ANADOLU ÜNİVERSİTESİ

AÜSBD

Anadolu Üniversitesi Sosyal Bilimler Dergisi
Anadolu University Journal of Social Sciences

Makine Çevirisinde Şeffaflık ve Hesap Verebilirlik: Yapay Zeka Çağında Yeni Etik Normlar Ortaya Çıkıyor mu?

Transparency and Accountability in Machine Translation: Are New Ethical Norms Emerging in the Age of AI?

Gülfidan AYTAŞ[1]

Öz

Bu çalışma, Makine Çevirisi (MÇ) alanında şeffaflık ve hesap verebilirlik kavramlarının politika belgelerinde, kurumsal söylemlerde ve akademik yazında nasıl tanımlandığını, uygulandığını ve dönüştüğünü incelemektedir. Veri seti; altı uluslararası politika belgesi (örneğin UNESCO, OECD, AB Yapay Zekâ Yasası taslakları), yirmi üç hakemli akademik makale ve 2018–2024 yılları arasında önde gelen MÇ sağlayıcıları tarafından yayımlanan dokuz kurumsal açıklama ve teknik yönergeden oluşmaktadır. Belgeler, MÇ sistemlerinde şeffaflık, sorumluluk veya açıklanabilirlik kavramlarına açık atıf yapılması ölçüt alınarak amaçlı örnekleme yöntemiyle seçilmiştir. Analiz, Braun ve Clarke'ın tematik analiz çerçevesi izlenerek yürütülmüş ve NVivo yazılımı ile desteklenen yinelemeli, tümevarımsal bir kodlama süreci uygulanmıştır. Bulgular üç ana tema ortaya koymaktadır: (1) şeffaflığın politik söylemde söylemsel olarak genişlemesi, şeffaflığın evrensel bir güvence olarak sunulmasına karşın uygulama ölçütlerinin belirsiz kalması; (2) kurumsal iletişimde hesap verebilirliğin kaydırılması, değerlendirme emeği ve risk yönetiminin kullanıcıya yüklenmesi; ve (3) akademik yazında şeffaflığın daralması, kavramın kuramsal olarak tartışılmasına rağmen MÇ iş akışlarına ampirik düzeyde sınırlı entegrasyon sağlanması. Bu bulgular doğrultusunda çalışma, MÇ'de teknik, etik ve bağlamsal anlam katmanlarını bütünleştiren ilişkisel bir model olarak Hermeneutik Şeffaflık kavramını önermektedir. Çalışma, yüksek düzeyli söylem ile uygulamadaki gerçeklik arasındaki boşluklara dikkat çekmekte ve şeffaflığın durağan bir gereklilik değil; duruma bağlı, yorumlayıcı ve diyalojik bir süreç olarak yeniden düşünülmesi gerektiğini savunmaktadır. MÇ araştırmaları, yönetişim mekanizmaları ve platform tasarımı için sonuçlar da tartışılmaktadır.

Anahtar Kelimeler: Makine çevirisi, şeffaflık, hesap verebilirlik, yapay zekâ etiği, çeviri etiği

JEL Kodu: L86

Abstract

This study examines how transparency and accountability are articulated, operationalized, and transformed across policy documents, corporate communication, and academic discourse on Machine Translation (MT). The dataset consists of six international policy documents (e.g., UNESCO, OECD, EU AI Act drafts), twenty-three peer-reviewed academic articles, and nine corporate statements and technical guidelines issued by leading MT providers between 2018 and 2024. Documents were selected through purposive sampling based on explicit references to transparency, responsibility, or explainability in MT systems. The analysis follows Braun and Clarke's thematic framework and was conducted through an iterative, inductive coding process supported by NVivo. Three overarching themes emerged: (1) the rhetorical expansion of transparency in policy discourse, where transparency is framed as a universal safeguard yet operational criteria remain vague; (2) the displacement of accountability in corporate communication, which locates evaluative labour and risk management at the user's end; and (3) the narrowing of transparency in academic texts, where the concept is discussed conceptually but seldom integrated into empirical MT workflows. Building on these findings, the article proposes Hermeneutic Transparency as a relational model that integrates technical, ethical, and contextual layers of meaning-making in MT systems. The study highlights gaps between high-level discourse and actual implementation practices and argues that transparency should be reconceived not as a static requirement but as a situated, interpretive and dialogic process. Implications for MT research, governance, and platform design are also discussed.

Keywords: Machine translation, transparency, accountability, AI ethics, translation ethics

JEL Code: L86

[1]Giresun University, School of Foreign Languages, Department of Foreign Languages, Lecturer Dr., gulfidan.aytas@giresun.edu.tr, ORCID: 0000-0003-1566-1592

Introduction

The rapid mainstreaming of neural machine translation (NMT) and, more recently, large language model (LLM)-based translation has transformed translation workflows, public-facing communication, and specialized domains such as law, health, and administration. Despite this technological shift, traditional translation ethics—centered on the human translator's duties of fidelity, responsibility to clients and readers, and professional accountability (Chesterman, 2001, 2016)—has struggled to keep pace. Translation production now involves multiple actors, including model developers, data providers, deployers (platforms or institutions), post-editors, and end-users, creating a dispersed agency that complicates two canonical ethical pillars: transparency and accountability.

Policy regimes have responded to these challenges by elevating these pillars. The OECD AI Principles (2019) emphasize "transparency and responsible disclosure around AI systems" to allow affected parties to understand and contest outcomes (p. 1), and call for accountability among actors "developing, deploying or operating AI systems" (p. 1). UNESCO's Recommendation on the Ethics of AI (2021) extends this logic, highlighting that transparency and explainability are often essential preconditions for rights protection, and must be proportionate to context and impact (pp. 9–10). The EU Artificial Intelligence Act (European Union, 2024) codifies these principles in law, clarifying that fulfilling transparency obligations does not automatically render a system lawful and distinguishing risk tiers, explicitly listing AI systems used for translation of initial documents as low-risk while still requiring documentation, logging, and proportional transparency (p. 49, 69–70, 121).

Empirical research, however, reveals persistent gaps between these policy ideals and MT practice. For example, gender-ambiguous inputs are disproportionately rendered as masculine by Google Translate (~90%) and DeepL (~85–88%), diverging from benchmark distributions (Rescigno & Monti, 2023, pp. 5–6). Users are rarely informed about these biases or how to contest outputs, highlighting an ethical fault line recognized by OECD (2019) and UNESCO (2021). In legal translation, liability for machine-assisted outputs remains unclear, particularly when post-editors or human certifiers are involved, as existing frameworks may not account for machine-generated omissions or distortions (GenLaw Blog, 2023, p. 1).

These observations motivate the present study, which aims to investigate how transparency and accountability are defined and operationalized in MT, identify tensions or failures in practice, and propose actionable norms for distributing responsibility across developers, deployers, and human translators/post-editors. Specifically, the study addresses three research questions: (1) How are transparency and accountability articulated in MT policy and practice? (2) Where do these principles conflict with real MT behaviors such as bias, hallucination, or domain drift? (3) What role-sensitive ethical norms can be implemented to ensure responsible practice across the MT lifecycle?

