CZU: 81’322.4:004.8:174:001.89 https://doi.org/10.59295/teia2025.09

THE TRIPARTITE CHALLENGE: ETHICS, QUALITY, AND RESPONSIBILITY IN AI TRANSLATION

Irina BAZAROVA

Free International University of Moldova (ULIM)

https://orcid.org/0000-0002-2446-3005

The rapid integration of Artificial Intelligence (AI) and Large Language Models (LLMs) into the language services industry has created a critical tripartite challenge concerning ethics, quality, and responsibility in translation. This paper, grounded in recent academic literature from 2023 to 2025, analyzes the emergent ethical conflicts surrounding data usage and algorithmic bias, the evolving, often contested, metrics of translation quality (particularly in high-stakes legal and medical domains), and the restructuring of professional responsibility among developers, users, and human translators. Key findings demonstrate a prevailing low trust in Generative AI for core translation tasks, persistent deficiencies in accuracy, particularly in specialized fields like legal documentation, and urgent demands for algorithmic transparency and human agency in the face of data exploitation and bias. Ultimately, navigating the future of AI translation necessitates not just technological refinement, but the urgent establishment of comprehensive ethical and professional guidelines, promoting AI literacy, and ensuring that human oversight remains the decisive factor in high-stakes linguistic mediation.

Keywords: *AI Translation, Ethics, Quality Assessment, Responsibility, Algorithmic Bias, Data Governance, Translator Agency, Neural Machine Translation (NMT), Hybrid Translation.*

PROVOCAREA TRIPARTITĂ: ETICĂ, CALITATE ŞI RESPONSABILITATE ÎN TRADUCEREA ASISTATĂ DE INTELIGENŢA ARTIFICIALĂ

Integrarea rapidă a Inteligenței Artificiale (IA) și a Modelelor de Limbaj de Mari Dimensiuni (LLM) în industria serviciilor lingvistice a generat o provocare tripartită critică în ceea ce privește etica, calitatea și responsabilitatea în procesul traducerii. Lucrarea de față, fundamentată pe literatura academică recentă din perioada 2023–2025, analizează conflictele etice emergente legate de utilizarea datelor și prejudecățile algoritmice (*algorithmic bias****), metricele în continuă evoluție și adesea contestate ale calității traducerii (în special în domeniile juridic și medical, unde mizele sunt ridicate), precum și restructurarea responsabilității profesionale între dezvoltatori, utilizatori și traducători umani. Rezultatele principale evidențiază un nivel scăzut de încredere în IA generativă pentru sarcinile de traducere de bază, deficiențe persistente în privința acurateței — în special în domenii specializate, cum este documentația juridică — și cerințe urgente pentru transparență algoritmică și autonomie umană în fața exploatării datelor și a subiectivității. În cele din urmă, gestionarea viitorului traducerii prin IA necesită nu doar rafinare tehnologică, ci și stabilirea urgentă a unor ghiduri etice și profesionale cuprinzătoare, promovarea alfabetizării digitale în domeniul IA și asigurarea faptului că supravegherea umană rămâne factorul decisiv în medierea lingvistică de înaltă responsabilitate.***

Cuvinte-cheie: *Traducere prin IA, Etică, Evaluarea calității, Responsabilitate, Prejudecată algoritmică, Guvernanța datelor, Autonomia traducătorului, Traducere automată neuronală (NMT), Traducere hibridă.*

1. Introduction: The Restructured Landscape of Translation

The advent of Neural Machine Translation (NMT) and subsequent Generative AI models marks a definitive paradigm shift in the translation profession. AI tools have moved from being mere computational aids to powerful, fluent content generators, fundamentally changing professional workflows and market dynamics. This technological revolution, while promising unprecedented efficiency and accessibility, has simultaneously generated a new wave of critical academic inquiry focused on three interconnected pillars: ethics, quality, and responsibility. The challenge is no longer merely to achieve technical accuracy but to ensure the resulting linguistic output is fair, culturally sound, and professionally accountable.

