

ETHICAL CHALLENGES IN ARTIFICIAL INTELLIGENCE-BASED TRANSLATION

Dr. Ananya Sharma

Assistant Professor

Department of Computer Science and Artificial Intelligence

School of Computing and Information Technology

University of Technology and Innovation, New Delhi, India

Abstract

Artificial intelligence-based translation is now used in education, cross-border commerce, public services, health communication, humanitarian response, research, and everyday interpersonal exchange. Neural machine translation systems and large language model interfaces can reduce language barriers at remarkable speed, yet their apparent fluency can hide ethical questions about meaning, representation, privacy, responsibility, and access. Translation is not merely a word-substitution task; it involves social context, cultural knowledge, power relations, and consequences for people who must rely on the translated output.

This paper examines the principal ethical challenges associated with artificial intelligence-based translation. It considers semantic fidelity, cultural and linguistic representation, gender and social bias, low-resource language inequality, privacy and confidentiality, intellectual property, labour implications, transparency, accountability, and human oversight. The analysis treats translation systems as socio-technical infrastructures whose impacts depend on training data, model design, interface choices, deployment settings, institutional policies, and the opportunities available to users to question or correct harmful outputs.

The study uses an integrative review and conceptual analysis of scholarship on machine translation, natural language processing ethics, translation studies, data governance, and trustworthy artificial intelligence. It distinguishes performance quality from ethical acceptability, because high automatic evaluation scores cannot establish that a system preserves dignity, protects confidential information, or represents communities fairly. The review also examines how risk varies across high-stakes domains, including legal, medical, educational, and public-service communication.

The paper argues that responsible translation requires governance across the full system life cycle: lawful and documented data practices, community-aware language coverage, bias evaluation, context-sensitive human review, clear user communication, privacy controls, auditability, accessible redress, and fair treatment of professional translators and language communities. Ethical AI translation should therefore be evaluated not only by speed and fluency, but also by whether it expands communication without weakening rights, accuracy, cultural integrity, or public trust.

Keywords: Artificial Intelligence, Machine Translation, Translation Ethics, Language Bias, Cultural Representation, Privacy, Low-Resource Languages, Human Oversight, Algorithmic Accountability, Trustworthy AI

1. INTRODUCTION

Artificial intelligence-based translation has become a familiar layer of digital communication. It appears in search engines, mobile applications, messaging platforms, social-media interfaces, customer-service systems, document workflows, online learning environments, and speech technologies. Neural machine translation and more recent generative AI systems can produce a translation within seconds, making multilingual communication more immediate for people who would otherwise lack access to

a human translator. This capability has significant social value, especially where language barriers limit participation in education, healthcare, justice, migration services, or economic activity.

However, the social importance of translation also makes errors ethically significant. A fluent sentence can still distort a speaker's intent, omit a condition, soften a warning, assign an unintended gender, or remove a culturally meaningful expression. In routine and low-risk settings, such problems may be inconvenient. In legal, clinical, emergency, or

administrative settings, they may influence rights, safety, eligibility, consent, or access to public services. The ethical question is therefore not only whether an output is grammatically acceptable, but whether the translation is appropriate for the people, context, and decision at stake.

The risks are intensified by unequal linguistic resources. Systems generally perform better for languages represented by abundant digitised text, professional parallel corpora, standardised scripts, and commercial demand. Many regional, Indigenous, minority, and low-resource languages have limited data, unstable orthographies, scarce evaluation materials, or no meaningful community role in system development. A technology promoted as universally inclusive can thus reproduce existing linguistic hierarchies by giving some communities reliable access while offering others incomplete, culturally insensitive, or unavailable service.

Ethical concerns also arise from the data and interfaces behind translation. Users may enter personal messages, business documents, clinical notes, immigration records, legal materials, unpublished research, or culturally sensitive content into cloud-based tools without knowing how text is retained, processed, or reused. Developers and deploying organisations may rely on training data whose origins, licences, consent conditions, and representational limits are not fully documented. The distribution of responsibility can become unclear when an error is produced through the interaction of a model, a platform, a user interface, a vendor policy, and an institutional workflow.

The objective of this paper is to examine the ethical challenges that shape the use and legitimacy of artificial intelligence-based translation. The paper reviews relevant literature, defines the object of study, explains a review-based methodology, analyses major ethical risks, and proposes principles for responsible, transparent, inclusive, and human-centred translation systems.

