, and become susceptible to AI hallucinations, reacting to them as if they were physical realities.

This cognitive flaw is amplified by the speed paradox of AR translation. Conversational large language models (LLM) deliver responses through iterative prompting, which allows users to negotiate and specify translation purposes, consequently fostering cognitive engagement by allowing for doubts creating a temporal 'space for doubt' (Asscher, 2026, p. 52; Chang et al., 2024, p. 5; Sun & Muñoz Martín, 2025, p. 42). In contrast, AR applications such as GL function as near-real-time visual reflexes by generating instantaneous, seamless overlays that preclude negotiation. By their very design, AR interfaces do not offer opportunities for evaluation. Thus they suppress critical user engagement and inevitably entrench reliance on System 1 heuristics.

4. Context-Blindness

Comprehending the reliability trap entails acknowledging that physical signs do not operate as simple neutral backgrounds. Instead, their physical properties, including spatial orientation, shape, and colour, function as semiotic modes whose pragmatic force stems from their material emplacement (Kress & Van Leeuwen, 2001, p. 28; Scollon & Scollon, 2003, pp. 129–130, 142). This multimodal complexity is problematic for modern translation as human recipients reflexively combine these environmental cues to assess the stakes of the situation (O'Sullivan, 2013, pp. 2, 11).

Some scholars (e.g. Asscher, 2026, p. 51; Tieber, 2025, p. 221) point out that traditional functional approaches in translation studies, including Skopos theory, are inherently deficient in assessing modern translation technologies and automated systems, as they assume that translation is based on conscious, goal-oriented human cooperation and individualised transactions. Conversely, AI systems lack consciousness, intentionality and moral agency (Moor, 2006, p. 20). In fact, they function as mechanical black boxes, in which inanimate algorithms blindly and indifferently calculate statistical word probabilities devoid of any apprehension of communicative aims or the capacity to bear human accountability (Forcada, 2017, p. 296; UNESCO, 2022, para. 36). As a result, the theoretical models anchored in human agency are inadequate to account for this phenomenon.

Evaluation of real-world stakes should supersede psychological intention; therefore, translation is a decision-making process under uncertainty (Pym, 2015, pp. 70–71; Pym, 2025a, p. 1). While Pym emphasizes the risks and uncertainties involved in translation decisions, Gile (2021, p. 59) approaches the issue from a different perspective, highlighting human translators' concern with complying with norms and achieving specific goals. Paradoxically, this emphasises the exact limitations of AI: since these systems lack professional ethics and motivations,

and are deprived of consciousness, they must be assessed through the lens of risk understood as the probability of communicative failure (Pym, 2025b, p. 1). Unlike technological tools, human translators reflexively default to multimodal cues to optimally allocate cognitive effort, strategically expending more mental resources on high-stakes elements to prevent catastrophic pragmatic failure (Pym, 2025a, p. 18).

Mobile MT tools (e.g., Apple Live Text, DeepL, Google Lens or Microsoft Translator) employing optical character recognition (OCR) inherently undermine risk management due to context-blindness. As AR systems detect and single out linguistic characters from the environmental plane, they simultaneously discard multimodal risk markers, treating them as mere visual noise. By disregarding geosemiotic indexicality, AI operates within a complete situational vacuum, rendering it incapable of distinguishing a high-voltage fence from a crumpled leaflet in a puddle (Scollon & Scollon, 2003, p. 110).

The multidimensional quality metrics (MQM) framework, as a traditional MT quality assessment model, lacks the appropriate apparatus to address this danger. The loss of physical context is completely ignored in the industry norm, which entails the evaluation of isolated and decontextualised text segments (Doherty, 2017, p. 133; Lommel et al., 2014, p. 460; Vieira, 2020b, p. 327). As a result, machines allocate the processing load equally between life-threatening warnings and bubble gum stickers. Incapable of calibrating computational effort to different risks, their operational malfunctions are unpredictable and pervasive (Pym, 2025a, p. 19).

This danger is profoundly amplified by the very architecture of MT and generative AI systems (Sun et al., 2025, pp. 3–4). They are fundamentally misleading, producing grammatically correct outputs that stimulate overtrust (Pym & Hao, 2025, p. 6). Whereas older statistical MT provoked scepticism and low-trust reception due to obvious errors, the modern AI models offer seamless fluency, which translates into a high-trust regime; critical pragmatic failures masquerade as deceptively convincing lies, thereby outsourcing the onus of verification to the end user (Pym, 2025a, pp. 40, 44–45, 50).

Casting the user in a notary-like role – expected to authenticate and verify truth – exposes a complex ethical vacuum in AR translation. Divested of moral agency, these systems cannot bear legal responsibility, which invalidates them in high-stakes contexts necessitating human accountability. In turn, algorithmic inscrutability and the absence of legal responsibility render societies vulnerable by diluting accountability for pragmatic failures, such as the mistranslation of medical instructions that may lead to incorrect treatment (Monzó-Nebot, 2025, p. 187; Ren & Yin, 2021, p. 206; Tasa-Fuster, 2025, p. 162). By releasing context-blind systems, algorithmic gatekeepers delegate the ethical mandate to users, exploiting them to perform linguistic janitorial work to mitigate machine errors (Kelly, 2014). Deprived of professional verification capabilities, unprepared users facing profound communicative interferences are forcefully transformed into incidental translators, coerced to resolve urgent interlingual crises (Bowker,

2025a, p. 272; Tryuk, 2021, p. 398). This contravenes a basic risk management directive: high communicative risks require symmetrical investments in translator trustworthiness (Pym, 2021, p. 152). Fundamentally, AR translation democratizes the illusion of comprehension in tandem with a total shift of personal physical risk, saddling unaware users entirely with the responsibility for machine failures.

5. Methodology and Corpus Design

To empirically operationalize context-blindness and the reliability trap, this observational field study assessed how AR MT mediates incidental texts and triggers pragmatic hazards in situ. Aiming to maximise ecological validity and evaluate the algorithms under non-laboratory conditions, the investigation encompassed varied lighting, winter weather conditions, and complex real-world physical substrates. The photographic corpus was compiled across urban and semi-urban areas in Rzeszów, Poland, during February 2026.

The selection of the corpus, comprising 50 incidental texts, was driven by the epistemic opacity encountered by foreign users who are unable to determine the text's pragmatic stakes prior to decoding it. To reproduce this state, the sample embraces a broad spectrum of pragmatic risk, ranging from low-stakes informational leaflets and plaques to high-stakes safety warnings. The sampled texts manifest clear, embedded pragmatic functions – specifically directive, commissive, or informative. This deliberate heterogeneity provides a framework for evaluating AI systems to assess their capacity to calibrate computational effort across varying risk levels.