By integrating policy analysis, academic scholarship, and corporate practice, this study contributes to the field by: (a) clarifying the regulatory baseline for transparency and accountability in AI/MT; (b) illustrating how opacity manifests in MT through analytic vignettes; and (c) proposing a role-based ethical framework that operationalizes high-level principles into actionable guidance aligned with UNESCO's call for auditability and traceability (2023, p. 10). Ultimately, this research situates MT at a normative inflection point, arguing that translation ethics must evolve from individual, compliance-based frameworks to relational, system-aware, and interpretive practices capable of addressing distributed agency and socio-technical risks.

Literature Review and Theoretical Framework

Evolution of Translation Ethics: From Human Agency to Technological Mediation

The concept of ethics in translation studies has traditionally centered on the human translator as the primary ethical agent, responsible for fidelity, faithfulness, and accountability to the source text, the target audience, and the commissioner. Anthony Pym (2012) identifies translator responsibility as "*a social negotiation of trust in communicative acts*" (p. 46), positioning ethical decision-making within interpersonal relations. Andrew Chesterman (2001) further operationalizes this with his well-known "Hieronymic Oath," in which translators pledge to act with honesty, impartiality, and competence toward all parties involved in the translation act (pp. 147–148). These frameworks, deeply rooted in human agency, assume a clear locus of ethical responsibility: the translator.

However, as translation increasingly becomes mediated by technology — from computer-assisted translation (CAT) tools to neural machine translation (NMT) and large language models (LLMs) — the singular focus on the human translator is no longer sufficient. Baker (2018) observes that "*translation is no longer a discrete act performed by a single subject but a distributed process across human and non-human actors*" (p. 112). This distributed agency blurs traditional boundaries of ethical responsibility. Translators today often serve as post-editors or quality controllers rather than sole producers of the target text, raising questions about how responsibility is shared among developers, data curators, platform providers, and end-users.

This shift has prompted scholars to revisit classical translation ethics. Chesterman (2016) argues that "*ethical norms must evolve to encompass institutional and technological actors*" (p. 59). The question is no longer merely how a translator behaves ethically but how the entire translation ecosystem — including algorithms — adheres to ethical standards. Such a reconceptualization requires importing concepts from adjacent fields, most notably AI ethics.

Core Principles of AI Ethics: Transparency and Accountability

The field of AI ethics, emerging prominently in the last decade, provides a robust vocabulary for addressing the challenges posed by machine translation. Luciano Floridi (2019) proposes five fundamental principles for the ethical development and deployment of AI: beneficence, non-maleficence, autonomy, justice, and explicability (pp. 39–40). Among these, explicability — often operationalized as *transparency* and *accountability* — is especially pertinent to machine translation. Floridi and Cowls (2021) define explicability as "*the ability to render AI systems understandable and to hold relevant actors answerable for their outputs*" (p. 12).

Transparency has two dimensions: *technical transparency,* which refers to the comprehensibility of algorithmic decision-making processes, and *procedural transparency*, which involves disclosure to users about system capabilities, limitations, and use contexts. Boddington (2017) stresses that "*without transparency, informed consent and public trust in AI are impossible*" (p. 85). In the context of MT, this translates into the need to disclose whether a text was machine-translated, the system's training data limitations, and known biases.

Accountability, meanwhile, concerns the assignment of responsibility when AI outputs cause harm or produce undesirable effects. Jobin, Ienca, and Vayena's (2019) global survey of AI guidelines shows that accountability appears in over 80% of them, typically defined as the "*obligation of actors involved in the AI lifecycle to justify and assume responsibility for decisions and outcomes*" (p. 396). Yet as Mittelstadt (2022) notes, accountability in AI is often "*diffuse and diluted across complex socio-technical systems*" (p. 5), a challenge that becomes particularly acute in machine translation, where errors can result from multiple layers of interaction: biased data, flawed model design, inadequate post-editing, or misuse by end-users.

Intersecting Ethical Frameworks: Translation Meets AI

The convergence of translation ethics and AI ethics reveals both complementarities and tensions. Both traditions value responsibility, trust, and fairness, but they conceptualize them differently. Translation ethics historically emphasizes *relational ethics* — trust between translator, author, and audience — whereas AI ethics often operates in a *systemic register*, focusing on governance, regulation, and institutional accountability.

Scholars have begun articulating new frameworks to bridge these paradigms. Kenny and Doherty (2020) argue that "*machine translation ethics must synthesize the translator's normative obligations with the systemic accountability of AI systems*" (p. 27). They propose a layered model in which developers, providers, and translators share responsibility at different points of the translation lifecycle. Similarly, O'Hagan (2022) suggests that transparency in MT should not be understood solely as disclosure of machine involvement but also as "*making visible the social and cultural biases embedded in training data and system outputs*" (p. 61).

This expanded view of transparency and accountability is echoed in international policy documents. UNESCO's *Recommendation on the Ethics of Artificial Intelligence* (2021) emphasizes that transparency and explainability are "*essential preconditions for accountability and for the exercise of human rights*" (pp. 9–10). The OECD (2020) calls for "*responsible disclosure*" and stresses that individuals and organizations "*should be held accountable for*

the proper functioning of AI systems" (p. 1). The EU Artificial Intelligence Act (2024) codifies these principles in binding law, requiring documentation, user disclosure, and traceability mechanisms — even for lower-risk applications such as machine translation (pp. 49, 69–70).

Despite these advances, significant gaps remain. As Hovy and Spruit (2023) point out, transparency is often treated as a binary property — a system is either transparent or opaque — whereas in practice it exists on a spectrum, requiring contextual calibration (p. 8). Furthermore, accountability mechanisms often fail to map onto the actual distribution of agency in MT workflows. Translators may be held responsible for errors originating in biased datasets, while developers remain shielded from liability. This misalignment underscores the urgent need for a new, role-sensitive ethical model.

Contextualizing Transparency: From Technical Disclosure to Hermeneutic Visibility

While transparency in AI is frequently discussed in technical or procedural terms, its ethical significance within translation lies in its interpretive dimension — how meaning, mediation, and power are rendered visible. Drawing from Ricoeur's (1976) hermeneutic model, translation can be understood as a process of *making meaning appear* through interpretive labor. When this process is delegated to non-human agents, transparency extends beyond code-level explainability; it involves clarifying how algorithmic mediation shapes interpretive outcomes.

Olohan (2021) argues that "the opacity of machine translation is not merely a technical limitation but an epistemic condition of mediated communication" (p. 88). That is, even if developers disclose architectures or datasets, users still face interpretive opacity — they cannot fully trace how linguistic, cultural, or ethical norms are encoded. This suggests a need for *hermeneutic transparency*: an awareness of how technological mediation reframes linguistic agency and cultural values.

In this sense, transparency in translation ethics cannot be equated with full disclosure. Rather, as Ananny and Crawford (2018) note, transparency must coexist with "meaningful opacity," recognizing that interpretability is situated, partial, and culturally mediated (p. 977). For machine translation, this means communicating *the conditions of mediation* — how systems interpret and transform meaning — rather than merely providing technical documentation.

Accountability Beyond the Human: Distributed Responsibility and Shared Agency

The rise of neural and generative translation systems challenges traditional notions of accountability premised on singular human intention. Latour's (2005) actor-network theory offers a useful framework for understanding responsibility as a distributed property of socio-technical assemblages. Within this perspective, translators, developers, data annotators, corporate platforms, and even users collectively participate in shaping translation outcomes.