2. LITERATURE REVIEW

Research on machine translation has traditionally concentrated on model architecture, data availability, evaluation metrics, and error reduction. These priorities remain important, but recent scholarship shows that translation quality cannot be reduced to automatic scores or average benchmark performance. A system may score well on common test sets while producing misleading gender assignments, weakening register, mishandling cultural references, or failing on the kinds of texts that users need most. Human-centred evaluation is therefore increasingly important, particularly where translation is used in sensitive or consequential settings.

Studies of bias in machine translation demonstrate that target-language grammar and statistical regularities can convert ambiguous or gender-neutral source expressions into stereotyped outputs. Research also identifies broader representational concerns, including unequal treatment of names, dialects, honorifics, identities, and culturally situated meanings. These findings connect machine translation to wider debates in natural language processing about dataset provenance, social bias, model documentation, and the exclusion of communities whose languages or communication practices are poorly represented in technological systems.

Policy and governance literature contributes an additional perspective. UNESCO and the OECD emphasise human rights, fairness, transparency, accountability, and inclusion in AI governance, while the NIST AI Risk Management Framework offers a structured approach to identifying and managing risks. Legal and regulatory developments, including data-protection rules and risk-based AI requirements, reinforce the need for organisations to understand their role when translation tools are deployed in professional workflows. The following table summarises key contributions that inform the analysis of ethics in AI-based translation.

Table 1: Summary of Key Literature on Ethics in AI-Based Translation

S. No.	Author(s) / Year	Ethical Focus	Study Context	Key Contribution
1	Hovy & Spruit (2016)	Social impacts of NLP	Natural language processing	Shows why language technologies must be assessed

				for social consequences, not only technical performance.
2	Bender et al. (2021)	Data and model risks	Large language models	Explains how scale, undocumented data, and detached deployment can create foreseeable harms.
3	Blodgett et al. (2020)	Language and power	NLP research	Connects language technology harms to social structures, power, and unequal representation.
4	Stanovsky et al. (2019)	Gender bias	Machine translation evaluation	Introduces an evaluation approach for analysing gender bias across translation systems.
5	Prates et al. (2020)	Occupational stereotypes	Gender-neutral source languages	Demonstrates how translation may assign gendered roles based on social stereotypes.
6	Savoldi et al. (2021)	Bias research review	Machine translation	Synthesises conceptual, methodological, and mitigation challenges in gender-bias research.
7	Saunders & Byrne (2020)	Bias mitigation	Neural machine translation	Examines domain-adaptation approaches to reducing gender bias in translation.
8	Orife et al. (2020)	Participation and inclusion	African-language translation	Shows the value of participatory work for low-resource translation development.
9	Kreutzer et al. (2022)	Dataset documentation	Web-crawled multilingual data	Identifies quality, provenance, and representational concerns in large multilingual datasets.
10	Mager et al. (2023)	Indigenous language ethics	Community-facing translation	Argues that language speakers should be central to the design and governance of translation systems.
11	UNESCO (2021)	Human rights and dignity	Global AI governance	Provides values and policy principles for ethical AI development and use.
12	NIST (2023)	Risk management	AI system governance	Offers a framework for governing, mapping, measuring, and managing AI risks.

3. OBJECT OF STUDY

The object of this study is the ethical governance of artificial intelligence-based translation. The analysis includes neural machine translation engines, large language model translation interfaces, speech-translation applications, automated subtitle systems, enterprise translation platforms, and embedded translation features in digital services. It considers translation as a process involving data collection, model development, evaluation, deployment, human use, feedback, monitoring, and revision.

The study examines the interaction between linguistic performance and ethical responsibility. Its principal areas include semantic fidelity, cultural context, gender and social bias, language inequality, privacy, confidentiality, consent, intellectual property, transparency, accountability, professional work, and human oversight. These areas are interdependent: a model may be accurate on average yet still harm a user through a biased assumption, a privacy failure, or an output that is unsuitable for a high-stakes decision.

The analysis also considers the relationship between translation technology and language communities. A responsible system should not treat people only as data providers, error reporters, or end users. It should recognise their role in defining acceptable use, identifying context-specific risks, evaluating quality, shaping language representation, and receiving a fair share of the benefits created when their linguistic resources are incorporated into technological products.

