



LANGUAGE
AND LAW

LINGUAGEM
E DIREITO

VOLUME 12.1

ISSN 2183-3745

Language and Law Linguagem e Direito

ISSN: 2183-3745 (online)
Special Issue, 2025

Editors / Diretores

Malcolm Coulthard & Rui Sousa-Silva
Aston University, UK & Universidade do Porto, Portugal

Special Issue / Número especial

Edited by / Editado por

Angela Soltan, Rebekah Rousi & Lucía Ruiz Rosendo
State University of Moldova, University of Vaasa, Finland
& University of Geneva, Switzerland

Book Reviews Editors / Editores de Recensões

Rita Faria (Portugal) & Luciane Fröhlich (Brasil)
Universidade do Porto, Portugal &
Universidade Federal de Santa Catarina, Brasil

Cover / Capa

Rui Effe

Publisher / Editora

Faculdade de Letras da Universidade do Porto

This publication was supported by project UID/00022/2025, FCT-Fundação para a Ciência e Tecnologia, Portugal, through national funds.

Esta publicação é financiada por fundos nacionais através da FCT - Fundação para a Ciência e a Tecnologia, I.P., no âmbito do projeto UID/00022/2025.

International Editorial Board / Conselho Editorial Internacional

Janet Ainsworth, *University of Washington, USA*

Carmen Rosa Caldas-Coulthard, *University of Birmingham, UK*

Le Cheng, *Zhejiang University, China*

Virginia Colares, *Universidade Católica de Pernambuco, Brasil*

Diana Eades, *University of New England, Australia*

Debora Figueiredo, *Universidade Federal de Santa Catarina, Brasil*

Maribel del Pozo Triviño, *Universidad de Vigo, Spain*

Ed Finegan, *University of Southern California, USA*

Tammy Gales, *Hofstra University, USA*

Núria Gavaldà, *Universitat Autònoma de Barcelona, Spain*

Maria Lúcia Gomes, *Universidade Tecnológica Federal do Paraná, Brasil*

Tim Grant, *Aston University, UK*

Patrick Juola, *Duquesne University, USA and Juola Associates*

Krzysztof Kredens, *Aston University, UK*

Iman Laversuch, *University of Cologne, Germany*

Janny Leung, *University of Hong Kong, Hong Kong*

Belinda Maia, *Universidade do Porto, Portugal*

Fernando Martins, *Universidade de Lisboa, Portugal*

Karen McAuliffe, *University of Birmingham, UK*

Frances Rock, *Cardiff University, UK*

Paolo Rosso, *Polytechnic University of Valencia, Spain*

Susan Sarcevic, *University of Rijeka, Croatia*

Roger Shuy, *Georgetown University Washington, USA*

Editorial Assistant / Assistente Editorial

Sofia Meneses-Silva, *Universidade do Porto, Portugal*

Copyright / Direitos de autor

The articles published in this volume are covered by the Creative Commons "Attribution-NonCommercial" (CC-BY-NC) license (see <http://creativecommons.org>). They may be reproduced in its entirety as long as Language and Law / Linguagem e Direito is credited, a link to the journal's web page is provided, and no charge is imposed. The articles may not be reproduced in part or altered in form, or if a fee is charged, without the journal's permission. Copyright remains solely with individual authors. The authors should let the journal Language and Law / Linguagem e Direito know if they wish to republish.

Os artigos publicados neste volume estão cobertos pela licença Creative Commons "Attribution-NonCommercial" (CC-BY-NC) (consultar <http://creativecommons.org>) e podem ser reproduzidos na íntegra desde que seja feita a devida atribuição à Language and Law / Linguagem e Direito, com indicação do link para a página da revista e desde que não sejam cobradas quaisquer taxas. Os artigos não podem ser parcialmente reproduzidos, o seu formato não pode ser alterado, e não podem ser cobradas taxas sem a autorização da revista. Os direitos de autor dos trabalhos publicados nesta revista pertencem exclusivamente aos seus respetivos autores. Os autores devem informar a revista Language and Law / Linguagem e Direito se pretenderem submeter o artigo noutra a outra publicação.

Language and Law / Linguagem e Direito

Language and Law / Linguagem e Direito is a free, exclusively online peer-reviewed journal published twice a year. It is available on the website of the Faculty of Arts of the University of Porto, at <http://ojs.letras.up.pt/index.php/LLLD/>.