As Krüger (2022) observes, "ethical agency in AI-assisted translation is an emergent property of interlinked human and non-human actors" (p. 312). This challenges the juridical tendency to localize blame or responsibility in a single point of failure. Instead, accountability must be conceptualized as *shared accountability*, emphasizing coordination, traceability, and responsiveness throughout the translation lifecycle.

Recent proposals advocate multi-level accountability structures:

1. **Operational accountability** — ensuring that system design and data curation meet ethical standards;
2. **Translational accountability** — maintaining editorial oversight during post-editing and deployment;
3. **Institutional accountability** — ensuring organizational transparency in data governance and model training.

Together, these levels instantiate what Crawford and Paglen (2021) term the "ecology of accountability," wherein responsibility becomes a relational network rather than a linear chain. This view aligns with the European Union's (2024) emphasis on "traceable accountability" in AI governance — a requirement that responsibility be *distributable yet identifiable* at every node of system interaction.

Methodology and Analytical Framework

This study employs a qualitative, document-based, and conceptual research design that integrates ethical theory, policy analysis, and critical discourse examination in order to investigate how transparency and accountability are articulated and operationalized within machine translation ecosystems. Rather than focusing on the technical performance of MT systems, the study adopts a descriptive and interpretive orientation, consistent with methodological recommendations in translation studies and AI ethics that emphasize the importance of analyzing socio-technical systems, institutional discourses, and distributed forms of agency (Cronin, 2022; Floridi, 2019; Kenny & Doherty, 2020). Such an approach is particularly suited to capturing the multi-actor configuration of contemporary MT infrastructures in which developers, platform providers, post-editors, and end-users share— though asymmetrically—responsibility for meaning production and error management (Baker, 2018; Krüger, 2022). Accordingly, the study examines how principles of transparency and accountability are defined in policy discourse, reproduced or diluted in corporate communication, and critiqued within scholarly debates, while also identifying points at which these ideals come into conflict with observed MT behaviours such as bias, hallucination, and contextual drift.

The dataset was constructed through purposive sampling and consists of three interrelated bodies of material. Policy documents were primarily published between 2019 and 2024, reflecting the period in which international AI governance frameworks began to emerge.(A detailed overview of the dataset is provided in Appendix B (Table B1)): six major AI policy and governance documents, including the OECD AI Principles (2019, 2020, 2023), UNESCO's Recommendations on the Ethics of Artificial Intelligence (2021, 2023), and the EU Artificial Intelligence Act (2024); twenty-three peer-reviewed studies were collected and compiled between September 2024 and January 2025 that specifically address ethics, responsibility, or transparency in AI-mediated translation; and nine publicly available corporate documents produced by Google Translate, DeepL, and ChatGPT-based translation materials, such as transparency statements, privacy policies, and model documentation. Sources were selected based on accessibility, recency, and direct relevance to the normative concerns at the centre of this research. The material was gathered between September 2024 and January 2025 through searches conducted in Scopus, Web of Science, Google Scholar, and the official repositories of OECD, UNESCO, and the European Union. A three-stage screening process—title review, abstract evaluation, and full-text reading—was used to ensure conceptual coherence and thematic alignment, resulting in the exclusion of documents that focused solely on technical performance, general AI ethics unrelated to language technologies, or historical analyses outside the 2018–2024 timeframe.

Data analysis followed Braun and Clarke's (2006) six-phase thematic analysis model and was conducted using NVivo to support systematic coding and theme development. To enhance the transparency of the study, a representative NVivo codebook containing three primary coding nodes, along with illustrative coded data excerpts, is provided in Appendix A. An initial round of open coding generated a corpus of inductive and theoretically informed codes related to the framing of transparency, interpretations of accountability, and manifestations of institutional opacity. These codes were refined into a set of overarching themes that captured the discursive patterns across policy, academic, and corporate layers, with particular attention to the "rhetoric–reality gap" identified in prior work (Fairclough, 1995; Merriam & Tisdell, 2016). Analytical triangulation was central to the study's methodology: policy texts were examined for normative prescriptions, scholarly literature provided conceptual and critical depth, and corporate materials offered empirical insight into the operationalization—or strategic reframing—of transparency and accountability in practice. Corporate transparency documents from major machine translation platforms were examined, primarily published between 2020 and 2024, reflecting the increasing institutionalization of AI transparency policies. The triangulated structure made it possible to identify tensions between formal commitments to ethical governance and the practical limitations of proprietary MT systems that often resist full disclosure or traceability (Hovy & Spruit, 2023; O'Hagan, 2022).

To ensure trustworthiness, the analysis incorporated multiple strategies aligned with Lincoln and Guba's (1985) criteria for credibility and confirmability. Coding was repeated at two different points in time to enhance consistency; an analytic trail documenting coding decisions, theme consolidation, and interpretive choices was maintained; and reflexive engagement was sustained throughout the process, acknowledging the researcher's interpretive stance and the epistemic constraints inherent in examining opaque, commercially protected AI

systems. Following the calls of O'Hagan (2022) and Mittelstadt (2022) for a critical and reflexive orientation in AI ethics research, the study positions itself not as an empirical evaluation of MT performance but as an interpretive inquiry into the normative structures, discursive practices, and socio-technical dynamics that shape ethical responsibility in machine-mediated translation.

Findings and Discussion

The findings presented in this section draw on a triangulated analysis of policy documents, scholarly literature, and corporate materials, each offering a distinct but interconnected perspective on how transparency and accountability are articulated, operationalized, and reframed within the machine translation ecosystem. Rather than treating these layers as discrete domains, the analysis traces recurring discursive patterns—particularly the widening gap between ethical rhetoric and practical implementation—across all three. The themes that follow synthesize these patterns by examining the normative framing of ethical principles in policy texts, the critical interrogations developed in academic discourse, and the operational choices embedded in the design and communication strategies of major MT providers. Together, these thematic clusters illuminate the systemic tensions that shape the ethical landscape of contemporary MT, as well as the relational dependencies that complicate attempts to assign responsibility within complex socio-technical environments.

The Rhetoric–Reality Gap in AI Policy Discourse

Analysis of the six policy documents indicates a persistent gap between the normative prominence of transparency and accountability and the mechanisms offered to enact them in machine translation systems. Both UNESCO (2021, 2023) and the OECD AI Principles (2020, 2023) repeatedly frame these values as foundational ethical imperatives; yet these affirmations remain largely declarative rather than operational. For instance, the UNESCO Recommendation states that "transparency and explainability are prerequisites for safeguarding human rights" (2021, p. 10) and later reiterates that "member states should ensure traceability in AI decision-making processes," but provides no accompanying guidance on how such traceability could be achieved in neural MT systems whose training corpora, model architectures, and domain adaptation procedures are proprietary and inaccessible. Similarly, the OECD guidance calls for "responsible disclosure to stakeholders," yet its implementation annex specifies detailed procedures primarily for high-risk sectors such as healthcare and critical infrastructure, while machine translation is categorised among "limited risk" systems for which disclosure remains voluntary and undefined. The EU AI Act (2024) mirrors this pattern: although the Act mandates documentation and post-market monitoring for high-risk AI, MT systems fall under minimal-obligation categories, resulting in only generic requirements such as "providing users with appropriate information," without specifying what constitutes appropriateness for a system that transforms meaning across languages.