This qualitative study utilizes purposive sampling (Saldanha & O'Brien, 2014, p. 34) guided by predetermined critical parameters to propose a taxonomy of errors. Sampling continued until theoretical saturation (Guest et al., 2006, p. 65) was achieved, that is, when new examples yielded no novel pragmatic hazard categories. To demonstrate pragmatic density of the augmented linguistic landscape (ALL), the assembled corpus was categorized based on primary illocutionary force and risk potential (Table 1).

Pragmatic Category (Illocutionary Force)	Examples from the physical environment	Risk Level / Stakes	Number of Texts (N=50)
Hazard & Safety Warnings (Directives/Assertives)	Warning tapes, falling ice notices, biological hazard alerts (e.g., ASF), construction site hazards.	High (Physical Safety)	6
Regulatory & Prohibitive (Directives/Commissives)	Private property signs, parking gates, municipal bans.	High (Legal/Financial)	6
Commercial, Agricultural & Employment (Informative/Commissives)	Local produce boards, employment offers, rural services.	Medium/High (Socio-economic)	27
General Informative & Community (Informative)	Missing pet posters, cultural events, charity collections.	Low (Orientation)	11

Table 1. Pragmatic categorization of incidental texts.

The data distribution reflects the ecological reality of the investigated environment, dominated by commercial and agricultural discourse. By capturing this continuum, purposive sampling mitigates confirmation bias and identifies pivotal pragmatic hazards for a rigorous in-depth qualitative assessment.

5.1. Technological Apparatus and Procedure

To mitigate hardware limitations and avoid language misidentification, GL (v.1.18.250731009)[1] was employed on a Google Pixel 9 Pro XL, with the auto-detect function manually overridden to enforce a strict Polish-to-English translation direction. This chosen direction faithfully replicates the anticipated experience of a foreign user dependent on a global lingua franca.

The pair and translation direction were intentionally chosen to expose structural neural machine translation (NMT) deficiencies. Polish, as a morphologically rich and relatively low-resource language compared to global lingua francas, poses a complex challenge for NMT algorithms. As established in the literature, systems relying on byte pair encoding mechanisms to process such languages are inherently prone to hallucinating artificial, yet deceptively fluent outputs (K. Lee et al., 2018, p. 2; Macken et al., 2019, p. 68; Rothwell et al., 2023, pp. 105–106). While English privileges indirectness, Polish administrative discourse inclines towards blunt, authoritative tones, typically realized through impersonal constructions or bare infinitives (Wierzbicka, 2003, pp. 36, 64).

In the course of the study, GL AR translations overlaid on physical signs were visually documented via screenshots and further evaluated against the source texts (ST) to detect instances of context-blindness and assess pragmatic hazards. As users process AR overlays as a single communicative act, this research mirrors this perspective adopting an integrative approach to evaluate the holistic experience of the end-user. Consequently, the apparatus is perceived as a single agent, and therefore, the analysis does not artificially deconstruct errors within the black-box pipeline, nor does it determine whether they are caused by image capture, OCR, layout reconstruction, or NMT. Accordingly, the analysis of errors is founded on the final communicative outputs and the subsequent hazards they pose to the user.

5.2. Error Classification Framework

Traditional MT quality assessment metrics, predominantly focused on sentence-level linguistic equivalence, are inherently ill-equipped to evaluate physically embedded incidental texts. Hence, this research moves beyond standard error quantification and adopts a model anchored in pragmatics and risk management. By applying the notion of pragmatic failure (Thomas, 1983, p. 94), the analysis reorients away from prescriptive grammatical, syntactic and semantic errors to the failure to achieve intended communicative goals (Desilla, 2025, pp. 16–22). The gravity of AI hallucinations is governed by real-

world consequences rather than syntax. In this view, a linguistic error is critical only when it invalidates the main function of the text, whereas it is of lesser importance when affecting peripheral elements that do not compromise overall communicative success (Pym, 2015, p. 71). Consequently, AR overlay errors are assessed through the lens of risk, conceptualised as the estimated probability of communicative failure (Pym, 2025b, p. 1).

Departing from a purely frequency-driven model, this qualitative study proposes a typology of AR pragmatic hazards. Reflecting the interpretive nature of qualitative pragmatic analysis (Saldanha & O'Brien, 2014, pp. 22, 38–41), the categorization and annotation of errors were conducted directly by the researcher acting as an expert analyst. Methodological rigor and analytical validity were ensured through the systematic application of explicit pragmatic criteria (illocutionary point and real-world risk gravity) combined with theoretical data saturation (Guest et al., 2006, p. 65). The analysis focuses on six prototypical cases, purposively sampled from a saturated dataset, to demonstrate three distinct vectors of pragmatic failure, providing a rigorous conceptualization of spatial, legal, and socio-economic hazards faced by users of AR translation.

6. Case Studies

MT offered by AR tools effectively renders texts characterized by low morphological complexity and high semantic redundancy (Bowker & Buitrago Ciro, 2019, pp. 49–53; Rothwell et al., 2023, pp. 105–106), including standard agricultural notices, but at the same time systemically breaks down when confronted with high pragmatic density. The most critical algorithmic failures arise when strict directives or messages carrying high social stakes are highly elliptical and embedded in physical space (Scollon & Scollon, 2003, pp. 28–31; Thomas, 1983, p. 94).

In the case of AR translations, errors are not generated at random; rather they systematically invalidate pragmatic functions - the illocutionary force - within the LL. By moving beyond sentence-level frameworks like MQM, this research frames these MT hallucinations as pragmatic hazards, delineated exclusively by the communicative stakes they jeopardize for the end-user. Fundamentally, the three vectors of failure do not reproduce the initial risk level of the sign, but rather trace the definitive pragmatic impact of the failure, especially as generated hallucinations may escalate even low-stakes informational notices into critical socio-economic breakdowns. The findings uncover three distinct categories of pragmatic failure: the destruction of directives, pragmatic inversion, and geosemiotic erasure.

6.1. Category 1: The Destruction of Directives

Within the LL, regulatory signs and safety warnings constitute classic examples of directive speech acts, demanding recipient compliance through

a world-to-word direction of fit that requires physical execution to align the world with propositional content (Searle, 2012, pp. viii, 14). The validity of these directives is fundamentally contingent upon indexicality, as they derive their illocutionary force directly from material emplacement (Scollon & Scollon, 2003, p. 29). AR tools, afflicted by context blindness, decouple the text from its material bonds, engendering profound pragmatic failure which manifests not as grammatical inaccuracy, but rather as the target output's inability to realize the ST's real-world risk-mitigation purpose.

Example 1A: Directive Inversion



ST: Brama Wjazdowa. Nie zastawiać. [Entry Gate. Do not block.]