All articles should be submitted via the system. See the guidelines for submission at the journal website.

Requests for book reviews should be sent to rssilva@letras.up.pt.

Language and Law / Linguagem e Direito é uma revista gratuita publicada exclusivamente online, sujeita a revisão por pares, publicada semestralmente e disponível no website da Faculdade de Letras da Universidade do Porto, em <http://ojs.letras.up.pt/index.php/LLLD/>.

Os materiais para publicação deverão ser enviados através do sistema, e devem seguir as instruções disponíveis na página da revista.

As propostas de recensão de livros devem ser enviadas para rssilva@letras.up.pt.

Indexing and abstracting / Indexação e bases de dados bibliográficas

Language and Law / Linguagem e Direito is covered by the following abstracting and indexing services:

A Language and Law / Linguagem e Direito encontra-se indexada e catalogada nas seguintes bases de dados:

CrossRef

EBSCO Legal Source

ERIH PLUS: European Reference Index for the Humanities and the Social Sciences

Google Scholar

Journals for Free

JURN

Portal RCAAP

QUALIS Periódicos (A2)

Sherpa Romeo

PUBLISHED BIANNUALLY ONLINE / PUBLICAÇÃO SEMESTRAL ONLINE

ISSN: 2183-3745

D.O.I.: <https://doi.org/10.21747/21833745/lanlaw>

D.O.I. Volume 12(1): https://doi.org/10.21747/21833745/lanlaw12_1

THE ARTICLES ARE THE SOLE RESPONSIBILITY OF THEIR AUTHORS.

THE ARTICLES WERE PEER REVIEWED.

OS ARTIGOS SÃO DA EXCLUSIVA RESPONSABILIDADE DOS SEUS AUTORES.

OS ARTIGOS FORAM SUBMETIDOS A ARBITRAGEM CIENTÍFICA.

Contents / Índice

ARTICLES / ARTIGOS

- Introduction / Introdução** 1-9 / 10 - 20
- No language left behind? Towards an integrated framework for linguistic rights, human rights and technology regulation** 21-38
Ingeborg Birnie
- Handling multilingualism by the EU institutions at the time of AI-driven language technologies** 39-60
Aneta Skorupa-Wulczynska and Anna Zannina
- Can machine translation really help minority languages in Europe? An analysis with value scenarios** 61-78
Sergi Alvarez-Vidal and Maarit Koponen
- When machine-generated mistranslation on social media becomes misinformation: Risks to users, corporate responsibility, and legal implications** 79-99
Khetam Al Sharou
- Deaf in AI: AI language technologies and the erosion of linguistic rights** 100-121
Maartje De Meulder
- The complexity of gender and language: Digitization of a physical board game deconstructing gender stereotypes** 122-143
Dimitra Anastasiou, Christian Moll, Marie Gallais, Laurence Johannsen, and Carole Blond-Hanten
- Professor: Who holds the copyright for AI-assisted and AI-generated contents?** 144-161
Katerina Zdravkova

Language, Law and Rights: Balancing AI-Driven Technology and Equity

Introduction

The rise of artificial intelligence (AI) presents a profound challenge to the equitable application of law and the protection of language rights. The technological landscape is undergoing game-changing shifts at an incredible pace. What seems new today will be either old or obsolete in one or two years. This dynamic landscape therefore demands constant monitoring and critical investigation.

This Special Issue, *Language, Law and Rights: Balancing AI-Driven Technology and Equity*, examines the intricate relationship between these areas, exploring how AI-driven language technologies can both empower and marginalize, particularly for speakers of minority and minoritised languages. This collection of articles stems from discussions initiated at the “Voices on Language Rights: Public Perceptions and Attitudes” workshop, which took place at the University of Geneva, February 12-13, 2024¹. This event, organised within the Language in the Human-Machine Era (LITHME) project – funded by the European Cooperation in Science and Technology (COST) – brought together experts from different European countries to explore the impact of new technologies on language rights. The workshop aimed to raise awareness about the potential effects of emerging technologies on both major and minority languages, as well as their speakers, in diverse contexts. Understanding these impacts is essential as we continue to explore the complexities associated with the use of new and emerging technologies on a global scale.