KEY POINTS OF OBJECT OF STUDY

- To map ethical risks across the complete life cycle of AI-based translation systems.
- To examine semantic fidelity, contextual suitability, and the consequences of translation errors.
- To analyse gender, social, cultural, and dialectal bias in translated outputs.
- To evaluate language-resource inequality and the exclusion of low-resource or marginalised language communities.
- To investigate privacy, confidentiality, consent, data retention, and cross-border data handling.
- To assess data provenance, licensing, intellectual-property concerns, and documentation practices.
- To consider transparency, explainability, auditability, and the allocation of responsibility for harm.
- To examine the role of professional translators, reviewers, interpreters, and community evaluators.
- To identify human-oversight requirements for high-stakes domains and vulnerable users.
- To develop practical principles for trustworthy, inclusive, and rights-respecting AI translation.

Table 2: Object of Study Framework

S. No.	Study Component	Description	Purpose
1	Translation Life Cycle	Data sourcing, model training, evaluation, deployment, feedback, and revision.	Locate ethical responsibility across the system rather than only at the final output.
2	Semantic Fidelity	Meaning preservation, register, context, ambiguity, and error severity.	Assess whether translated text is reliable and appropriate for its intended use.
3	Bias and Representation	Gender, social identity, cultural references, dialects, and language hierarchy.	Identify harms caused by stereotyped or exclusionary translations.
4	Language Equity	Coverage, quality gaps, community participation, and access for low-resource languages.	Prevent unequal benefits and burdens across linguistic communities.
5	Privacy and Data Rights	Confidential text, consent, retention, access control, and lawful processing.	Protect users and organisations from unauthorised disclosure or reuse.
6	Transparency and Accountability	Notices, documentation, audit trails, reporting, and redress.	Make system limits and responsibility visible to users and institutions.
7	Human Oversight and Work	Review, escalation, professional roles, labour conditions, and user control.	Ensure that automation supports rather than displaces accountable expertise.
8	Public Trust and Inclusion	Accessibility, language accessibility, participation, and accessible complaint channels.	Ensure that translation benefits are meaningful, fair, and socially legitimate.

4. RESEARCH METHODOLOGY

This paper uses an explanatory and integrative review design. Rather than collecting a new

corpus or conducting an experiment with a single translation engine, it synthesises scholarship on machine translation,

computational linguistics, translation studies, AI ethics, data governance, and human-rights-oriented AI policy. The approach is appropriate because the research question concerns the organisational, technical, cultural, and ethical conditions through which translation systems are developed and used across diverse settings.

The review is organised around four analytical domains: linguistic fidelity and contextual adequacy; bias, representation, and language equity; privacy, data governance, and intellectual property; and accountability, human oversight, labour, and public trust. Sources are considered for their contribution to understanding translation-system risks, the kind of evidence provided, the linguistic or institutional context addressed, and the implications for implementation. Particular attention is given to the distinction between performance metrics and the lived consequences of translated outputs.

To preserve analytical clarity, the paper distinguishes between translation assistance and autonomous decision support, between a minor stylistic error and a materially harmful error, between public text and confidential text, and between broad language coverage and meaningful linguistic inclusion. A platform may support many languages while offering unreliable quality for some of them; it may disclose an AI label without providing useful information about limitations; or it may invite user feedback without offering a process for redress. These distinctions guide the interpretation of evidence and the development of recommendations.

4.1 RESEARCH DESIGN

- Study Type: Explanatory and literature-based research.
- Approach: Integrative review and conceptual synthesis.
- Unit of Analysis: AI-based translation systems, multilingual datasets, deployment contexts, and governance arrangements.
- Primary Ethical Domains: Fidelity, representation, privacy, accountability, labour, and inclusion.
- Context Variables: Language-resource level, domain risk, user vulnerability, data sensitivity, and institutional capacity.

- Outcome Variables: Translation reliability, rights protection, equitable access, transparent use, and user trust.
- Ethical Position: No personal, confidential, or human-subject data were collected for this study.

4.2 METHODOLOGICAL STEPS

- Step 1: Define AI-based translation as a socio-technical process rather than a standalone software output.
- Step 2: Identify major stages of the translation system life cycle, from data acquisition to post-deployment monitoring.
- Step 3: Review literature on translation quality, social bias, language inequality, privacy, and data documentation.
- Step 4: Classify ethical challenges according to linguistic, technical, legal, cultural, organisational, and social dimensions.
- Step 5: Compare proposed controls, including dataset documentation, risk assessment, bias testing, human review, and user notices.
- Step 6: Interpret evidence in relation to high-stakes domains, low-resource languages, and platform-based deployment.
- Step 7: Identify limitations, including varied language contexts, rapidly evolving models, and incomplete public information about proprietary systems.
- Step 8: Develop principles for responsible, transparent, participatory, and human-centred AI translation.