AR Translation Output: Entry Gate. Do not pay.
Emplacement: A notice on a garage gate.
Pragmatic Analysis: The MT entirely subverts the illocutionary force, altering a spatial prohibition into financial guidance. AR translation decouples the text from the physical gate and violates the principle of indexicality. The context-

blind algorithm retrieves the financial sense of the Polish verb "zastawiać" (to pawn/pledge) based on its training data, and produces a syntactically fluent but pragmatically misleading translation.

Pragmatic Hazard: The AR translation, offering a high cognitive fluency, invokes unwarranted trust and actively encourages the user to park their car and inadvertently violate traffic regulations. This translation outcome exposes the user to potential fines and the impoundment of the user's vehicle. The case provides an example of a critical commission error, wherein users reflexively follow incorrect automated directives.

Example 1B: Morphological Hallucination



ST: UWAGA SOPLE! [CAUTION ICICLES!]



AR Translation Output: ATTENTION ICILS!

Emplacement: A warning notice on the building façade, directly beneath an area where dangerous ice overhangs form.

Pragmatic Analysis: The MT system experiences a morphological breakdown caused by data sparsity, a fundamental flashpoint for hallucinations. By rendering the source plural noun "sople", the tool generates "icils" —a non-existent word that creates absolute epistemic opacity. From a cognitive processing standpoint,

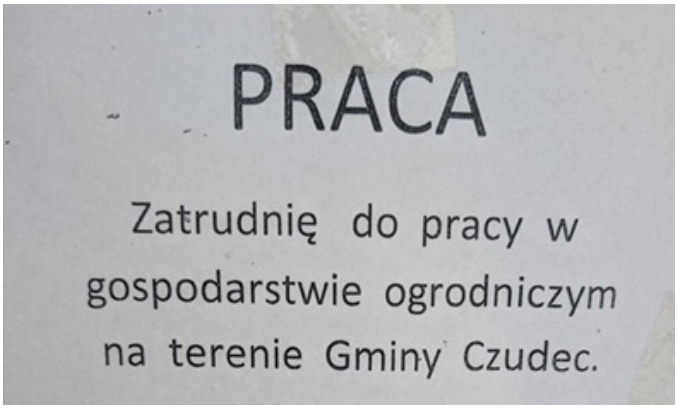
safety warnings should immediately activate the automatic, effortless System 1 to evoke a rapid risk-mitigation response.

Pragmatic Hazard: The non-existent target "icils" compels the user to engage the effortful, analytical System 2 to decipher the MT error. Even a short delay within the hazard may increase the risk of injury. A general pragmatic failure amplifies a safety-critical hazard by failing to deliver a real-time warning message.

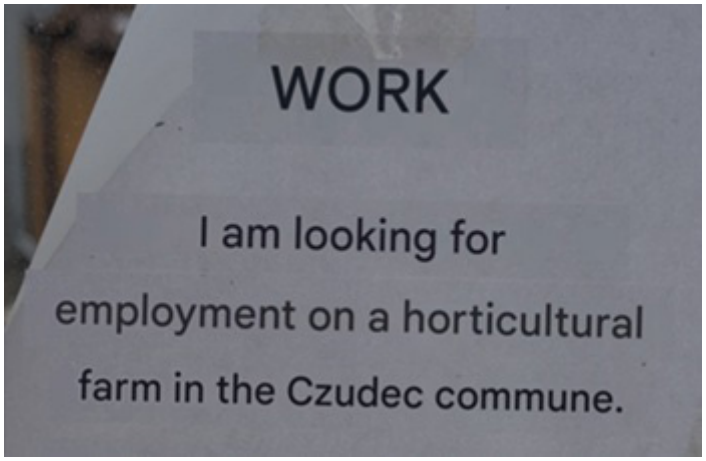
6.2. Category 2: Pragmatic Inversion

Category 2 explores the structural breakdown of commissive acts, that is, illocutions committing users to a future course of action, including promises and contracts, which function as mechanisms for socio-economic integration (Searle, 2012, p. 14). The generated hazard, namely pragmatic inversion, invalidates translation's moral and ethical imperative to facilitate mutual benefits through cooperation (Pym, 2025a, pp. 10–11). AR translation transcends mere linguistic non-equivalence, as it changes the direction of fit and misattributes agentive roles (Searle, 2012, pp. 3–4). Consequently, GL outputs diminish foreign users' interaction opportunities, thereby leading to their socio-economic exclusion.

Example 2A: The Reversal of Agency



ST: PRACA. Zatrudnię do pracy w gospodarstwie ogrodniczym... [WORK. I will hire for work in a horticultural farm...]



AR Translation Output: WORK. I am looking for employment on a horticultural farm...
Emplacement: A recruitment notice at a public bus stop.
Pragmatic Analysis: The MT engine precipitates failure of the communicative act, as it erroneously decodes the Polish verb "zatrudnię" – "I will hire". The complete inversion of the world-to-word direction of fit alters a potential livelihood opportunity into an announcement of potential market competition, directly inverting the roles of employer and job-seeker.
Pragmatic Hazard: The misinterpretation effectively paralyzes cooperation. High cognitive fluency drives user trust to a far greater extent than underlying semantic fidelity. As a consequence, the aesthetically coherent AR overlay functionally blocks the user from applying for the available job, and the neutralized commissive act results in a definitive collapse in potential mutual economic benefits.

Example 2B: The Hallucination of Crisis Resolution



ST: !!!ZAGINĄŁ KOT!!! To jest Loki. Mniejszy niż większość czarny kot z białymi wąsikami, oraz białym futerkiem na brzuszku i końcach łapek. Jeśli posiadasz informacje dotyczące jego miejsca przebywania, albo co się z nim stało, proszę zadzwoń na numer telefonu: [!!!CAT IS MISSING!!! This is Loki. A black cat smaller than most, with white whiskers, and white fur on his belly and the tips of his paws. If you have information regarding his whereabouts or what happened to him, please call the phone number:]



AR Translation Output: !!! THE CAT IS BACK!!! This is Loki. Smaller than most, this black cat has white whiskers and white fur on its belly and the tips of its paws. If you have any information regarding his whereabouts or what happened to him, please call the following number:

Emplacement: A community appeal posted on an electric pole.

Pragmatic Analysis: The MT generates a temporal and situational hallucination. While the AR output preserves visual urgency through the replication of exclamation marks, the hallucinated crisis resolution in the target output completely destroys the source illocutionary point, which is an expressive communal plea for vigilance.