This Special Issue expands upon some of the foundational work presented in the chapter on “Language Rights” (Soltan et al., in press) within the forthcoming LITHME edited volume, *Language in the Human-Machine Era* (MIT Press). While the book provides a broad overview of language in the human-machine era, this special issue scrutinizes specific legal and ethical ramifications of these technological advancements. Central to our inquiry is the question: does AI serve humanity, or does it inadvertently exacerbate existing inequalities? The Special Issue examines the multifaceted impact of AI on language, law, language rights, human rights, and linguistic/social justice, with particular focus on the experiences of those who speak minority or minoritised languages. It aims to provide a critical assessment of these developments from various perspectives, both in terms of their benefits and advantages, as well as their weaknesses, challenges, and points for development.

Numerous indicators suggest that the reflections presented in this volume may be relevant not only to academia but also to society at large, especially during a time when human interaction with technology poses significant challenges (Zanatto, Bifani,

¹<https://lithme.eu/2024/04/17/voices-on-language-rights-public-perceptions-and-attitudes/>

& Noyes, 2023). We are currently navigating a period marked by substantial change and uncertainty, driven by new technologies whose implications remain largely ambiguous. The technological landscape is experiencing transformative shifts that extend beyond commonly used digital applications to encompass those propelled by artificial intelligence. AI is increasingly becoming integrated into all facets of human activity. We find ourselves in an “AI race”, echoing the space race of the 20th century. As a result, we are witnessing profound changes in how we interact with, collaborate via, and co-create using this technology. At its core, AI functions as a tool for processing and generating human-like language, and its influence as well as significance are ultimately shaped by the design and deployment choices made by humans.

When considerations of equity and ethics are neglected in the development of AI, the technology can exacerbate existing social inequalities and further marginalise vulnerable communities. Notably, large language models (LLMs) may perpetuate those inequalities due to biases present in their training data and insufficient representation of diverse languages (see De Meulder, 2021; Krausneker & Schügerl, 2022; Quandt et al., 2022, for a discussion on the ethical and responsible development of such technologies, particularly regarding sign language avatars, and also refer to De Meulder 2025, Birnie 2025, and Skorupa-Wulczyńska & Zannina 2025 in this Special Issue). Moreover, recent technological advancements may deepen inequalities among vulnerable communities, potentially infringing on their rights. Over the past decade, humanitarian actors and affected communities have increasingly incorporated advancements in information and communication technologies (ICTs) and the resulting digital data into their crisis responses. However, these operational and technological advancements are developing without a universally accepted rights-based approach (RBA) for conducting humanitarian information activities in the contemporary context (Greenwood et al., 2017).

LLMs have become the focus of most discussions, as this technology enables the generation of content that previously demanded significant human effort. Additionally, LLMs operationalise language in a manner that empowers non-expert users to access, interact with, and control various other types of pervasive technologies, such as chatbots, avatars, robots, and Internet of Things (IoT) devices. However, there is a critical caveat: LLMs inherit biases from the data on which they are trained, perpetuating unequal representation across languages. These biases stem from both the quality of the data and representation within the training data, as well as the specific languages used in gathering the data.

Against this backdrop, not all languages are treated equally, despite the fact that different international organisations have developed instruments to promote the implementation of language rights principles. While international law provides a foundation for language rights, these protections are rarely at the forefront. Instead, they are typically included within broader standards for minority rights. A clear example is the case of Sign languages in the United Nations’ Convention on the Rights of Persons with Disabilities (UNCRPD 2006), where linguistic rights are acknowledged but positioned within a wider disability-rights agenda rather than recognised as full linguistic rights. Foundational documents like the Declaration on the Rights of Persons Belonging to National or Ethnic, Religious and Linguistic Minorities (UNGA, 1992), and the Council of Europe’s Framework Convention for the Protection of National Minorities (Council of Europe, 1995) establish the right of minorities to use their own language, but