5. DATA ANALYSIS

The data analysis is based on qualitative synthesis of the reviewed literature. The evidence is interpreted by tracing how data practices, model behaviour, interface design, and institutional decision-making influence the ability of a translation system to preserve meaning, represent people fairly, protect confidential information, and remain accountable when errors occur. The analysis does not report newly collected translation observations; instead, it consolidates recurring patterns and identifies the conditions under which ethical weaknesses are most likely to affect users.

Across the reviewed studies, the most consistent message is that translation quality

cannot be separated from social context. A model can perform well on benchmark sentences while failing to preserve meaning in a court notice, a medical discharge instruction, a culturally marked statement, or a dialogue where gender, formality, kinship, or respectful address matter. The following sections organise the analysis around four challenge areas that are especially important for artificial intelligence-based translation.

5.1 SEMANTIC FIDELITY, CONTEXT, AND CULTURAL REPRESENTATION

The first ethical challenge concerns the relationship between linguistic fluency and semantic fidelity. Translation systems often produce smooth, confident sentences even when they have misunderstood ambiguity, domain terminology, negation, legal modality, emotional tone, sarcasm, politeness, or culturally specific references. This can encourage overreliance because users may equate natural-sounding output with correctness. The risk is particularly serious where a user cannot independently assess the target language or where the translation is used to make a decision affecting health, rights, safety, or employment.

Translation also involves choices about register and relationship. A literal rendering may remove respectful address, change a request into an instruction, soften an obligation, or make a speaker appear less certain or less credible than intended. Cultural expressions, local concepts, kinship terms, humour, and religious references may require explanation, adaptation, or human consultation rather than surface-level substitution. A system that ignores this context can produce a grammatically valid output that still misrepresents the person or community being translated.

Ethically responsible evaluation should therefore include context-sensitive human assessment. Automatic metrics are useful for engineering comparison, but they cannot by themselves determine whether the output is safe, fair, culturally appropriate, or sufficient for a particular purpose. Deployers should identify high-risk use cases, define escalation routes, communicate limitations, and require qualified review when the potential consequence of an error is high. Translation assistance can be valuable, but it should not be

presented as equivalent to certified interpretation or professional translation when the setting requires accountable expertise.

KEY OBSERVATION:

- Fluency can conceal serious meaning errors, especially when users cannot verify the target language.
- Register, cultural context, and domain-specific terminology are ethical as well as linguistic concerns.
- High-stakes translation requires contextual human review, clear limitations, and proportionate escalation procedures.

5.2 BIAS, LANGUAGE INEQUALITY, AND REPRESENTATIONAL HARM

Bias in machine translation is often visible when a system must translate between languages with different gender-marking conventions. A gender-neutral sentence may be rendered with a masculine or feminine occupational role based on learned stereotypes, and a translation may assign family roles, status, or personality traits in ways that do not appear in the source text. These outcomes are not harmless technical curiosities. They can normalise social stereotypes, misrepresent individuals, and reinforce discriminatory assumptions in communication that may influence education, recruitment, public services, or personal relationships.

Representational harm extends beyond gender. Systems can mishandle racialised names, dialects, code-switching, indigenous terms, social identities, disabled people's communication, and culturally situated language. Translation models generally reflect the distributions and editorial conventions of their training data, which may privilege dominant languages, standard forms, and commercially visible content. When a language community is represented only through sparse, extracted, or externally collected text, the model may fail to recognise ordinary forms of expression or may treat the community's language as a deficit rather than a living source of knowledge and identity.

Language inequality is therefore a governance problem, not only a resource problem. Improved coverage should not be defined solely by the number of languages listed in an interface. It should consider quality, accessibility, community authority, data

provenance, local evaluation, and whether people can contest harmful translations. Participatory development, community-based evaluation, fair compensation, and locally relevant documentation can help avoid extractive approaches in which linguistic data are taken from communities while the benefits, control, and economic value remain elsewhere.