Pragmatic Hazard: The optimistic target headline introduces a highly salient contextual assumption that fundamentally conditions how the subsequent text is cognitively processed. Guided by the relevance-comprehension heuristic via the path of least effort (Wilson & Sperber, 2002, pp. 259–260; Gutt, 2014, p. 28), the user processes the body text against this newly established false frame. Because the body text is rendered with high fluency, it induces cognitive ease and strengthens automation bias. Rather than prompting the reader to reject the hallucinated headline, the request regarding the cat's whereabouts is pragmatically re-interpreted as a secondary, retrospective inquiry about the animal's past movements during its absence. This derivation of a distorted contextual implicature drastically devalues the social urgency of the notice, leading passersby to dismiss the appeal and effectively excluding foreign-language residents from civic and neighbourly participation. Deprived of validation mechanisms, users are profoundly vulnerable to being misled by algorithmic deception stemming from MT hallucinations.

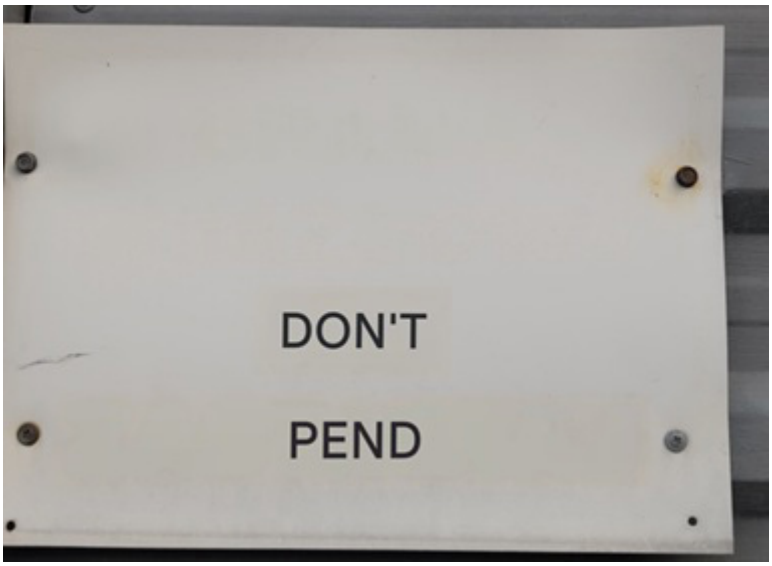
6.3. Category 3: Geosemiotic Erasure

Category 3 focuses on the destruction of spatial and cultural indexicality. Within the ALL, signs derive illocutionary force not strictly from their lexicon, but from precise physical emplacement and reliance on localized cultural scripts which govern unencoded meaning (Scollon & Scollon, 2003, pp. x, 142; Wierzbicka, 2003, pp. xvi–xvii). When the MT system abstracts the text from material and cultural coordinates, it precipitates geosemiotic erasure. This decoupling leads to epistemic opacity, a condition that confronts users with an archaic or culturally anomalous, or incongruous translated reality that makes the immediate physical environment completely illegible.

Example 3A: Boundary Disorientation



ST: Nie zastawiaj. [Do not block / Do not obstruct.]



AR Translation Output: Don't pend.
Emplacement: A directive on a private property gate.
Pragmatic Analysis: Confronted with a highly elliptical directive, the MT system commits a fatal lexical selection error by generating the English archaic verb "to pend" (meaning to hang/remain undecided). Violating the principle of indexicality, this non-functional message forces users attempting to decode it to transition from the intuitive, automatic System 1 and engage the effortful, analytical System 2.
Pragmatic Hazard: Public signs delineate spatial zones and define allowed actions within them. When lexical hallucinations neutralize these boundaries, they completely strip private owners of their property rights and regulatory authority. As a direct result, the user commits an error of omission by overlooking the prohibition, which simultaneously triggers a physical error of commission - blocking the gate. Finally, the algorithmic breakdown manifests as a spatial violation, exposing the user to the risk of inadvertent interpersonal conflict and potential financial penalties.

Example 3B: Cultural Misalignment



ST: Owies Bingo po Centrali Cena 80 zł/100kg Nienadowa... [Bingo Oats after the Central (Seed Station) Price 80 PLN/100kg Nienadowa (Village)...]



AR Translation Output: Oat Bingo after Centralia Price: PLN 80/100kg Unpleasant...
Emplacement: A localized advertisement for agricultural products posted near a bus stop.
Pragmatic Analysis: The MT misinterprets the Polish agricultural shorthand "po Centrali". In local farming parlance, this ellipsis denotes certified grain procured from a central seed station, literally phrased as "after the Central." The algorithm's literal rendering demonstrates severe context-blindness when confronted with specialized cultural scripts. Driven by anglocentric training data, the system introduces a foreign element absent from the ST – namely, Centralia, a famous American ghost town frequently associated with the Silent Hill horror franchise. Additionally, the MT strips the village name, "Nienadowa", of its proper noun status by translating it as "Unpleasant". This specific hallucination lacks any discernible semantic or morphological motivation, rendering the entire augmented message utterly absurd.
Pragmatic Hazard: The potential commercial transaction is thwarted by the replacement of geographical data and quality indicators with dystopian hallucinations and pejorative descriptors. The eradication of spatial indexicality elicits an affective rejection of the offer included in the ST. Consequently, algorithmic noise overrides the communicative goal, resulting in total pragmatic failure.

7. Discussion

A fundamental dichotomy in HCI manifests within the ALL: an inverse correlation between grammatical fluency and pragmatic reliability. Although GL delivers morphologically sound and syntactically fluent outputs, such linguistic perfection gives rise to the reliability trap - a condition in which high textual fluency conceals severe context-blindness and geosemiotic erasure. Through the analysis of automation bias (7.1), geosemiotic erasure (7.2), high-stakes pragmatic hazards (7.3), and societal impacts and erosion of human agency (7.4), this study ultimately advocates for MT literacy.

7.1. Automation Bias and the Ergonomics of Trust

Situating AR translation within HCI underscores the importance of cognitive ergonomics, which investigates how technological interfaces influence mental processing and cognitive load (O'Brien, 2012, p. 103). In the context of global mobility, the seamless interface design of GL activates System 1, prompting users to cede unquestioned authority to the application.

This dynamic promotes a hazardous state of automation bias, a psychological tendency to omit information gathering, defaulting to generated outputs as cognitive shortcuts (Mosier & Skitka, 2009, p. 205). This algorithmic hyper-dependence materializes in cascading failure trajectories. Initially, users commit

errors of omission by overlooking contradictory geosemiotic cues, which subsequently pave the way for errors of commission as they reflexively follow the application's false directives (Skitka et al., 1999, p. 993) - a pattern illustrated in Case Study 3A, where a hallucination masks the boundary, causing gate obstruction.

The hallucinatory façade of NMT (Koehn & Knowles, 2017, p. 30) reinforces this bias through the illusion of fluency, capitalizing on the cognitive tendency to prioritize grammatical smoothness over meaning accuracy (Martindale and Carpuat, 2018, p. 22). Indeed, although these technological tools propose grammatically flawless AR overlays, AI systems inherently lack the structural capacity to render the actual pragmatic intent of the ST (Sheng & Kong, 2023, as cited in Sun et al., 2025, p. 5). Since the system is prone to generating highly deceptive, fluently inadequate translations (Martindale et al., 2019, p. 233), the application's superficial flawlessness serves as a dangerous cognitive trap.