treat it as one component of cultural identity rather than a central, standalone right. The result is a set of principles that are often difficult to enforce directly for linguistic justice alone. While more recent, follow-on tools, such as the UN Practical Guide of Implementation of Language Rights for Linguistic Minorities (2017), or the Council of Europe Guide on Facilitating the implementation of the European Charter for Regional or Minority Languages through artificial intelligence (Council of Europe, 1992; Gerken, 2022, see also Birnie 2025 in this Special Issue) aim to translate these principles into practice, the fundamental challenge remains. Nascent regulations like the European Union's AI Act (European Parliament and Council 2024, see also Skorupa-Wulczyńska 2025 in this Special Issue) are designed to govern technology, but it remains an open question whether they can effectively address the deep-seated linguistic inequalities that older human rights frameworks have struggled to resolve. This is further complicated by the digital divide. As of 2024 (International Telecommunication Union), a third of the global population remains offline, especially in low-income regions and rural areas. The fact is that a considerable number of minority and minoritised languages are at a disadvantage in the AI era. Consequently, many language communities, with their rich variations, do not have equal access to the technological advancements that define our era. This disparity raises significant concerns about equitable participation and opportunities. LLMs exemplify how information technology can influence the very essence of our being, shaping language, culture, identity, and our collective sense of belonging. An increased reliance on technology may also create social pressures to keep pace, which could marginalise individuals lacking access or the necessary skills. This situation highlights the need for equitable access to information and services, ensuring that all communities can contribute to and thrive in society. Special attention must be given to the rights and wellbeing of those in vulnerable situations. In a nutshell, we must consider how technological developments impact languages, cultures, identities, and a sense of belonging within communities.

Through investigating the complex interconnections between language, law, and rights in the age of AI-driven technologies, we hope to contribute to four ongoing discussions. First, we hope to facilitate an exploration of the intricacies of human-machine collaboration by examining how humans and AI co-create meaning, the impacts of AI on human agency, and the ethical considerations surrounding human-machine interaction. Second, the Special Issue seeks to enhance our understanding of AI's implications for equity and justice, particularly regarding access to information and services, the digital divide, and the rights of marginalised language communities, along with the specific challenges faced by vulnerable populations. Third, it contributes to contemporary debates on language policy and planning. Lastly, the Special Issue represents investigations of the broader societal impacts of language technology, including AI's role in shaping cultural identity, its use in public services, and the ethical implications of AI-driven communication.

The six articles carefully selected for this Special Issue of *Language Law / Linguagem e Direito* bring together a diversity of reflections on the central theme, "Balancing AI-Driven Technology and Equity", in dialogue with the fields of law and language. The articles address the theme in various contexts and social practices, ranging from institutional legal frameworks to protect smaller state, regional, minority, and indigenous languages to the issue of mistranslation, the impact of technology on deaf communities'

rights, learning minority languages, perpetuation of gender stereotypes, and intellectual property. The Special Issue leaves plenty of room for discussion around AI technology and its impact on language, law, and rights, demonstrating, for example, the challenges that societies face nowadays to make the most of AI technologies, while preserving the rights of their speakers.

The issue opens with an article by Inge Birnie entitled “No language left behind? Towards an integrated framework for linguistic rights, human rights and technology regulation”. The article highlights the fact that access to technology is not universal, examining the various institutional legal frameworks that regulate AI at multiple levels. It particularly emphasises the shortcomings of these frameworks in addressing the needs of minority and minoritised languages, which are often at a disadvantage compared to dominant languages like English regarding access to online spaces. This situation contributes to digital imperialism and colonialism, adversely affecting smaller state, regional, minority, and indigenous languages that typically have fewer speakers. Birnie additionally highlights the challenges faced by languages not traditionally classified as ‘at risk’ (yet, non-English), while discussing the implications of excluding most languages from digital and online spaces, as well as from the regulatory frameworks designed to support and protect them. The article also offers potential avenues for reflection on those issues.

The article “Handling multilingualism by the EU institutions at the time of AI-based language technologies”, by Aneta Skorupa-Wulczynska and Anna Zannina, explores multilingualism as a fundamental principle of language policy within the EU and examines the various challenges brought about by translation as well as technological advances. With the advent of AI and innovative tools such as translation memory systems, neural machine translation, and speech-to-text technologies, new opportunities have emerged alongside significant challenges and risks. The authors investigate whether these technologies improve citizens’ access to EU institutions and subsequently enhance their engagement in democratic processes by reviewing a variety of tools used by EU institutions. They highlight the risks posed by AI language technologies, while recognising their potential to promote equality in the context of the EU’s growing multilingualism. The article suggests mitigation measures, emphasising that such technologies should be operated by humans who adhere to ethical, moral, and legal standards.