KEY OBSERVATION:

- Gender-neutral source text can be translated into stereotyped target-language forms when systems rely on biased statistical patterns.
- Language coverage is not equivalent to language equity; reliability and community participation matter.
- Bias evaluation should include social identities, dialects, cultural expressions, and real user contexts, not only benchmark averages.

5.3 PRIVACY, CONFIDENTIALITY, AND DATA GOVERNANCE

AI translation services often operate through cloud infrastructure and are designed for quick copy-and-paste use. This convenience can create a confidentiality risk when users submit personal correspondence, health information, legal documents, financial records, unpublished manuscripts, commercial contracts, or internal government material. The ethical issue is not merely whether the output is accurate. It is whether users understand what happens to their text, who can access it, how long it is retained, whether it may be used for product improvement, and whether the transfer is lawful under the relevant data-protection and confidentiality obligations.

Responsible data governance requires clear purpose limitation, data minimisation, privacy notices written for actual users, security controls, access restrictions, retention schedules, secure deletion procedures, and audit trails. Organisations should distinguish between public or low-sensitivity text and material that requires protected translation environments. They should provide staff with clear policies about when consumer tools are unsuitable and should assess whether enterprise services offer contractual safeguards that match the sensitivity of the data and the jurisdictional context.

Data governance also includes provenance and intellectual-property questions. Training data

can contain copyrighted text, community-generated material, translations created by professionals, and content collected from the web without a direct relationship to the original authors or speakers. Documentation should clarify known sources, licences, limitations, filtering decisions, and use restrictions. While transparency may be constrained by commercial secrecy or data-scale complexity, a total absence of documentation makes it difficult for communities, researchers, and deployers to evaluate risk or seek correction.

KEY OBSERVATION:

- Users may unintentionally expose confidential content when translation tools are embedded in routine digital workflows.
- Privacy-by-design requires minimisation, clear notices, secure processing, retention limits, and usage policies tailored to risk.
- Dataset provenance and licensing documentation are necessary for accountable translation-system development and deployment.

5.4 ACCOUNTABILITY, HUMAN OVERSIGHT, AND THE FUTURE OF TRANSLATION WORK

Responsibility becomes difficult to locate when a translation error emerges from a complex chain of actors. A model developer may claim that the system is a general-purpose tool; a platform may describe the output as informational; an organisation may integrate it into a workflow; and a frontline employee or user may rely on it under time pressure. Yet the person affected by the error may experience the result as a denial of service, a breach of confidence, or a misrepresentation of their voice. Ethical governance requires that responsibility be assigned before harm occurs, not negotiated only after a complaint is made. Human oversight should be matched to the purpose and consequence of the translation. In low-risk situations, users may reasonably choose automated assistance with visible uncertainty and easy correction. In high-stakes settings, systems should support qualified translators, interpreters, reviewers, or domain experts rather than replace them without safeguards. Human involvement must be meaningful: reviewers need adequate time, language competence, access to source

context, authority to correct or reject output, and an escalation process for cases where the model's limitations could cause harm.

The wider translation profession is also affected by automation. AI can reduce repetitive work, improve terminology management, and help translators explore alternatives, but it may also intensify productivity pressure, shift risk onto freelancers, reduce compensation for post-editing, or encourage organisations to treat linguistic expertise as interchangeable. Fair deployment should recognise professional judgement, protect labour conditions, disclose AI use to clients and end users where relevant, and avoid business models that extract value from translators' work while weakening their ability to control how their translations or feedback are reused.

KEY OBSERVATION:

- Accountability must be allocated across developers, vendors, deployers, and institutional users before a harmful error occurs.
- Human review is meaningful only when reviewers have context, competence, time, authority, and a clear escalation path.
- Responsible adoption should support professional translators and community evaluators rather than treating their expertise as a disposable safeguard.

6. DISCUSSION

The reviewed literature demonstrates that ethical translation is a foundational condition for legitimate AI-mediated communication. Translation systems can expand access to information, reduce practical barriers, and support multilingual exchange, but these benefits are not automatic. The quality of outcomes depends on whether developers and deployers can establish trusted practices for data sourcing, language representation, evaluation, user communication, privacy protection, and error management. A technology-first approach may produce impressive demonstrations while leaving unresolved questions about whose language is represented, who bears the risk of failure, and who can seek correction.