This illustrative pattern directly aligns with Mittelstadt's (2022) critique of ethics-washing, wherein institutions foreground strong moral language but refrain from specifying the operational responsibilities needed to enact those values. In the context of MT, this produces a structural paradox: translation systems are normatively branded as "responsible AI," yet the documents governing them disclose almost nothing about how meaning is represented, how bias is mitigated, or how errors and context drift are identified and corrected. The absence of process-level guidance results in de facto self-regulation, leaving downstream actors; translators, deployers, and end-users to assume interpretive and corrective responsibilities without access to the technical conditions that shape system behaviour.

A systematic comparison of these cross-layer patterns is presented in Table 1.

Table 1

Comparative Analysis of Transparency and Accountability Across Policy, Academic, and Corporate Layers

Theme / Layer	Policy Documents (UNESCO, OECD, EU AI Act)	Academic Literature (2018–2024)	Corporate Materials (Google Translate, DeepL, ChatGPT Translate)
Framing of Transparency	Emphasizes transparency and explainability as ethical imperatives, yet provides limited operational guidance for proprietary MT systems.	Highlights rhetorical uses of transparency and calls for contextually grounded, relational conceptualizations.	Presents transparency through high-level assurances and UX cues rather than substantive disclosure of training data or model logic.
Accountability Distribution	Advocates traceability and documentation, but implementation protocols focus on high-risk sectors; MT remains under-defined.	Critiques fragmented accountability and argues for shared responsibility frameworks across actors.	Shifts responsibility to users through statements urging them to verify outputs; developer liability remains ambiguous.
Practical Prescriptions	Suggests documentation, logging, and risk-based transparency, though MT receives minimal explicit direction.	Proposes model cards, dataset disclosure, and translator-oriented ethical guidelines.	Offers interface-level transparency tools (editable outputs, feedback channels) with no accompanying process-level disclosures.
Evidence of Action	Articulates norms and principles but lacks enforcement mechanisms for low-risk systems such as MT.	Produces theoretical and empirical frameworks but notes absence of institutional implementation.	Demonstrates visible UI features but provides negligible detail on audits, dataset composition, or error-handling pipelines.

Corporate Transparency: Branding Rather Than Disclosure

The corporate documents examined system cards, privacy statements, and transparency reports from Google Translate, DeepL, and ChatGPT Translate reveal a consistent tendency to equate transparency with high-level assurances rather than epistemic disclosure. All three platforms affirm commitments to responsible AI, yet none provide concrete details regarding data provenance, audit procedures, domain-specific vulnerabilities, or error-handling protocols. This tendency becomes particularly clear in statements such as "we ensure our models are trained responsibly" and "users are responsible for verifying outputs," which recur across policy and corporate documentation and illustrate how responsibility is rhetorically shifted toward the end user rather than grounded in institutional disclosure. (See Appendix C, C2 for representative excerpts.). Similarly, DeepL's repeated emphasis on providing "safe and secure translation" operates as a generalized ethical promise rather than an explanation of how translations are generated, curated, or quality-controlled. The vagueness of these formulations aligns directly with Bietti's (2020) argument that ethical language in AI companies often functions as a competitive branding strategy rather than a mechanism of public accountability.

Corporate discourse further exhibits a shift from explainability to user-oriented empowerment narratives: features such as editable outputs, reporting channels, and confidence indicators are framed as transparency mechanisms, even though they reveal nothing about model decision processes or training dynamics. These interface-level gestures exemplify what Ananny and Crawford (2018) term "performative transparency," offering visibility at the level of user interaction while preserving the opacity of core algorithmic operations. In practice, such framing discursively relocates responsibility from system producers to end-users, who are implicitly tasked with identifying and correcting errors without access to the conditions under which those errors arise.

Distributed Accountability in MT Ecosystems

The triangulation of policy, scholarly literature, and corporate materials demonstrates that accountability in MT ecosystems is increasingly distributed across actors whose access to information and decision-making power is asymmetrical. Empirical findings in the reviewed scholarship (Kenny & Doherty, 2020; Krüger, 2022) show that

translators are positioned primarily as post-editors who intervene after automated processing yet lack visibility into model parameters, training data composition, or system limitations. This reduced epistemic agency constrains their capacity for informed ethical judgement, even as they are held responsible for the final quality of the output. Corporate documentation reinforces this model through user-facing guidance that encourages error reporting or corrective feedback, placing evaluative labour at the periphery of the translation pipeline. For example, platforms explicitly state: 'Users are responsible for verifying outputs' (Appendix C, C1). This multi-layered diffusion of responsibility echoes Han's (2021) theorization of the post-truth condition in digital systems, in which partial visibility coexists with systemic opacity. In MT, such dynamics normalize a fragmented accountability structure where errors are jointly produced but not jointly traceable.

Hermeneutic Transparency and the Reconfiguration of Interpretation

Despite these structural limitations, the dataset also reveals emergent practices that point toward a more interpretive, hermeneutic model of transparency. Several recent platform updates introduce contextual cues such as "automatically translated," side-by-side source–target views, and indications of model-generated reformulations. Although these measures fall short of algorithmic disclosure, they create interpretive awareness by alerting users to the mediated and contingent character of MT output. Such practices resonate with Cronin's (2022) proposal for a posthuman translation ethics grounded in humility toward technological mediation and with Ricoeur's (1976) view of interpretation as an act of responsibility. By signalling the conditions of production rather than claiming neutral linguistic fidelity, these gestures foster a form of transparency that enhances user literacy rather than attempting (or promising) full procedural openness. This suggests a conceptual shift: transparency becomes less about exposing internal logic and more about enabling informed engagement with the interpretive nature of machine-generated meaning.

Synthesis and Implications

The triangulated findings across policy, academic, and corporate layers reveal that the ethical governance of machine translation is shaped less by the absence of standards than by their uneven articulation and operationalization. The policy analysis demonstrated that although transparency and accountability are repeatedly positioned as normative imperatives, the documents offer limited procedural guidance for low-risk yet high-impact systems such as MT. As shown in the excerpts from UNESCO, OECD, and the EU AI Act, the emphasis on transparency remains predominantly declarative, leaving the mechanisms required to achieve it unspecified. The academic corpus further highlighted this tension by identifying the "rhetoric–reality gap," wherein ethical principles circulate as symbolic commitments without corresponding institutional structures capable of enforcing them. Corporate materials reinforced this pattern: statements such as "we ensure our models are trained responsibly" or "users are responsible for verifying outputs" illustrate how responsibility is rhetorically shifted toward end-users while core model operations remain opaque.

Taken together, these findings indicate that ethical incoherence emerges not from conceptual ambiguity but from structural asymmetries in who possesses epistemic access to MT systems and who bears the burden of mitigating their risks. In light of this evidence, the study argues for a relational reframing of MT ethics that directly addresses these asymmetries. First, transparency should be reconceptualized as interpretive disclosure, meaning that platforms must provide contextual information—such as dataset provenance summaries, domain limitations, and known error patterns—rather than relying solely on interface-level cues that simulate openness. Second, accountability should be treated as a negotiated relational practice, requiring explicit allocation of roles and responsibilities among developers, deployers, and translators; this entails moving beyond user-directed disclaimers toward traceable documentation and shared decision-making structures. Third, ethics should be approached as an ongoing socio-technical process, rather than a static compliance exercise, recognizing that meaning-making, error detection, and risk assessment in MT systems are co-produced by human and non-human agents.