Recent research spanning 2023 to 2025 highlights these tensions vividly. The literature reveals that AI's powerful capabilities are deeply intertwined with complex moral dilemmas. For instance, the very foundation of AI translation—the immense datasets used for training—is identified as a core ethical problem, creating conflicts over data ownership and informed consent [1]. Simultaneously, the assessment of quality has moved beyond simple fluency to rigorous evaluations of accuracy and cultural appropriateness, especially in critical contexts [2]. The question of responsibility has also been restructured: who is liable when a fluent, yet erroneous, AI translation leads to professional or personal harm? The consensus in this body of literature suggests that the future sustainability and integrity of the translation industry depend entirely on establishing robust frameworks that govern these three areas. The following sections systematically analyze the ethical, quality, and responsibility challenges in AI translation, arguing that while AI offers undeniable benefits, its adoption must be conditional on prioritizing human agency, developing rigorous and domain-specific quality metrics, and enforcing transparent accountability structures.

2. The Ethical Imperative: Bias, Transparency, and Data Ownership

Ethical scrutiny of AI translation systems primarily focuses on the upstream mechanisms of data acquisition, processing, and the downstream effect of algorithmic outputs on fairness and human rights. Two major ethical battlegrounds dominate the discussion: algorithmic bias and data legitimacy.

2.1 Algorithmic Bias and the "Black-Box" Problem

One of the most pressing ethical risks in AI-assisted translation is the reproduction and amplification of societal biases present in the massive training corpora [3]. Neural Machine Translation (NMT) systems, trained on large, unchecked datasets, tend to reflect cultural, gender-based, or racial prejudices. For example, if a corpus contains disproportionate instances of male pronouns associated with professional roles (e.g., 'doctor,' 'engineer'), the AI system may systematically output biased translations that reinforce gender stereotypes when translating from gender-neutral languages.

The resulting translations can perpetuate harmful stereotypes, render minorities invisible, or distort sociocultural contexts, leading to outcomes that are professionally unacceptable and ethically damaging. As noted in *Ethical Challenges in AI-Assisted Translation* [3], the absence of control mechanisms can lead to an overreliance on technology, exacerbating this risk when users lack training in critical evaluation. The paper argues that this lack of control risks the deprofessionalization of the craft if the translator loses the critical distance required to interrogate the AI's underlying assumptions [3].

Compounding the issue of bias is the problem of transparency. Generative AI systems often operate as "black boxes," making it exceedingly difficult to determine precisely how a particular translation was generated [3]. This lack of explicability—or the inability to reverse-engineer the decision-making path—prevents human translators from justifying their interventions or identifying the source of deep-seated bias. This is not merely a technical annoyance but a fundamental ethical challenge, as transparency is a core principle in modern AI governance frameworks. Without understanding the *why* behind a translation error, mitigating bias becomes a futile exercise, and accountability becomes impossible to assign. The call for transparent, auditable, and culturally sensitive tools is therefore a non-negotiable requirement for ethically acceptable translation practice [3].

2.2 Data Legitimacy, Copyright, and Professional Rights

The second major ethical front concerns the origins and use of the data that powers AI models. The study *Ethical Conflicts, Legal Risks, and Translator Agency in the Algorithmic Era* [1], anchored in the critical question "Whose data?", investigates the ethical conflicts generated by data utilization. The paper identifies key risks, including unauthorized data extraction and a severe lack of informed consent regarding the training data sources, leading to a "data alienation" where the content is stripped of its original context [1].

Translation data, often stripped of its contextual and cultural features during computational processing, is frequently used without the consent of the original translators or copyright holders [1]. This constitutes an infringement of data rights, posing serious ethical challenges to the profession and marginalizing the intellectual and moral rights of human professionals. Furthermore, the use of proprietary client data for training models raises severe confidentiality and legal risks, particularly in sensitive sectors like corporate or government translation. The debate makes it clear that the translation industry cannot simply accept technological efficiency at the expense of ignoring fundamental issues of intellectual property and professional rights.