Semantic fidelity should be understood as a public-interest issue. In many settings, an inaccurate translation does not remain a private inconvenience; it can change how a

person is perceived, whether they understand a warning, or whether they can exercise a right. Risk-based deployment is therefore necessary. Organisations should determine which uses are low risk, which require contextual review, and which should not rely on automated translation at all. This requires more than a generic disclaimer. It requires documented use cases, staff training, appropriate procurement, and clear routes for challenging errors.

Bias and language inequality require a similarly broad response. Technical debiasing can be helpful, but it cannot fully resolve representational harms when data, benchmarks, or product priorities exclude communities. Translation development should involve speakers, educators, translators, researchers, and local institutions in ways that give them influence over data use, evaluation criteria, and acceptable deployment. Participatory governance is especially important for Indigenous, minority, and low-resource languages, where the apparent availability of data may not establish community consent or cultural appropriateness.

Privacy and accountability must be treated as design requirements rather than post-deployment compliance tasks. Translation interfaces should make data practices intelligible, distinguish sensitive from non-sensitive use, and offer organisational settings that protect confidential content. Providers should document important limitations, provide mechanisms for reporting material errors, and preserve sufficient auditability for independent assessment. Institutions should avoid procurement arrangements that obscure data handling, prevent quality review, or make continued access dependent on opaque proprietary systems.

The discussion also shows why success metrics need to extend beyond speed, language count, and benchmark scores. Responsible evaluation should include error severity, user comprehension, equity across language varieties, privacy impact, cultural adequacy, accessibility, labour effects, and the availability of redress. Ethical AI translation is not achieved by eliminating every error, which is unrealistic, but by creating systems and institutions that identify foreseeable harms, make limitations visible, protect people from

disproportionate risk, and learn from community feedback without exploiting it.

7. CONCLUSION

This paper examined the ethical challenges associated with artificial intelligence-based translation. The analysis shows that translation systems are shaped not only by computational performance but also by the quality and provenance of data, semantic and cultural context, bias, language inequality, privacy safeguards, intellectual-property practices, accountability arrangements, human oversight, and the distribution of benefits and risks. These factors explain why a system can be technically impressive while still being unsuitable for a particular community, domain, or decision.

The evidence reviewed in this paper indicates that ethical AI translation requires documented data practices, contextual and community-aware evaluation, bias testing across relevant identities and language varieties, privacy-by-design, transparent user communication, risk-based human review, auditability, accessible

redress, and fair recognition of translator and community expertise. These measures do not prevent innovation. Rather, they create the conditions under which translation technology can be used reliably and in ways that respect human dignity, cultural integrity, and democratic accountability.

The paper concludes that trustworthy translation should be treated as a continuing governance responsibility. Developers, platform providers, organisations, and public institutions should combine technical improvement with rights-respecting design, participatory language governance, proportionate safeguards, and continuous monitoring of real-world effects. Future research should investigate ethical evaluation across additional language pairs, community-led models of data stewardship, translation in high-stakes public services, and the implications of generative AI systems that combine translation with summarisation, rewriting, and conversational assistance.

8. REFERENCES

1. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the dangers of stochastic parrots: Can language models be too big?* In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). <https://doi.org/10.1145/3442188.3445922>
2. Blodgett, S. L., Barocas, S., Daume III, H., & Wallach, H. (2020). *Language (technology) is power: A critical survey of bias in NLP*. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5454-5476). <https://doi.org/10.18653/v1/2020.acl-main.485>
3. Bostrom, N., & Yudkowsky, E. (2014). *The ethics of artificial intelligence*. In K. Frankish & W. M. Ramsey (Eds.), *The Cambridge Handbook of Artificial Intelligence* (pp. 316-334). Cambridge University Press.
4. Costa-jussa, M. R., Cross, J., Celebi, O., Elbayad, M., Heafield, K., Heffernan, K., Kalbassi, E., Lam, J., Maillard, J., & others. (2022). *No language left behind: Scaling human-centered machine translation*. [arXiv:2207.04672](https://arxiv.org/abs/2207.04672).
5. European Parliament and Council. (2016). *Regulation (EU) 2016/679 (General Data Protection Regulation)*. *Official Journal of the European Union*, L119, 1-88.
6. European Parliament and Council. (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. *Official Journal of the European Union*.
7. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daume III, H., & Crawford, K. (2021). *Datasheets for datasets*. *Communications of the ACM*, 64(12), 86-92. <https://doi.org/10.1145/3458723>
8. Guerberof-Arenas, A., & Moorkens, J. (2023). *Ethics and machine translation: The end user perspective*. In *Towards Responsible Machine Translation: Ethical and Legal Considerations in Machine Translation* (pp. 113-133). Springer.
9. Hovy, D., & Spruit, S. L. (2016). *The social impact of natural language processing*. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (pp. 591-598). <https://doi.org/10.18653/v1/P16-2096>