Continuing the theme of minority languages, Sergi Alvarez-Vidal and Maarit Koponen, in “Can machine translation really help minority languages in Europe? An analysis with value scenarios” reflect on the advances of machine translation (MT) across various sectors of society, thanks to the advent of neural MT (NMT) and more recently, LLMs. However, the authors argue that many MT models remain English-centric, delivering high-quality results predominantly for languages with abundant data, while for minority languages, the issue is often perceived as one of data scarcity. The authors invite us to look beyond the technical narrative of “data scarcity” to the deeper systemic realities that shape language technologies. Drawing on value scenarios for Catalan and Karelian – two languages with contrasting sociolinguistic trajectories – they explore how the deployment of MT reinforce or disrupt existing power relations between languages. The article delineates the primary challenges and possible drawbacks associated with MT for minority languages, and proposes general guidelines to inform future research and applications. This work foregrounds the need for design approaches that integrate

sociolinguistic diversity, ethical reflection, and participatory input from affected communities. It exemplifies the broader aim of this volume: to reframe AI not as a tool for efficiency, but as a field of responsibility.

In the article “When machine-generated mistranslation on social media becomes misinformation: Risks to users, corporate responsibility, and legal implications”, Khetam Al Sharou focuses on the impact of machine translation (MT) on misinformation, by drawing on real-life examples of how MT can distort content and disproportionately affect marginalised language communities that rely on it for information. The unchecked proliferation of MT on social media platforms presents a critical threat, where algorithmic errors can rapidly evolve into harmful misinformation. Al Sharou demonstrates that as users increasingly rely on MT to access information, they often lack the source-language knowledge to detect subtle but significant inaccuracies, leading to distorted meanings. The article emphasises the ethical responsibility of tech companies to address these risks and explores how regulatory frameworks like the EU’s Digital Services Act and the UK Online Safety Act can protect language rights online. Furthermore, the article advocates for increased accountability among service providers and the responsible use of MT services to ensure that accurate information is accessible. This argument reframes the issue as a matter of language rights, the right to access accurate information without being misled by flawed AI.

A discussion of the complex relationship between AI language technologies, sign language interpreting, and linguistic access is the objective of “Deaf in AI: AI language technologies and the erosion of linguistic rights”, by Maartje De Meulder. She emphasises how these advancements pose a risk to the hard-won linguistic rights of deaf communities. While AI tools promise innovation and resilience, they also perpetuate biases, reinforce technoableism, and exacerbate inequalities due to systemic and design flaws. By drawing on insights from Deaf Studies, Sign Language Interpreting Studies, and crip technoscience, the author critiques the perception of AI as a replacement for interpreters, underscoring its broader ramifications for access frameworks. The article advocates for deaf-led approaches to ensure that AI promotes equitable, ethical, and trustworthy accessibility practices that protect the linguistic and social rights of deaf communities. The author concludes with the powerful message that it is crucial to engage critically with technological changes as they emerge, in order to cultivate AI systems that are not only innovative but also equitable, inclusive, and trustworthy. Such systems should enhance linguistic practices rather than subordinate them to technological demands.

Stereotypes and terminology go hand-in-hand with language, and how the technological world is socially shaped. LLMs risk amplifying societal biases, particularly those rooted in gender stereotypes, posing a significant threat to equity in our automated world. Dimitra Anastasiou, Christian Moll, Marie Gallais, Laurence Johannsen, and Carole Blond-Hanten’s article entitled “The complexity of gender and language: Digitization of a physical board game deconstructing gender stereotypes”, a blended digital-physical tool, as a human-centered intervention to counteract this trend. Rooted in the social sciences, the GG is designed to deconstruct these stereotypes by engaging the public with scientific research on gender equality. A core component of the game directly confronts gendered language, serving as a practical tool for promoting linguistic rights. To test the prevalence of these biases within AI itself, the authors conducted an exploratory study using generative AI. The paper discusses these findings, revealing how current

LLMs navigate complex issues of gender and language. By contrasting AI-generated content with the human-led discussions that the game facilitates, the study underscores the importance of proactive educational tools. This positions the GG as a necessary intervention designed to address the root cause of the biases that AI systems inherit. Ultimately, the paper argues for the critical role of such human-centered tools in fostering a more equitable foundation for our digital future.

The rise of generative AI has created a fundamental conflict at the heart of intellectual property law, challenging our very definitions of authorship and creation. This legal dilemma operates on two fronts: the contentious use of copyrighted human expression for model training, and the unresolved question of who owns the rights to AI-generated content. In her article, “Professor: Who holds the copyright for AI-assisted and AI-generated contents?”, Katerina Zdravkova confronts these issues, providing a critical legal analysis of this new technological landscape. Through a detailed examination of nine case studies, she investigates the crucial distinction between AI-assisted and fully AI-generated works. Her findings reveal an emerging legal consensus where human creativity remains the cornerstone of copyright. The author argues that works created with AI as a tool can be granted protection, as the human author’s intellectual input is central. In contrast, she demonstrates that fully autonomous works generated without direct human creativity cannot currently be granted the same intellectual property rights. Zdravkova’s analysis provides an essential framework for navigating the rights of human creators in a world increasingly populated by automated content.