This relational, posthumanist orientation provides a more actionable and context-sensitive model of ethical MT governance. It bridges the gap identified in this study between normative aspirations and practical implementation by emphasizing that ethical commitments must be accompanied by structures that distribute knowledge, responsibility, and interpretive capacity more equitably across the MT ecosystem.

Conclusion and Implications

This study examined how transparency and accountability are constructed, operationalized, and strategically reframed across policy, academic, and corporate discourses surrounding machine translation. By triangulating six major policy documents, twenty-three scholarly publications, and nine corporate materials, the analysis demonstrated that ethical governance in MT is shaped not by the absence of regulatory principles but by the inconsistent translation of those principles into actionable practices. Transparency is normatively codified across all three layers, yet remains largely symbolic in operational terms, while accountability is invoked frequently but distributed in ways that burden those with the least epistemic access to system behavior.

At the policy level, the findings indicate that global frameworks such as UNESCO, OECD, and the EU AI Act articulate transparency and accountability as essential ethical imperatives but fail to provide procedural guidance adequate for proprietary MT architectures. The absence of explicit standards for dataset provenance, auditability, or error traceability allows such systems to be normatively governed yet practically opaque. The academic literature, in turn, amplifies this structural gap by critiquing the widening distance between ethical rhetoric and institutional practice, identifying how ethical language can serve as an instrument of legitimacy rather than a vehicle for enforceable obligation. Corporate documents reinforce this pattern through high-level assurances that foreground responsible AI while withholding substantive detail about model operations, risk mitigation, or internal quality-control processes. Together, these findings demonstrate that ethical incoherence in MT arises from relational asymmetries—specifically, disparities in who holds knowledge of system behavior and who is assigned responsibility for its consequences.

The study contributes to translation and AI ethics by proposing a relational reframing of transparency and accountability suited to the socio-technical nature of contemporary MT. First, the findings support conceptualizing transparency as interpretive disclosure rather than procedural revelation. This approach recognizes that complete algorithmic visibility is neither feasible nor sufficient; instead, users, translators, and deployers require contextual information—dataset summaries, known risk profiles, domain-specific limitations—that enables informed interpretive judgment. Second, the study advances an understanding of accountability as a negotiated practice distributed across developers, deployers, and human translators, rather than a unilateral obligation shifted onto end-users through disclaimers or interface features. Third, it emphasizes that ethics in MT cannot be reduced to compliance-based frameworks, but must instead be treated as an ongoing socio-technical process shaped by dynamic interactions between human and non-human agents.

These conclusions have several implications for research, policy, and professional practice. For policymakers, the findings highlight the need to extend transparency requirements beyond high-risk AI systems to include widely deployed but understudied domains such as MT. Minimal disclosure frameworks—covering dataset provenance, error typologies, and domain constraints—would support more equitable epistemic access across stakeholders. For corporate actors, the study underscores the ethical limitations of branding-based transparency and calls for the development of accessible model documentation that communicates epistemically relevant information without compromising proprietary rights. For translators and professional communities, the relational reframing proposed here suggests the need for updated training protocols that foreground the interpretive challenges and ethical risks of working with MT outputs under conditions of partial opacity. Finally, for researchers, the study demonstrates the value of cross-layer discourse analysis as a methodological approach for uncovering systemic ethical tensions that remain invisible when examining technical performance alone.

In sum, the study shows that achieving meaningful transparency and accountability in machine translation requires an ethical model that recognizes the distributed, interpretive, and relational nature of contemporary translation ecosystems. Closing the gap between ethical rhetoric and operational reality will depend not only on refining regulatory standards but also on cultivating shared epistemic capacities and collaborative responsibility across developers, institutions, translators, and end-users. By reframing transparency and accountability as relational practices, the study offers a pathway toward more equitable and context-sensitive ethical governance of MT systems.

References

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Baker, M. (2018). *Translation and conflict: A narrative account.* Routledge.

Bietti, E. (2020). From ethics washing to ethics bashing: Regulating AI in Europe. *Computer Law & Security Review, 35*(6), 105–114. https://doi.org/10.1016/j.clsr.2020.105383

Boddington, P. (2017). *Towards a code of ethics for artificial intelligence.* Springer.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27–40. https://doi.org/10.3316/QRJ0902027

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chesterman, A. (2001). Proposal for a hieronymic oath. *The Translator, 7*(2), 139–154. https://doi.org/10.1080/13556509.2001.10799099

Chesterman, A. (2016). *Memes of translation: The spread of ideas in translation theory* (Revised ed.). John Benjamins.

Crawford, K., & Paglen, T. (2021). *Excavating AI: The politics of images in machine learning training sets.* AI Now Institute. https://www.excavating.ai

Cronin, M. (2017). *Eco-translation: Translation and ecology in the age of the Anthropocene.* Routledge.

Cronin, M. (2022). *Translation in the digital age: Ethics, agency, and posthumanism.* Routledge.

European Union. (2024). *Artificial Intelligence Act (Regulation (EU) 2024/1689).* Official Journal of the European Union. https://eur-lex.europa.eu

Fairclough, N. (1995). *Critical discourse analysis: The critical study of language.* Longman.

Floridi, L. (2019). *The logic of information: A theory of philosophy as conceptual design.* Oxford University Press.

Floridi, L., & Cowls, J. (2021). A unified framework of five principles for AI in society. *Harvard Data Science Review, 3*(1). https://doi.org/10.1162/99608f92.8cd550d1

GenLaw Blog. (2023). *Liability and post-editing in AI-assisted legal translation.* https://blog.genlaw.org/CameraReady/16.pdf

Han, B.-C. (2021). *Infocracy: Digitization and the crisis of democracy.* Polity Press.

Hovy, D., & Spruit, S. L. (2023). The social impact of natural language processing. *Computational Linguistics, 49*(1), 1–27. https://doi.org/10.1162/coli_a_00468

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kenny, D., & Doherty, S. (2020). Ethics in machine translation: A humanist perspective. *Translation Studies, 13*(1), 19–35. https://doi.org/10.1080/14781700.2019.1689878

Krüger, R. (2022). Translation ethics and the posthuman turn: Agency, responsibility, and technology. *Target, 34*(2), 301–324. https://doi.org/10.1075/target.21035.kru

Latour, B. (2005). *Reassembling the social: An introduction to actor-network theory.* Oxford University Press.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Sage.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.

Mittelstadt, B. D. (2022). The value of AI ethics. *AI and Ethics, 2*(1), 1–9. https://doi.org/10.1007/s43681-021-00084-x

O'Hagan, M. (2022). Transparency in the age of neural MT: Revisiting translation ethics. *Machine Translation, 36*(1), 57–70. https://doi.org/10.1007/s10590-021-09317-0

OECD. (2019). *OECD Principles on artificial intelligence.* https://www.oecd.org/going-digital/ai/principles/

OECD. (2020). *Implementing trustworthy AI: AI principles in practice.* https://doi.org/10.1787/cb6b6e5f-en

OECD. (2023). *AI, data and public policy: Building more trustworthy artificial intelligence systems.* https://doi.org/10.1787/1f0a2d20-en

Olohan, M. (2021). *Technology, translation, and society: A sociological perspective.* Routledge.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research, 42*(5), 533–544. https://doi.org/10.1007/s10488-013-0528-y

Pym, A. (2012). *On translator ethics: Principles for mediation between cultures.* John Benjamins.