10. Kreutzer, J., Caswell, I., Wang, L., Wahle, J. P., Jansen, A., Riezler, S., & others. (2022). Quality at a glance: An audit of web-crawled multilingual datasets. *Transactions of the Association for Computational Linguistics*, 10, 50-72. https://doi.org/10.1162/tacl_a_00447
11. Larson, B. (2017). Gender as a variable in natural-language processing: Ethical considerations. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing* (pp. 1-11). <https://doi.org/10.18653/v1/W17-1601>
12. Liebling, D. J., Heller, K., Deng, W. H., & others. (2022). Opportunities for human-centered evaluation of machine translation. In *Findings of the Association for Computational Linguistics: NAACL 2022* (pp. 245-259).
13. Mager, M., Mager, E., Kann, K., & Vu, N. T. (2023). Ethical considerations for machine translation of Indigenous languages: Giving a voice to the speakers. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*. <https://doi.org/10.18653/v1/2023.acl-long.268>
14. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220-229). <https://doi.org/10.1145/3287560.3287596>
15. National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.100-1>
16. OECD. (2019). Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
17. Orife, I., Ezeani, I., Abdulkadir, D., Kamper, H., Elshahar, H., Duru, G., & others. (2020). Participatory research for low-resourced machine translation: A case study in African languages. In *Findings of EMNLP 2020* (pp. 2144-2160). <https://doi.org/10.18653/v1/2020.findings-emnlp.194>
18. Prates, M. O. R., Avelar, P. H., & Lamb, L. C. (2020). Assessing gender bias in machine translation: A case study with Google Translate. *Neural Computing and Applications*, 32(10), 6363-6381. <https://doi.org/10.1007/s00521-019-04144-6>
19. Prudhommeaux, E. (2022). Not always about you: Prioritizing community needs when developing endangered language technology. In *Proceedings of ACL 2022* (pp. 3931-3942). <https://doi.org/10.18653/v1/2022.acl-long.272>
20. Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33-44). <https://doi.org/10.1145/3351095.3372873>
21. Saunders, D., & Byrne, B. (2020). Reducing gender bias in neural machine translation as a domain adaptation problem. In *Proceedings of ACL 2020* (pp. 7724-7736). <https://doi.org/10.18653/v1/2020.acl-main.690>
22. Savoldi, B., Gaido, M., Bentivogli, L., Negri, M., & Turchi, M. (2021). Gender bias in machine translation. *Transactions of the Association for Computational Linguistics*, 9, 845-874. https://doi.org/10.1162/tacl_a_00401
23. Stanovsky, G., Smith, N. A., & Zettlemoyer, L. (2019). Evaluating gender bias in machine translation. In *Proceedings of ACL 2019* (pp. 1679-1684). <https://doi.org/10.18653/v1/P19-1164>
24. UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. United Nations Educational, Scientific and Cultural Organization.
25. United Nations. (2007). United Nations Declaration on the Rights of Indigenous Peoples. United Nations General Assembly Resolution 61/295.
26. Kenny, D. (Ed.). (2022). *Machine translation for everyone: Empowering users in the age of artificial intelligence*. Language Science Press. <https://doi.org/10.5281/zenodo.6759982>
27. Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector: Applications and challenges. *International Journal of Public Administration*, 42(7), 596-615. <https://doi.org/10.1080/01900692.2018.1498103>

-
28. World Intellectual Property Organization. (2020). Revised issues paper on intellectual property policy and artificial intelligence. WIPO Conversation on Intellectual Property and Artificial Intelligence.
 29. Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., Goodman, A., Hollander, R., Koenig, B. A., Metcalf, J., Narayanan, A., Nelson, A., & Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLOS Computational Biology*, 13(3), e1005399. <https://doi.org/10.1371/journal.pcbi.1005399>
 30. Zoph, B., Yuret, D., May, J., & Knight, K. (2016). Transfer learning for low-resource neural machine translation. In *Proceedings of EMNLP 2016* (pp. 1568-1575). <https://doi.org/10.18653/v1/D16-1163>