The comprehensive examination of the articles prompted us to further reflect on the white paper that emerged as a tangible outcome of the previously mentioned workshop. At the conclusion of the workshop, participants considered various principles relevant to different stakeholders, as a collective reflection on the ethical and practical principles needed to guide AI and language governance. Participants agreed that real progress depends on shared responsibility among technology developers, researchers, policy makers, and affected communities.

For technology developers, this means going beyond regulatory compliance to design systems that actively protect human and linguistic rights. Profitability and public value can coexist when innovation begins with an understanding of who is affected, how access is distributed, and what new inequalities might emerge. Development should be grounded in genuine human needs, with continuous input from the communities that the technology seeks to serve. This involves identifying affected communities from the outset, understanding their linguistic realities, assessing inequalities of access, and anticipating new disparities that emerging technologies might introduce. To that end, inclusive design requires attention to embedded biases, related to gender, age, ethnicity, political viewpoint, and other identity markers, and the proactive engagement of communities throughout the design and evaluation process. This participatory approach must be supported by education and mutual learning: developers should not only teach users how to engage with new technologies but also learn continuously from their evolving experiences and expectations.

For their part, researchers have several potential avenues to explore, regarding equity and equality. Key considerations include issues of positionality and reflexivity, as well as the importance of self-knowledge in understanding the actual needs of individuals. Emphasising these aspects fosters epistemic humility (Medina, 2013, p. 43, see also

Peled, 2018) – recognising what we do not know and beginning the process of discovery – and encourages the adoption of a critical lens that strengthens our commitment to diversity, not only in languages but also in language varieties. Additionally, cultivating meta-linguistic awareness is vital, allowing for higher-order reflection and helping to prevent modality chauvinism. This means avoiding the treatment of languages merely as systems of rules and protocols without a true understanding of their functioning. Research topics could also encompass interactivity and intercultural communication. Meaningful progress also depends on interdisciplinary collaboration that brings together researchers, developers, and communities. Such dialogue not only strengthens our collective capacity to face emerging challenges but also raises a deeper question of legitimacy – who authorises and benefits from research on language and technology. Upholding the principle of “*nothing about us, without us*” ensures that ethical reflection and moral responsibility take precedence over procedural compliance.

Policy makers, in turn, play a central role in creating conditions for responsible innovation: fostering transparency, supporting linguistic diversity, embedding fairness audits in regulation, and ensuring the financial sustainability of tools developed for under-resourced languages. Additionally, establishing enforcement mechanisms is vital to hold developers and technological businesses, and even owners, accountable for upholding standards of equity and linguistic justice. It is crucial for policymakers to sustain an ongoing dialogue with technology developers, communities, and researchers. Embedding this co-creation approach within the legal framework that encompasses human and language rights is the foundation for the commitment to these principles. Finally, it is essential to remember that technology is not always the best, or only, solution. Every product must be measured against its human alternative: does it genuinely enhance linguistic rights, equity, and agency, or does it simply automate what should remain a human act of connection? Communities have a defining role in shaping how language technologies evolve. Their participation is key not only in expressing needs and aspirations but also in ensuring these perspectives are reflected – both technically and legally – throughout the development cycle, from data collection to deployment. By engaging in the assessment of existing tools and the analysis of policy frameworks, communities can help reveal where access remains uneven and how language technologies might better serve equity and inclusion.

We hope this volume encourages further inquiry into the themes explored here and into other domains where AI increasingly shapes linguistic rights and practices. Continued reflection and collaborative research can help surface new challenges, deepen understanding, and contribute fresh perspectives to the broader conversation on how emerging technologies reshape language, equity, and society.