Rescigno, A., & Monti, J. (2023). Gender bias in machine translation: a statistical evaluation of google translate and deepl for english, italian and german. In INTERNATIONAL CONFERENCE Human-informed Translation and Interpreting Technology (HiT-IT 2023) Proceedings (pp. 1-11). Incoma Ltd..

Ricoeur, P. (1976). *Interpretation theory: Discourse and the surplus of meaning.* Texas Christian University Press.

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence.* https://unesdoc.unesco.org/ark:/48223/pf0000380455

UNESCO. (2023). *Guidance for the ethical use and governance of artificial intelligence in education.* https://unesdoc.unesco.org/ark:/48223/pf0000385457

Venuti, L. (1995). *The translator's invisibility: A history of translation.* Routledge.

Wolfe, C. (2010). *What is posthumanism?* University of Minnesota Press.

Appendix A. NVivo Codebook Excerpts and Illustrative Coded Data

The following excerpts provide a representative sample of the NVivo codebook and coded data used in the thematic analysis. These examples aim to enhance methodological transparency by clarifying how analytical categories were constructed and how specific textual segments were coded.

Code 1: Transparency_As_Rhetoric

Definition:

Instances where policy or corporate documents invoke transparency or explainability as ethical principles without specifying operational mechanisms such as data provenance, audit procedures, or traceability pathways.

Inclusion Criteria:

Statements referencing transparency, explainability, or clarity as values but lacking concrete guidance for implementation.

Exclusion Criteria:

Technical documentation outlining dataset composition, audit methods, or system-level monitoring.

Illustrative Coded Excerpt:

"Transparency and explainability are prerequisites for safeguarding human rights."

Code 2: User_Empowerment_As_Transparency

Definition:

Cases in which user-interface features (editable outputs, feedback forms, confidence indicators) are presented as substitutes for system-level transparency, shifting interpretive responsibility onto users.

Inclusion Criteria:

Corporate text framing user interaction features as mechanisms of transparency.

Exclusion Criteria:

Descriptions of internal logging systems, audit trails, or technical accountability mechanisms.

Illustrative Coded Excerpt:

"Users are responsible for verifying outputs."

Code 3: Distributed_Accountability_PostEditing

Definition:

Segments emphasizing the shifting of responsibility from system developers to translators, post-editors, or end-users, particularly in workflows where MT output must be corrected without access to model-level information.

Inclusion Criteria:

Discussions of accountability fragmentation, post-editing burdens, or unequal knowledge distribution among actors.

Exclusion Criteria:

Clear statements assigning responsibility to developers, model designers, or deployment entities.

Illustrative Coded Excerpt:

"Translators are positioned as post-editors who evaluate system output without insight into how the model generates or filters meaning."

Appendix B. Dataset Overview Table

Table B1.

Overview of the documents included in the dataset

Category	Number of Documents	Publication Years	Notes / Examples
Policy and governance documents	6	2019–2024	OECD AI Principles, UNESCO Recommendations, EU AI Act
Peer-reviewed academic studies	23	2018–2024	Ethics, AI-mediated translation, MT transparency
Corporate and user-facing documents	9	2020–2024	Google Translate, DeepL, ChatGPT-based translation materials guidelines and transparency statements
Total	38	2018–2024	Selected through purposive sampling based on relevance and accessibility

Appendix C. Representative Verbatim Excerpts

The following excerpts illustrate how key concepts transparency, accountability, user responsibility, and operational opacity were coded during the thematic analysis. Each excerpt is brief (≤25 words) and linked to its corresponding data layer (policy, academic, corporate).

C1 — Corporate Disclaimers (Responsibility Shift)

"Users are responsible for verifying outputs." *Corporate user guidance statement.*

C2 — Corporate Ethical Branding

"We ensure our models are trained responsibly."

Corporate transparency/ethics declaration.

C3 — Corporate Interface-Level Transparency

Label examples: "Automatically translated" / side-by-side source and target views.

User-facing UI cues presented as forms of transparency.

C4 — Policy Discourse (UNESCO/OECD)

"Transparency and explainability are prerequisites for safeguarding human rights." *From international policy documents emphasizing normative transparency.*

C5 — Policy (EU AI Act Summary Wording)

"Users must be provided with appropriate information." *High-level transparency requirement for low-risk AI systems, including MT.*

C6 — Academic Coding Definition (NVivo)

"Transparency_As_Rhetoric: references to transparency without operational mechanisms." *NVivo codebook category used during thematic analysis.*

Genişletilmiş Özet

Amaç

Makine çevirisinin (MÇ), özellikle sinirsel makine çevirisi (NMT) ve büyük dil modeli (LLM) temelli sistemlerin hızlı yayılımı, küresel iletişimi ve profesyonel çeviri uygulamalarını köklü biçimde dönüştürmüştür. Ancak bu ilerleme, şeffaflık (algoritmik kararların nasıl alındığı ve iletildiği) ve hesap verebilirlik (makine tarafından üretilen sonuçların ahlaki ve hukuki sorumluluğunu kimin üstlendiği) konularında temel etik soruları da beraberinde getirmektedir. Uzun süredir hem çeviri etiğinin (Chesterman, 2001; Pym, 2012) hem de yapay zekâ (YZ) yönetişiminin (OECD, 2019; UNESCO, 2021; Floridi & Cowls, 2021) merkezinde yer alan bu iki ilke, insan ve insan-olmayan aktörlerin sürekli etkileşim hâlinde olduğu bir ekosistemde yeniden tanımlanmaktadır.

Bu çalışmanın temel amacı, söz konusu dönüşümler karşısında yeni etik normların ortaya çıkıp çıkmadığını ve eğer çıkıyorsa, şeffaflık ile hesap verebilirliğin geliştiriciler, uygulayıcılar ve çevirmenler arasında nasıl yeniden kavramsallaştırıldığını, işletimselleştirildiğini ve dağıtıldığını incelemektir. Çalışma üç temel araştırma sorusuna odaklanmaktadır:

1. Şeffaflık ve hesap verebilirlik güncel makine çevirisi politikalarında ve uygulamalarında nasıl tanımlanmakta ve hayata geçirilmektedir?

2. Bu etik ilkeler, önyargı, halüsinasyon veya yorumlayıcı çarpıtma gibi gerçek dünya MÇ davranışlarıyla hangi noktalarda başarısız olmakta ya da çatışmaktadır?

3. Sosyo-teknik ağlar genelinde sorumlu, adil ve izlenebilir çeviri uygulamalarını güvence altına almak için hangi uygulanabilir çerçeveler önerilebilir?

Bu sorular, politika belgelerinde ilan edilen etik idealler ile ticari MÇ sistemlerinin işleyişindeki yapısal belirsizlik arasındaki giderek büyüyen bir "retorik–gerçeklik boşluğu"ndan kaynaklanmaktadır. OECD (2020, 2023), UNESCO (2021, 2023) ve Avrupa Birliği (2024) gibi kuruluşlar şeffaflık, açıklanabilirlik ve hesap verebilirliğin önemini vurgulasa da, ampirik bulgular bu ilkelerin yalnızca kısmen hayata geçirilebildiğini göstermektedir. Bu çalışma, makine çevirisinin yalnızca teknik bir araç değil, aynı zamanda etik anlamın, yazarlığın ve sorumluluğun dağıtık aktörler arasında sürekli olarak müzakere edildiği bir ahlaki altyapı olarak ele alınması gerektiğini savunmaktadır.