The research argues that to address these dilemmas, a new framework for fair, accountable, and human-centered data governance in AI translation must be proposed. This framework is essential not only for legal compliance but for safeguarding the translator's professional agency and the integrity of professional ethics in the digital age [1].

2.3 Ethical Implications for Low-Resource Languages and Linguistic Diversity

The ethics of AI translation extend beyond the protection of professionals to the protection of language itself. NMT systems are inherently data-hungry; consequently, high-quality AI translation is primarily optimized for high-resource language pairs (e.g., English, Spanish, Chinese). This creates a structural bias against low-resource languages which lack the extensive digital corpora needed for effective training.

The ethical implication is twofold: first, speakers of low-resource languages receive translation outputs of demonstrably lower quality due to the scarcity of data, which compounds social and economic disadvantages. Second, prioritizing the dominant languages risks the homogenization of linguistic output globally, potentially eroding linguistic diversity and cultural distinctiveness. The responsibility, in this case, falls on developers and funding bodies to implement strategies (such as zero-shot learning or synthetic data generation) that promote fair and equitable access to high-quality translation technologies across all linguistic communities. The failure to do so transforms a technological shortcoming into an act of linguistic injustice [3]; [1].

3. The Evolving Definition and Assessment of Quality in the AI Era

The concept of "quality" in translation has always been complex, involving fluency, accuracy, domain-specificity, and cultural appropriateness. With AI, this definition has become even more volatile, splitting into debates over AI's intrinsic accuracy and its reliability for high-stakes, specialized content.

3.1 Trust, Reliability, and the Cognitive Load of Fluent Errors

The human perception of AI-generated output is a direct measure of its perceived quality [4], a mixed-methods study involving professional translators, reveals a critical insight: **trust in Generative AI for core translation tasks is generally low**. This lack of trust is a direct reflection of historical and persistent quality concerns.

One key challenge is the nature of NMT errors. Unlike older rule-based systems that often produced clunky or obviously incorrect translations, NMT outputs are characterized by high fluency. This fluency can be deceptive, masking subtle, contextually inappropriate, or factually wrong mistranslations—what is often called "fluent error" or "hallucination." For translators, the usefulness of an AI tool is inextricably linked to its accuracy and the reliability of its output, which is highly sensitive to the nature and timing of system failures [4]. The process of identifying these fluent errors (post-editing) can be more cognitively demanding and time-consuming than translating from scratch, effectively undermining the efficiency benefit of the AI tool. This high cognitive load associated with deep correction is a significant factor contributing to professional distrust [4].

3.2 Quality in High-Stakes Domains: Legal and Medical Translation

The limitations of AI quality become acutely visible in high-stakes specialized fields where linguistic precision and accountability are paramount.

In legal translation, where a single mistranslation can have severe contractual or diplomatic consequences, AI has been shown to be inadequate. *A Comparative Study of Accuracy in Human vs. AI Translation of Legal Documents into Arabic* [5], which compared AI (ChatGPT) and human translation of legal documents, revealed a significant gap in accuracy. Human translations were found to be superior in terms of legal terminology usage, clarity of complex legal constructs, and adherence to formal legal standards [5]. Specifically, the study noted that while AI could produce coherent text, it often failed to capture the culturally sensitive and legally defined precision required for effective legal communication in the Arabic context. This research confirms that while AI can handle general language, it often falls short of the precision required in contexts where terminology is legally defined and culturally sensitive. The study strongly sug-

gests that a hybrid approach is necessary, where AI supports but does not replace human expertise [5].

A similar imperative for caution is seen in medical writing. *Challenges, Ethical Considerations, and Best Practices of Using AI-Assisted Translation in Medical Writing* [2] emphasizes that while AI can save time, mistakes in AI-generated translations can cause confusion, damage credibility, and even affect patient safety. The research underscores that effective translation in this domain involves more than just word-matching; it requires an understanding of cultural context, professional norms, and writing style that advanced AI models still struggle with [2]. The authors advocate for best practices that mandate authors must clearly disclose the use of AI tools and carefully review all AI translations before submission, placing the burden of final quality squarely on the human professional [2].