7.2. Geosemiotic Erasure and the Flattening of Superdiversity

Transcending the cognitive illusion of fluency, geosemiotics explains what is erased within physical space. Automation bias fosters a technological dependence that isolates users form physical and temporal dimensions of meaning. As incidental texts derive their illocutionary force directly from their indexicality, the overlays generated by GL trigger geosemiotic erasure (Scollon & Scollon, 2003, p. 29). Detaching text from its material anchors and rendering it as isolated data strings dismantles the semiotic nexus - a unique intersection of discourses in time and space - which gives the signs their true meaning (Blommaert, 2013, p. 28; Scollon & Scollon, 2004, p. 19).

Context-blindness undermines the historical dimension of space. In complex environments, signs point backward to their origins and forward to their future uptake through the double arrow of indexicality (Blommaert, 2013, p. 44). The discussed commercial notice illustrates a deeply situated chronotope where semiotization actively embeds space in history, signifying the agricultural quality of the seed (Blommaert, 2013, p. 51). The application, blind to this retrospective grounding, compensates for data deficiency by resorting to the globally dominant training models (Rothwell et al., 2023, p. 106). The rendition of culturally bound jargon into "after Centralia" - which accidentally implies an American ghost town and a horror franchise – constitutes a profound instance of algorithmic delocalization. GL substitutes a situated cultural script with a globalized hallucination (Wierzbicka, 2003, pp. xvi–xvii).

Accordingly, the AR system flattens superdiversity by imposing a homogenized interpretative framework onto nuanced local realities (Blommaert, 2013, p. 51). By decoupling texts from authentic localized chronotopes, the system subjects users to topographic hallucinations that elicit an affective rejection. Ultimately, a tool designed to facilitate text comprehension transforms into an instrument of geosemiotic and socio-economic exclusion.

7.3. High-Stakes Hazards: From Equivalence to Risk Management

Automation bias and geosemiotic erasure underscore major limitations in MT quality assessment. Traditional frameworks, including MQM, reduce errors to mere localized deviations from predefined specifications (Lommel et al., 2014, pp. 459–460). Within these frameworks, an AR output demonstrating flawless grammar but reversed pragmatic intent might be assessed with only minor semantic penalties, thereby yielding an unjustifiably high quality score.

This paradigm is deficient for AR, in which translation transcended the boundaries of a simple linguistic transfer to function as a critical tool for legal, socio-economic, and spatial navigation. Therefore, AR translation should be evaluated through the lens of risk management, where risk is understood as the probability of a communicative act failing to achieve its cooperative goals (Pym, 2025a, pp. 1–2). This approach is further underscored by differences in translation stakes. While mistranslations in literary texts may cause low-stakes aesthetic losses, the ALL constitutes an inherently high-stakes environment where errors result in real-world consequences determined by the gravity of the failure (Pym, 2025a, p. 4).

When deployed to decode public directives, GL constitutes a form of safety-critical translation (Canfora & Ottmann, 2018). The resulting pragmatic hazards are not mere benign linguistic drifts. The failures in Category 1 expose users to physical threats or legal violations, and as such, they align directly with Canfora and Ottmann's (2018) taxonomy of damages. Furthermore, the failures presented in Categories 2 and 3 preclude socio-economic cooperation and elicit an affective rejection.

Fundamentally, the reliability trap necessitates a paradigm shift in translation evaluation. AR systems should be evaluated based on their capacity to mitigate pragmatic risk rather than their ability to generate fluent sentences. A grammatically flawed translation that successfully communicates high-stakes information represents successful risk mitigation. Conversely, a fluent translation that fails to convey a risk, obscures a hazard or inverts a directive represents a profound failure of risk management. Thus, in the AR-mediated translation, linguistic fluency must be subordinated to pragmatic adequacy.

7.4. Societal Impacts and the Erosion of Human Agency

Moving beyond simple algorithmic glitches, the failures of AR signify a deep structural shift leading to the erosion of human agency. Although AR tools dictate efficiency-driven workflows, they simultaneously degrade the cognitive and reflective dimensions of interaction. The fluency offered by MT deceives users, compelling them to relinquish critical reflection and cede social and spatial agency to the context-blind application. Therefore, users operating within ALL are not autonomous subjects, but rather passive recipients of automated heuristics

that obscure the fundamentally human core of communication (Pietrzak, 2025, p. 2).

The collected incidental texts empirically substantiate this marginalization. Traditionally, critical algorithmic failures were confined to formal legal or medical domains (Vieira et al., 2021, p. 1516); however, the research corpus demonstrates that AR prostheses introduce high-stakes hazards into everyday public spaces. Directive inversions generated by GL expose users to legal and financial penalties, geosemiotic erasures generate physical safety hazards, and hallucinated associations compromise economic opportunities and social integration. Ultimately, the reliability trap constitutes not a mere technological inconvenience, but an invisible wall that hinders safety, legal adherence, economic prospects, and social integration.

Assessing these impacts requires a shift from instrumental research on technical efficiency (Dizdar, 2014, as cited in Baumgarten & Tieber, 2025, p. 2) toward consequentialist ethics that trace the broader social ripple effects. Evaluating AR technology solely through the lens of immediate accuracy of isolated translations should therefore be replaced by a long-term assessment of its networked implications for societies (Baumgarten & Tieber, 2025, p. 2; Jonas, 1984). From this perspective, the reliability trap materializes as a profound social phenomenon, as GL promotes blind trust in technology corporations and degrades human mediation. The AR application extends far beyond the mere generation of linguistic errors; it actively constructs a reality of geosemiotic and socio-economic marginalization.

8. Conclusion

The revolution of the digital age and AI necessitates a redefinition of translation - namely the rejection of isolated linguistic models in favor of a complex HCI paradigm, where mobile AR systems function as cognitive prostheses for decoding physical surroundings. However, this symbiosis introduces a critical fragility - the reliability trap. This phenomenon manifests in scenarios where context-blind algorithms process superdiverse physical chronotopes, as superficial grammatical fluency camouflages geosemiotic erasure and pragmatic hazards.

As has been demonstrated by the preceding analysis, the reliability trap exposes the fundamental inadequacies of traditional, equivalence-based quality assessments. These empirical findings directly corroborate Pym's (2025a, p. 17; 2025b, p. 4) theoretical assertion that evaluating automated translation necessitates a decisive shift towards a risk management paradigm. The discussed case studies indicate that AR translation errors are not defined merely by a lack of linguistic equivalence, but rather by their potential to trigger a failure of the entire communicative act (Pym, 2025a, p. 2). In this view, evaluating the risk of pragmatic harm and the corresponding probability of communicative failure

(Pym, 2025a, p. 11) should supersede formal equivalence as the fundamental non-essentialist guide for translation assessment in augmented environments.