Yöntem

Çalışma, belge analizi, eleştirel söylem analizi ve etik-yorumsamacı akıl yürütmeyi birleştiren karma niteliksel ve kavramsal bir araştırma tasarımını benimsemektedir. Bu yaklaşım, etik ilkelerin kurumsal, akademik ve ticari bağlamlarda nasıl ifade edildiğini ve nasıl hayata geçirildiğini bütüncül biçimde anlamayı mümkün kılmaktadır.

Araştırma Türü: Araştırma hem betimleyici hem de yorumlayıcı niteliktedir ve çeviribilimdeki niteliksel geleneklerle (Baker, 2018; Olohan, 2021) ile yapay zekâ etiği alanındaki yaklaşımlarla (Floridi, 2019; Mittelstadt, 2022) uyum göstermektedir. Hipotez test etmekten ziyade, çalışma, makine çevirisi (MÇ) uygulamalarında ortaya çıkan şeffaflık ve hesap verebilirliğin etik boyutlarını kavramsallaştırmayı ve eleştirel biçimde incelemeyi amaçlamaktadır.

Veri Kaynakları: Birbiriyle ilişkili üç veri alanı seçilmiştir:

1. **Politika ve yönetişim çerçeveleri:** OECD Yapay Zekâ İlkeleri (2019, 2020, 2023), UNESCO Yapay Zekâ Etiği Tavsiyeleri (2021, 2023) ve Avrupa Birliği Yapay Zekâ Tüzüğü (2024).

2. **Akademik literatür:** 2018–2024 yılları arasında yayımlanmış, yapay zekâ aracılı çeviride etik konusunu ele alan yirmiden fazla hakemli çalışma (örn. Kenny & Doherty, 2020; O'Hagan, 2022; Krüger, 2022; Floridi & Cowls, 2021).

3. **Kurumsal belgeler:** Google Translate, DeepL ve ChatGPT Translate'e ait kullanıcıya yönelik materyaller, şeffaflık raporları ve gizlilik politikaları.

Örnekleme: Amaçlı örnekleme yöntemi kullanılmıştır (Palinkas vd., 2015). 2018–2024 dönemine ait erişilebilir ve araştırma sorularıyla doğrudan ilişkili kaynaklara odaklanılmıştır ve hem normatif çerçeveleri hem de bunların pratik uygulamalarını ortaya koyan "bilgi açısından zengin örneklerin" yakalanması hedeflenmiştir.

Veri Toplama ve Analiz: Veriler belge analizi yoluyla toplanmış (Bowen, 2009) ve tematik içerik analizi (Braun & Clarke, 2006) kullanılarak incelenmiştir. Sorumluluğun atfedilmesi, önyargının azaltılması ve açıklanabilirlik gibi temel temalar tümevarımsal olarak kodlanmıştır. Politika söylemi, akademik kuram ve kurumsal iletişimi karşılaştırmak amacıyla katmanlar arası üçgenleme yapılmış; bu sayede tutarsızlıklar ve etik kör noktalar ortaya konmuştur. Verilerin düzenlenmesi ve tematik kümelendirilmesi sürecinde NVivo yazılımından yararlanılmıştır.

Araştırmanın güvenirliğini sağlamak amacıyla veri kaynakları arasında üçgenleme yapılmış; bu sayede inandırıcılık ve doğrulanabilirlik güçlendirilmiştir (Lincoln & Guba, 1985). Araştırmacının yorumlayıcı konumu açık biçimde kabul edilmiş ve yapay zekâ aracılı çeviriye yönelik etik incelemenin doğası gereği öznel boyutlar içerdiği vurgulanmıştır. Çalışma, O'Hagan (2022) ve Krüger'in (2022) savunduğu posthümanist etik yaklaşımı benimseyerek, bireysel sorumluluktan ziyade ilişkisel hesap verebilirliği merkeze almayı hedeflemektedir.

Bulgular

Yapılan analizler, şeffaflık ve hesap verebilirliğin yapay zekâ yönetişiminde merkezi bir söylemsel konuma sahip olduğunu; ancak kavramsal olarak belirsiz ve uygulama düzeyinde yetersiz kaldığını ortaya koymuştur. OECD Yapay Zekâ İlkeleri (2020) ve UNESCO'nun (2021) Yapay Zekâ Etiğine İlişkin Tavsiyesi gibi politika çerçeveleri, bu değerleri güvenin tesis edilmesi ve insan haklarının korunması için önkoşul olarak sıkça vurgulamakta; buna karşın, mülkiyete dayalı makine çevirisi (MÇ) sistemlerinde bu ilkelerin nasıl uygulanacağına dair somut yönlendirmeler sunmamaktadır. Bu durum, politika metinlerinde yer alan "şeffaflık ve açıklanabilirlik insan haklarının korunması için önkoşuldur" gibi ifadelerde açıkça görülmektedir (Appendix C, C4). Bu bulgu, Mittelstadt'ın (2022) etik söylemin yaptırım gücü olan hesap verebilirlik yapıları olmaksızın kullanılmasını tanımladığı "etik yıkama" eleştirisiyle örtüşmektedir.

Kurumsal MÇ sağlayıcıları olan Google, DeepL ve OpenAI, şeffaflığı etik bir ifşa mekanizmasından ziyade bir pazarlama aracı olarak kullanmaktadır. Kamuya açık açıklamalarında sıklıkla "sorumlu yapay zekâ" veya "etik çeviri" vaat edilmekte; ancak veri bileşimi, önyargı denetimlerinin sonuçları ya da yorumlayıcı sınırlılıklar gibi kritik ayrıntılara yer verilmemektedir (Bietti, 2020). Bunun yerine, güven puanları veya düzenlenebilir çıktılar gibi araçlar yoluyla etik yargının yükü kullanıcıya aktarılmaktadır. Ananny ve Crawford'un (2018) işaret ettiği bu söylemsel hamle, epistemik belirsizliği gizleyen bir performatif şeffaflık görünümü üretmektedir.

Bulgular, makine çevirisinde hesap verebilirliğin giderek daha fazla sayıda aktör arasında dağıtıldığını ve bunun da sorumluluğun parçalanmasına yol açtığını göstermektedir. İnsan çevirmenler çoğu zaman, sistemin temel işleyiş mantığını veya veri kaynaklı önyargılarını bilmeden, yapay zekâ tarafından üretilen çıktıları düzelten post-editörler olarak konumlanmaktadır. Bu durum, Krüger'in (2022) "posthüman etik paradoksu" olarak tanımladığı olguyla örtüşmektedir: çevirmenler, yetki ve öznellikten yoksun bırakılırken sorumluluk üstlenmektedir. Özellikle hukuki veya tıbbi çeviri gibi hata payının ciddi sonuçlar doğurabileceği alanlarda, hukuki sorumluluk zinciri büyük ölçüde belirsizliğini korumaktadır (GenLaw Blog, 2023).