3.3 The Promises and Pitfalls of Automated Quality Assessment (A-TQA)

Given the high volume of machine-translated content, a logical technological extension is the use of AI for Translation Quality Assessment (TQA) itself. *Using AI in Translation Quality Assessment* [5] explores this concept, investigating the feasibility and reliability of using AI, specifically ChatGPT, for TQA in legal translation.

The promise of A-TQA is that it could reduce instructor workload, promote evaluation consistency, and scale quality checks across large corpora. However, the study found that while some correlation exists between AI evaluations and human judgments, there are inherent challenges. AI-based TQA may struggle to assess the semantic complexity, stylistic appropriateness, and contextual nuance that human reviewers deem critical. If A-TQA systems are not calibrated to look for the subtle, fluent errors typical of NMT, they risk rubber-stamping low-quality, biased, or contextually inappropriate output. Therefore, while AI can assist in TQA by flagging mechanical issues, the ultimate responsibility for the definition and verification of quality remains a human, professional task [5].

4. Redefining Responsibility and Translator Agency

The integration of AI has radically challenged traditional notions of accountability, shifting the locus of responsibility from a single human translator to a complex network involving developers, platforms, and post-editors.

4.1 The Crisis of Accountability, Authorship, and the Ethical Buffer

A central challenge is determining who assumes responsibility when errors occur in machine-generated translations [3]. The lack of algorithmic transparency means that liability—whether legal, professional, or ethical—is diffused.

The papers stress that this ambiguity is highly problematic. The invisibility of the translator's effort is a related concern: At what point does the translator cease to be the author and become merely a reviewer or a "post-editor," a role that risks rendering their professional competencies invisible [3]? This concern extends to the legal field, where the concept of "authorship" has historically been tied to human intellectual creation.

The literature argues that the human professional must consciously assume the role of the "ethical buffer," responsible not only for translating but also for correcting and calling out the ethical failures of automated systems. This necessitates a proactive stance where the translator, armed with critical ethical and technical literacy, becomes the ultimate guarantor of fidelity, accuracy, and fairness. This shift implies a significant expan-

sion of the translator's professional role, demanding they acquire a new set of ethical, technical, and critical competencies beyond traditional linguistic mastery Translation [3].

4.2 The Role of AI Literacy and Professional Autonomy in Education

In the face of technological dependency, the preservation of translator agency is paramount. *Investigating novice translation students' AI literacy* [5] and *Ethical Challenges in AI-Assisted Translation* [3] both advocate for a full-scale revision of translation curricula.

The purpose of this revision is to transform the pedagogical approach from avoiding AI to fostering AI literacy—the ability to evaluate, collaborate with, and utilize AI responsibly. Studies show that novice students who lack the competence to identify covert AI errors risk developing inaccurate skills or over-relying on the tools [5]. By integrating AI ethics and critical evaluation training from the earliest stages, educational institutions can fulfill their responsibility to prepare professionals who can engage with automated systems without relinquishing their professional judgment. This ensures the future workforce is equipped with the cultural and moral competencies necessary to tackle the changing environment of AI-based translation [3] [5].

4.3 Socio-Economic Responsibility: Market Precarity and the Freelancer

The responsibility challenge is not purely ethical or technical; it has profound socio-economic dimensions, particularly impacting freelance translators. *Translators' Trust and Distrust in the Times of GenAI* [4] notes that ongoing concerns over market precarity and falling rates are major factors contributing to translator distrust of GenAI.

As AI efficiency drives down the perceived value of human input (even when that input involves necessary post-editing for quality assurance), the financial sustainability of the profession is threatened. The ethical responsibility, therefore, extends to pricing structures and client education. Developers and large language service providers (LSPs) have a socio-economic responsibility to ensure that their business models do not systematically exploit the human post-editors whose labor is essential for ensuring the final, high-quality, accountable translation product. Fair compensation for the cognitive load of error detection and correction must be recognized as integral to upholding professional ethics [4].