While delineating the theoretical framework for AR translation, certain limitations of this study must be acknowledged. First, this qualitative study favours analytical depth over quantitative generalizability and is rooted in a specific geosemiotic and linguistic context. Secondly, the rapidly evolving nature of NMT implies that certain algorithmic errors may be resolved by software updates. Nevertheless, context-blindness and users' tendency towards automation bias will persist as intractable challenges in HCI.

Further limitations stem from the reliance on GL, a single commercial platform delivering translations based on proprietary OCR and NMT architectures. Future studies should investigate whether similar pragmatic hazards are present in other AR translation ecosystems. Due to the rapid development of technologies, research should also encompass LLMs, as their multimodal capabilities allow for direct processing of visual inputs (Chang et al., 2024, p. 26). As pointed out by Chang et al. (2024, p. 15), LLMs frequently produce coherent yet factually inaccurate statements that are entirely ungrounded in reality. In the context of cross-lingual tasks, Asscher (2026, p. 79) argues that this problem is even more pronounced than in traditional NMT, systematically resulting in mistranslations that are semantically unrelated to the ST. Furthermore, as described by Shi et al. (2025, p. 144), such models are prone to generating "extrinsic hallucinations" by embellishing or inventing unverified information when confronted with ambiguous content. Collectively, these findings confirm that the discussed reliability trap and superficial grammatical fluency pose a universal challenge reaching far beyond NMT architectures.

Recognising these risks and limitations foregrounds a pedagogical imperative: the development of MT literacy specifically adapted to OCR-based translation of incidental texts. Such literacy should involve more than a general awareness of algorithmic limitations. As Bowker (2025b, p. 438) contends, MT literacy must transcend purely techno-procedural competencies to cultivate the critical judgment required to evaluate whether an automated tool is appropriate for a specific communicative context. Users need to understand that OCR may incorrectly recognise, omit or segment characters and words, while subsequent MT processing may produce fluent but pragmatically misleading interpretations. They should also recognise that the meaning of public signs depends not only on their linguistic content but also on their physical location, visual presentation and communicative purpose. In practical terms, MT literacy entails checking whether the translated message is consistent with the surrounding environment, paying particular attention to negation, prohibitions, warnings and other expressions that directly influence behaviour. When a translation appears ambiguous or contradicts visible contextual cues, users should refrain from treating it as authoritative and seek alternative means of verification, particularly in high-stakes situations. These competencies should be supported by interface features that preserve access to the ST and make potential recognition or

translation uncertainties visible. Thus, effective MT literacy involves not merely understanding how automated translation works, but knowing when its output can reasonably be relied upon and when further verification is necessary.

The neutralization of the reliability trap should not be completely relegated to end-users; the promotion of MT literacy must be coupled with the ethical design of HCI. Modern AR translation tools are developed to offer a seamless user experience by fluently superimposing perfectly integrated visual overlays on the ST. This very seamlessness engenders a black-box problem, as the interface actively conceals the collaborative character of the translation process (O'Brien, 2012, p. 117). The lack of transparency is ethically problematic. As pointed out by Lee and See (2004, pp. 55–56), proper trust calibration requires aligning user trust with the actual capabilities and limitations of the system; concealing algorithmic uncertainty produces miscalibrated trust (specifically overtrust). This imbalance fosters complacency and automation bias, ultimately eroding users' critical cognitive reflection (Parasuraman & Manzey, 2010, p. 403).

The mitigation of high-stakes pragmatic hazards is highly dependent on AI developers, who should reject rigid, dehumanizing design approaches and embrace interfaces promoting appropriate trust calibration (Lee & See, 2004, p. 55; O'Brien, 2012, pp. 108–109). This might be achieved by introducing epistemic friction into the AR interface - for instance, by retaining the ST, displaying visual confidence scores, or integrating algorithmic warnings. As technological design expresses the values of its developers, they should recognize that the ergonomics and usability of these tools affect societal vulnerabilities and power structures (De la Cruz Recio, 2025, p. 106). A genuinely human-centered augmented translation framework should not hide the technology's limitations but openly communicate them to empower human agency.

Ultimately, cross-cultural mediation in the era of AI mandates a shift towards augmented translation - a paradigm in which users actively interact with the system and employ their language skills to improve, rather than passively consume, automated suggestions (Pym & Hao, 2025, p. 153). By ensuring that human agency actively governs generated outputs, we will be able to secure the ALL from the risk of the flattening of superdiversity. In doing so, AI will emerge as a safe tool that genuinely facilitates translation, instead of functioning as a hidden reliability trap.

DATA AVAILABILITY STATEMENT: Data available upon contact with corresponding author.

End notes

1. Commercial MT systems function as black-box mechanisms dependent upon perpetual, opaque server updates that endanger computational reproducibility by precluding the consistent replication of results (Chan et al., 2020, p. 286; Chan et al., 2024, p. 17). By specifying the exact software version, this study acknowledges these limitations; therefore, it intentionally does not position its findings as universal computational axioms, but rather as a synchronic snapshot of technological capabilities at a particular point in their evolution.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the Tribes of Fluency to Form a Metacognitive Nation. *Personality and Social Psychology Review, 13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Asscher, O. (2026). *Machine Translation and Translation Theory.* Routledge. https://doi.org/10.4324/9781003607977

Azuma, R. T. (1997). A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments*, *6*(4), 355–385. https://doi.org/10.1162/pres.1997.6.4.355

Baumgarten, S., & Tieber, M. (2025). Translation Technology and Machine Agency in Contemporary Multilingual Societies. In S. Baumgarten & M. Tieber (Eds.), *The Routledge Handbook of Translation Technology and Society* (pp. 1–18). Routledge. https://doi.org/10.4324/9781003271314-1

Blommaert, J. (2013). *Ethnography, Superdiversity and Linguistic Landscapes: Chronicles of Complexity.* Multilingual Matters.

Bowker, L. (2025a). Risks for Lay Users in Machine Translation and Machine Translation Literacy. In E. Monzó-Nebot & V. Tasa-Fuster (Eds.), *The Social Impact of Automating Translation An Ethics of Care Perspective on Machine Translation* (pp. 60–76). Routledge. https://doi.org/10.4324/9781003465522-4

Bowker, L. (2025b). The Need for Machine Translation Literacy. In S. Baumgarten & M. Tieber (Eds.), *Routledge Handbook of Translation Technology and Society* (pp. 436–447). Routledge. https://doi.org/10.4324/9781003271314-38

Bowker, L., & Buitrago Ciro, J. (2019). *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community* (First edition). Emerald Publishing.