Çalışma ayrıca; teknik açıklıktan ziyade yorumlayıcı açıklığı merkeze alan hermeneutik şeffaflık kavramını ortaya koymaktadır. Ricoeur'ün (1976) yorumbilgisi yaklaşımı ve Cronin'in (2022) posthümanist çeviri etiğinden hareketle geliştirilen bu model, şeffaflığı algoritmaların tüm ayrıntılarıyla ifşası olarak değil, anlam üretimine ilişkin bir diyalog olarak yeniden çerçevelendirmektedir. "Otomatik çeviri" etiketlerinin kullanılması ya da kaynak ve hedef metnin yan yana gösterilmesi gibi yeni kurumsal uygulamalar, bu paradigmaya doğru atılan erken adımlar olarak değerlendirilmektedir. Bu tür uygulamalar, kullanıcıların çeviriyi nötr bir bilgi aktarımı değil; yorumlayıcı ve aracılanmış bir edim olarak kavramalarına olanak tanıyan bağlamsal okuryazarlığı teşvik etmektedir.

Sektörler arası sentez, mevcut etik düzenlemelerin teknik açıdan güçlü; ancak ilişkisel açıdan zayıf olduğunu göstermektedir. Politikalar, şeffaflık ve hesap verebilirliği çoğunlukla birer uyum (compliance) ilkesi olarak

kodlamakta; buna karşın güven, karşılıklılık ve diyalojik anlayış gibi ilişkisel etik boyutları göz ardı etmektedir. Bu nedenle çalışma, etiğin üç ana ilke temelinde yeniden düşünülmesini önermektedir:

- Yorum olarak şeffaflık: teknik ayrıntılar yerine yorumlayıcı çerçevelerin açığa çıkarılması,

- İlişki olarak hesap verebilirlik: tüm aktörler arasında paylaşılan ve izlenebilir sorumluluk,

- Uygulama olarak etik: sabit bir uyum listesi yerine sürekli müzakere edilen bir süreç.

Bu unsurlar bir araya geldiğinde, makine çevirisi etiği posthümanist bir ahlaki ekoloji içinde yeniden yapılandırılmakta; öznellik ve anlamın insanlar ile makineler tarafından birlikte inşa edildiği kabul edilmektedir.

Sınırlılıklar

Bu çalışma, ampirik ölçümler yerine kavramsal ve yorumlayıcı bir yaklaşımı bilinçli olarak benimsemektedir. Bunun temel nedeni, yapay zekâ aracılı çeviride şeffaflık ve hesap verebilirlik gibi etik ilkelerin nicel göstergelerle doğrudan ölçülebilir olmaktan ziyade, söylemsel, normatif ve bağlamsal yapılar olarak ortaya çıkmasıdır. Bu nedenle analiz, etik ilkelerin politika metinlerinde, akademik tartışmalarda ve kurumsal söylemlerde nasıl tanımlandığını ve çerçevelendiğini ortaya koymayı amaçlamaktadır.

Kamuya açık belgelerle çalışılması, mülkiyete tabi sistemlerin sınırlı erişilebilirliği göz önüne alındığında, alan yazında yaygın ve geçerli bir yöntemdir. Bu yaklaşım, kurumsal şeffaflık ve hesap verebilirliğe ilişkin değerlendirmelerin söylemsel düzeyde yapılmasını mümkün kılmakta ve ticari sistemlerin etik iddiaları ile pratikleri arasındaki yapısal tutarsızlıkları görünür kılmaktadır. Bu yönüyle çalışma, ampirik genelleme iddiasında bulunmaksızın, alandaki etik tartışmaların kavramsal derinliğini artırmayı ve gelecekte yapılacak ampirik araştırmalar için sağlam bir kuramsal çerçeve sunmayı hedeflemektedir.

Öneriler

Bu çalışmanın bulguları doğrultusunda, makine çevirisi ve yapay zekâ etiği alanında bütüncül bir yaklaşımın benimsenmesi önerilmektedir. Çeviri etiğinin, yapay zekâ yönetişimi ve posthümanist kuramla birlikte ele alınması; etik sorumluluğun bireysel aktörlerle sınırlı olmayan, sosyo-teknik sistemler içinde dağıtık bir yapı olarak yeniden kavramsallaştırılmasını gerekli kılmaktadır. Bu bağlamda hermeneutik şeffaflık, yapay zekâ destekli çeviride anlam üretimi ve aracılık süreçlerini açıklamak için kuramsal bir çerçeve sunmaktadır. Uygulama düzeyinde, makine çevirisi sağlayıcılarının şeffaflığı soyut etik taahhütlerin ötesine taşıyarak çeviri çıktılarının nasıl üretildiğini ve hangi sınırlılıkları içerdiğini açıklayan bağlamsal bilgilendirme mekanizmaları (üstveri etiketleri, model kartları ve köken göstergeleri) geliştirmesi önem taşımaktadır. Profesyonel çevirmenler açısından ise algoritmik önyargı ve yorumlayıcı aracılığa ilişkin farkındalığın, çevirmen eğitimi ve sertifikasyon süreçlerinde temel bir yeterlilik alanı hâline getirilmesi; yapay zekâ etiği ve dağıtık hesap verebilirlik konularının eğitim programlarına dâhil edilmesi önerilmektedir. Politika düzeyinde, düzenleyici çerçevelerin yalnızca teknik belgelendirmeyle sınırlı kalmayıp çevirinin yorumlayıcı ve kültürel boyutlarını da kapsayacak biçimde genişletilmesi; şeffaflık yükümlülüklerinin uzman olmayan kullanıcılar için de erişilebilir olması ve sorumluluk zincirlerinin insan ile insan-dışı aktörleri içerecek şekilde tanımlanması gerekmektedir. Son olarak, makine çevirisinde etik şeffaflığın güçlendirilmesi, kamusal güvenin artırılmasına ve özellikle marjinalleştirilmiş dil topluluklarını etkileyen dilsel eşitsizlik ve kültürel önyargı risklerinin azaltılmasına katkı sağlayarak dilsel adaletin desteklenmesine olanak tanıyabilir.

Özgün Değer

Bu çalışma, çeviribilim ve yapay zekâ etiği alanlarına hem kuramsal hem de uygulamaya dönük özgün katkılar sunmaktadır. Çalışmanın temel özgünlüğü, çeviri etiği, yapay zekâ yönetişimi ve posthümanist felsefeyi sistematik bir çerçevede bir araya getirerek, bugüne kadar çoğunlukla ayrı disiplinler içinde ele alınmış etik tartışmaları bütüncül bir perspektifle yeniden düşünmesinde yatmaktadır. Bu doğrultuda geliştirilen hermeneutik şeffaflık kavramı, şeffaflığı teknik bir özellik ya da algoritmik açıklanabilirlik meselesi olmaktan çıkararak, anlam üretimi, aracılık ve sorumluluğun yorumlayıcı boyutlarını merkeze alan yenilikçi bir etik kavramsallaştırma sunmaktadır. Ayrıca çalışma, etik sorumluluğu bireysel aktörlere indirgemek yerine, insan ve insan-dışı bileşenlerden oluşan sosyo-teknik ağlar içinde paylaşılan ve izlenebilir bir ilişki olarak ele alan ilişkisel bir etik çerçeve önermektedir.