4.4 Policy and Regulatory Demands: The Road to Accountable AI

Given the magnitude of the ethical and quality risks, the need for clear policy and regulatory frameworks is evident. The debate on ethics and artificial intelligence in translation is inseparable from the need for institutional implementation and consensus on fundamental principles. This includes:

1. **Mandatory Disclosure:** Requiring authors and service providers to clearly disclose the extent of AI use in the translation process [2].
2. **Data Governance:** Establishing legal mechanisms to ensure that training data is authorized and consent is obtained, protecting the intellectual property of translators [1].
3. **Auditing and Certification:** Implementing regular, independent audits of AI models to identify and mitigate biases, particularly in high-stakes domains.

These regulatory steps represent the institutional commitment required to ensure that technological advancement aligns with professional ethical standards and safeguards the public interest [2], [1].

5. Conclusion: Synthesizing the Future of AI Translation

The integration of AI into the translation field presents a complex and urgent tripartite challenge concerning ethics, quality, and responsibility. The research conducted between 2023 and 2025 provides a clear picture of this restructured landscape, underscoring that the core issues are less about technology development and more about professional governance.

Ethically, the conflicts revolve around the opacity of black-box algorithms and the violation of professional data rights stemming from massive, unauthorized training corpora [1] [3]. The industry must prioritize transparency, anti-bias filtering, and human-centric data governance to move forward responsibly and fairly address the historical marginalization of low-resource languages.

In terms of Quality, the academic consensus is that AI-generated output, while fluent, frequently fails to meet the stringent standards required in high-stakes domains like legal and medical translation, often lacking the contextual, cultural, and precise technical competence of human professionals [5] [2]). The enduring necessity of human oversight for quality assurance is undeniable, and the challenges of automating TQA must be approached with caution [5].

Regarding Responsibility, the burden has been complicated, necessitating a shift in the translator's role to that of an "ethical buffer" and critical evaluator. This demands enhanced professional AI literacy to resist over-reliance and ensure the final output aligns with deontological principles, as current studies show a deep-seated distrust in GenAI for core tasks [4]. Furthermore, the economic responsibility to ensure fair compensation for the critical, high-cognitive-load work of post-editing must be addressed.

The future of AI translation is not defined by the elimination of the human element, but by the successful integration of AI's efficiency within a framework rigorously governed by ethical principles and quality standards. This requires continuous technological refinement, comprehensive ethical regulation, and, most importantly, the empowerment of human translators as the final, accountable decision-makers in an increasingly automated world.

References:

1. ZIWEI C. Ethical Conflicts*, Legal Risks, and Translator Agency in the Algorithmic Era*. 2025. [Accessed 15.11.2025] Available: https://francis-press.com/uploads/papers/7a8ENUwrRP2GwUL2wAfVquXL0CWg8a35owaYdTjh.pdf
2. KOJIMA T., MORITA T., POPIEL H. *Challenges, Ethical Considerations, and Best Practices of Using AI-Assisted Translation in Medical Writing*. 2024. [Accessed 21.11.2025] Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12602140/]
3. ANDRADE PRECIADO J. S., RAMÍREZ HÉCTOR J. S., SÁNCHEZ BELÉM P., PÉREZ ESTHER E. G. *Ethical Challenges in AI-Assisted Translation.* 2025. [Accessed 21.11.2025] Available: https://www.researchgate.net/publication/392891457_Ethical_Challenges_in_AI-Assisted_Translation
4. RIVAS GINEL M. I. *Translators' Trust and Distrust in the Times of GenAI.* 2025. [Accessed 17.11.2025] Available: https://www.tandfonline.com/doi/full/10.1080/14781700.2025.2507594
5. ALGHAMDI FATIMAH A., ALOTAIBI H. *Using AI in Translation Quality Assessment: A Case Study of ChatGPT and Legal Translation Texts*. 2025. [Accessed 21.11.2025] Available: https://www.mdpi.com/2079-9292/14/19/389