Canfora, C., & Ottmann, A. (2018). Of Ostriches, Pyramids, and Swiss cheese: Risks in Safety-critical Translations. *Translation Spaces*, *7*(2), 167–201. https://doi.org/10.1075/ts.18002.can

Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., & Ivkovic, M. (2011). Augmented Reality Technologies, Systems and Applications. *Multimedia Tools and Applications, 51*(1), 341–377. https://doi.org/10.1007/s11042-010-0660-6

Cenoz, J., & Gorter, D. (2006). Linguistic Landscape and Minority Languages. In D. Gorter (Ed.), *Linguistic Landscape. A New Approach to Multilingualism* (pp. 67–80). Multilingual Matters.

Chan, C.-H., Tim Schatto-Eckrodt, & Johannes Gruber. (2024). What Makes Computational Communication Science (ir)reproducible? *Computational Communication Research, 6*(1), 1–30. https://doi.org/10.5117/CCR2024.1.5.CHAN

Chan, C.-H., Zeng, J., Wessler, H., Jungblut, M., Welbers, K., Bajjalieh, J. W., Van Atteveldt, W., & Althaus, S. L. (2020). *Reproducible Extraction of Cross-lingual Topics (rectr). Communication Methods and Measures, 14*(4), 285–305. https://doi.org/10.1080/19312458.2020.1812555

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., & Xie, X. (2024). *A Survey on Evaluation of Large Language Models. ACM Transactions on Intelligent Systems and Technology, 15*(3), 1–45. https://doi.org/10.1145/3641289

Cronin, M. (2020). Translation, Technology and Climate Change. In M. O'Hagan (Ed.), *The Routledge Handbook of Translation and Technology* (pp. 516–530). Routledge.

De la Cruz Recio, M. (2025). Translation and Critical Theories of Translation Technology. In S. Baumgarten & M. Tieber (Eds.), *The Routledge Handbook of Translation Technology and Society* (pp. 100–112). Routledge. https://doi.org/10.4324/9781003271314-8

Desilla, L. (2025). *Translation and Pragmatics: Theories and Applications.* Routledge. https://doi.org/10.4324/9781003213178

Doherty, S. (2017). Issues in Human and Automatic Translation Quality Assessment. In D. Kenny (Ed.), *Human Issues in Translation Technology* (pp. 131–148). Routledge.

Endsley, M. R. (2017). From Here to Autonomy: Lessons Learned From Human–Automation Research. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 59*(1), 5–27. https://doi.org/10.1177/0018720816681350

Forcada, M. L. (2017). Making Sense of Neural Machine Translation. *Translation Spaces, 6*(2), 291–309. https://doi.org/10.1075/ts.6.2.06for

Gile, D. (2009). *Basic Concepts and Models for Interpreter and Translator Training* (Revised edition). John Benjamins Pub. Co.

Gile, D. (2021). Risk Management in Translation: How Much Does It Really Explain? *International Journal of Interpreter Education, 13*(1), 56–65. https://doi.org/10.34068/ijie.13.01.08

Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods, 18*(1), 59–82. https://doi.org/10.1177/1525822X05279903

Gutt, E. A. (2014). *Translation and Relevance. Cognition and Context.* Routledge.

House, J. (1977). *Translation Quality Assessment. A Model Revisited.* Gunter Narr.

House, J. (2015). *Translation Quality Assessment: Past and Present*. Routledge.

Jonas, H. (1984). *The Imperative of Responsibility. In Search of an Ethics for the Technological Age.* The University of Chicago Press.

Kahneman, D. (2013). *Thinking, Fast and Slow.* Random House.

Karatzas, D., Gomez-Bigorda, L., Nicolaou, A., Ghosh, S., Bagdanov, A., Iwamura, M., Matas, J., Neumann, L., Chandrasekhar, V. R., Lu, S., Shafait, F., Uchida, S., & Valveny, E. (2015). I*CDAR 2015 Competition on Robust Reading.* 2015 13th International Conference on Document Analysis and Recognition (ICDAR), 1156–1160. https://doi.org/10.1109/ICDAR.2015.7333942

Kelly, N. (2014). *Why So Many Translators Hate Translation Technology*. HuffPost. https://www.huffpost.com/entry/why-so-many-translators-h_b_5506533

Koehn, P., & Knowles, R. (2017). Six Challenges for Neural Machine Translation. *Proceedings of the First Workshop on Neural Machine Translation,* 28–39. https://doi.org/10.18653/v1/W17-3204

Kress, G. R., & Van Leeuwen, T. (2001). *Multimodal Discourse: The Modes and Media of Contemporary Communication Discourse.* Arnold.

Laufer, B., & Hulstijn, J. (2001). Incidental Vocabulary Acquisition in a Second Language: The Construct of Task-induced Involvement. *Applied Linguistics, 22*(1), 1–26. https://doi.org/10.1093/applin/22.1.1

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lee, K., Firat, O., Agarwal, A., Fannjiang, C., & Sussillo, D. (2018). *Hallucinations in Neural Machine Translation*. 1–18.

Lommel, A., Uszkoreit, H., & Burchardt, A. (2014). Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. *Tradumàtica Tecnologies de La Traducció*, (12), 455–463. https://doi.org/10.5565/rev/tradumatica.77

Macken, L., Van Brussel, L., & Daems, J. (2019). NMT's Wonderland where People Turn into Rabbits. A Study on the Comprehensibility of Newly Invented Words in NMT Output. *Computational Linguistics in the Netherlands Journal, 9*, 67–80.

Martindale, M. J., & Carpuat, M. (2018). Fluency Over Adequacy: A Pilot Study in Measuring User Trust in Imperfect MT. *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas, 1,* 13–25.

Martindale, M. J., Carpuat, M., Duh, K., & McNamee, P. (2019). Identifying Fluently Inadequate Output in Neural and Statistical Machine Translation. *Proceedings of Machine Translation Summit XVII: Research Track, 1*, 233–243.

Maurya, A., Kumar, A., & Alimohammadi, D. (2023). Application of Google Lens to Promote Information Services beyond the Traditional Techniques. *Qualitative and Quantitative Methods in Libraries, 12(1)*, 111–136.

Monzó-Nebot, E. (2025). In Different Voices. The Roles of Public Machine Translation Studies. In E. Monzó-Nebot & V. Tasa-Fuster (Eds.), *The Social Impact of Automating Translation. An Ethics of Care Perspective on Machine Translation* (pp. 176–205). https://doi.org/10.4324/9781003465522-10

Moor, J. H. (2006). The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems, 21*(4), 18–21. https://doi.org/10.1109/MIS.2006.80

Mosier, K. L., & Skitka, L. J. (2009). Human Decision Makers and Automated Decision Aids: Made for Each Other? In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance. Theory and Applications* (pp. 201–220). CRC Press.

O'Brien, S. (2012). Translation as Human–Computer Interaction. *Translation Spaces, 1*, 101–122. https://doi.org/10.1075/ts.1.05obr

O'Sullivan, C. (2013). Introduction: Multimodality as Challenge and Resource for Translation. *The Journal of Specialised Translation, 20,* 2–14. https://doi.org/10.26034/cm.jostrans.2013.398

Parasuraman, R., & Manzey, D. H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Pietrzak, P. (2025). Personal Resources in AI-Assisted Translation. *Research in Language, 23,* 1–10. https://doi.org/10.18778/1731-7533.23.01

Pym, A. (2015). Translating as Risk Management. *Journal of Pragmatics, 85,* 67–80. https://doi.org/10.1016/j.pragma.2015.06.010

Pym, A. (2021). Translator Ethics. In K. Koskinen & N. Kocijančič Pokorn (Eds.), *Routledge Handbook of Translation and Ethics* (pp. 147–161). Routledge.

Pym, A. (2025a). *Risk Management in Translation.* Cambridge University Press. https://doi.org/10.1017/9781009546836

Pym, A. (2025b). *Risk Management in Translation: A Reply to Critics*. https://doi.org/10.17613/060CY-AMP18

Pym, A., & Hao, Y. (2025). *How to Augment Language Skills. Generative AI and Machine Translation in Language Learning and Translator Training*. Routledge. https://doi.org/10.4324/9781032648033

Ren, W., & Yin, M. (2021). Conference Interpreter Ethics. In K. Koskinen & N. Kocijančič Pokorn (Eds.), *Routledge Handbook of Translation and Ethics* (pp. 195–210). Routledge.

Radtke, O., & Yuan, X. (2011). 'Please Don't Climb Trees and Pick Flowers for the Sake of Life' – Making Sense of Bilingual Tourism Signs in China. *Language and Intercultural Communication, 11*(4), 389-407. https://doi.org/10.1080/14708477.2011.616592

Rothwell, A., Moorkens, J., Fernández-Parra, M., Drugan, J., & Austermuehl, F. (2023). *Translation Tools and Technologies.* Routledge. https://doi.org/10.4324/9781003160793

Saldanha, G., & O'Brien, S. (2014). *Research Methodologies in Translation Studies.* Routledge. https://doi.org/10.4324/9781315760100

Scollon, R., & Scollon, S. W. (2003). *Discourses in Place: Language in the Material World.* Routledge.

Scollon, R., & Scollon, S. W. (2004). *Nexus Analysis: Discourse and the Emerging Internet.* Routledge.

Searle, J. R. (2012). *Expression and Meaning: Studies in the Theory of Speech acts.* Cambridge University Press. https://doi.org/10.1017/CBO9780511609213

Shi, Y., Xu, H., Kwok, H. L., Liu, K., Sun, S., Liu, K., & Moratto, R. (2025). ChatGPT in Professional Translation. A Double-Edged Sword – Insights from Chinese Translators on Capabilities, Concerns, and Future Prospects. In T*ranslation Studies in the Age of Artificial Intelligence* (pp. 125–149). Routledge. https://doi.org/10.4324/9781003482369-7

Shohamy, E., & Gorter, D. (2009). Introduction. In *Linguistic Landscape. Expanding the Scenery* (pp. 1–10). Routledge.

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does Automation Bias Decision-Making? *International Journal of Human-Computer Studies, 51*(5), 991–1006. https://doi.org/10.1006/ijhc.1999.0252

Sun, S., Liu, K., & Moratto, R. (2025). Introduction. Navigating the Paradigm Shift–Translation Studies in the Age of AI. In S. Sun, K. Liu, & R. Moratto (Eds.), *Translation Studies in the Age of Artificial Intelligence* (pp. 1–17). Routledge. https://doi.org/10.4324/9781003482369-1

Sun, S., & Muñoz Martín, R. (2025). Reframing Translation Expertise for the AI Era. In S. Sun, K. Liu, & R. Moratto (Eds.), *Translation Studies in the Age of Artificial Intelligence* (pp. 42–61). Routledge. https://doi.org/10.4324/9781003482369-3

Tasa-Fuster, V. (2025). Democracy, Artificial Intelligence, and Machine Translation Some Legal Challenges. In E. Monzó-Nebot & V. Tasa-Fuster (Eds.), *The Social Impact of Automating Translation. An Ethics of Care Perspective on Machine Translation* (pp. 147–162). https://doi.org/10.4324/9781003465522-8

Thomas, J. (1983). Cross-Cultural Pragmatic Failure. *Applied Linguistics, 4*(3), 91–112. https://doi.org/10.1093/applin/4.2.91

Tieber, M. (2025). (Re)Conceptualising Translation in the Age of the Machine. In S. Baumgarten & M. Tieber (Eds.), *The Routledge Handbook of Translation Technology and Society* (pp. 219–232). Routledge.

Tryuk, M. (2021). Translating and Interpreting in Conflict and Crisis. In K. Koskinen & N. Kocijančič Pokorn (Eds.), *The Routledge Handbook of Translation and Ethics* (pp. 398–414). Routledge.

UNESCO. (2022). *Recommendation on the Ethics of Artificial Intelligence* (pp. 1–44). UNESCO.

Vieira, L. N. (2020a). Machine Translation in the News: A Framing Analysis of the Written Press. *Translation Spaces, 9*(1), 98–122. https://doi.org/10.1075/ts.00023.nun

Vieira, L. N. (2020b). Post-editing of Machine Translation. In M. O'Hagan (Ed.), *The Routledge Handbook of Translation and Technology* (pp. 319–335). Routledge.

Vieira, L. N., O'Hagan, M., & O'Sullivan, C. (2021). Understanding the Societal Impacts of Machine Translation: A Critical Review of the Literature on Medical and Legal Use Cases. I*nformation, Communication & Society, 24*(11), 1515–1532. https://doi.org/10.1080/1369118X.2020.1776370

Wierzbicka, A. (2003). *Cross-Cultural Pragmatics: The Semantics of Human Interaction* (2nd ed. 2003). Mouton de Gruyter. https://doi.org/10.1515/9783110220964

Wilson, D., & Sperber, D. (2002). *Relevance Theory*. UCL Working Papers in Linguistics, 14, 249–290.

Xiao, Y., Zhang, Y., Ki, D., Bao, C., Martindale, M. J., Vaughn, C., Gao, G., & Carpuat, M. (2025). *Toward Machine Translation Literacy: How Lay Users Perceive and Rely on Imperfect Translations.* Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, 33997–34014.

Submission date: 04/05/2026
Acceptance date: 08/10/2026
Editors: Lincoln P. Fernandes, Dimitris Asimakoulas, Ewerton Gleison Lopes Branco