Guest Editors

Angela Soltan

Rebekah Rousi

Lucía Ruiz Rosendo

References

- Council of Europe. (1992, November). *European Charter for Regional or Minority Languages* (Tech. Rep.). Retrieved from <https://www.refworld.org/legal/agreements/coe/1992/en/41001>
- Council of Europe. (1995, February). *Framework Convention for the Protection of National Minorities. ETS 157*. Retrieved from <https://rm.coe.int/16800c10cf>
- De Meulder, M. (2021). Is “good enough” dood enough? Ethical and responsible development of sign language technologies. In *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*.
- European Parliament and Council. (2024, July). *Artificial Intelligence Act. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June. Official Journal of the European Union, L 1689*. Retrieved from <https://artificialintelligenceact.eu/>.
- Gerken, M. (2022). *Facilitating the implementation of the European Charter for Regional and Minority Languages through artificial intelligence*. Council of Europe.
- Greenwood, F., Howarth, C., Poole, D. E., Raymond, N. A., & Scarnecchia, D. P. (2017). “The Signal Code: A Human Rights Approach to Information During Crisis.”. *Harvard Humanitarian Initiative*. Retrieved from <https://hhi.harvard.edu/publications/signal-code-human-rights-approach-information-during-crisis>
- International Telecommunication Union. (2024). *Facts and Figures*. Retrieved from <https://www.itu.int/en/mediacentre/Pages/PR-2024-11-27-facts-and-figures.aspx>
- Krausneker, V., & Schügerl, S. (2022). Avatars for sign languages: Best practice from the perspective of deaf users. *ICCHP-AAATE 2022 Open Access Compendium "Assistive Technology, Accessibility and (e)Inclusion, Part I*. Retrieved from <https://epub.jku.at/obvulioa/content/titleinfo/7945395> doi: 10.35011/ICCHP-AAATE22-P1-21
- Medina, J. (2013). Active Ignorance, Epistemic Others, and Epistemic Friction. In *The Epistemology of Resistance* (1st ed., pp. 27–55). Oxford University Press. Retrieved 2025-12-06, from <https://academic.oup.com/book/9202/chapter/155839040> doi: 10.1093/acprof:oso/9780199929023.003.0001
- Peled, Y. (2018). Language barriers and epistemic injustice in healthcare settings. *Bioethics*, 32(6), 360–367.
- Quandt, L., Lamberton, J., Leannah, C., Willis, A., & Malzkuhn, M. (2022, June). Signing Avatars in a New Dimension: Challenges and Opportunities in Virtual Reality. In E. Efthimiou, S.-E. Fotinea, T. Hanke, J. C. McDonald, D. Shterionov, & R. Wolfe (Eds.), *Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology: The Junction of the Visual and the Textual: Challenges and Perspectives* (pp. 85–90). Marseille, France: European Language Resources Association. Retrieved from <https://aclanthology.org/2022.sltat-1.13/>
- Soltan, A., Rousi, R., Prys, C., O’Brien, S., Ruiz Rosendo, L., & Rainò, P. (in press). Language rights in the human-machine era: A multifaceted inquiry. In D. Sayers, M. Glass, H. Kelly-Holmes, & R. Fuchs (Eds.), *Language in the Human-Machine Era: New technologies and the coming transformation of language*. Boston, MA: MIT University Press.
- United Nations. (1992, December). *Declaration on the Rights of Persons Belonging to National or Ethnic, Religious and Linguistic Minorities. General Assembly Resolution 47/135*. Retrieved from <https://www.ohchr.org/en/instruments-mechanisms/instruments/declaration-rights-persons-belonging-national-or-ethnic>.

- United Nations. (2006, December). *Convention on the Rights of Persons with Disabilities. UN Treaty Series, 2515*. Retrieved from <https://social.desa.un.org/issues/disability/crpd/convention-on-the-rights-of-persons-with-disabilities-crpd>.
- United Nations. (2017). *Language Rights of Linguistic Minorities: A Practical Guide for Implementation. Special Rapporteur on minority issues: Geneva*. Retrieved from http://www.ohchr.org/Documents/Issues/Minorities/SR/LanguageRightsLinguisticMinorities_EN.pdf.
- Zanatto, D., Bifani, S., & Noyes, J. (2023). Constraining the Sense of Agency in Human-Machine Interaction. *International Journal of Human-Computer Interaction, 40*(13), 3482–3493. Retrieved from <https://www.tandfonline.com/doi/full/10.1080/10447318.2023.2189815> doi: 10.1080/10447318.2023.2189815

Language, Law and Rights: Balancing AI-Driven Technology and Equity

Introdução

O desenvolvimento da inteligência artificial (IA) representa um desafio profundo para a aplicação equitativa do direito e para a proteção dos direitos linguísticos. O panorama tecnológico está a sofrer transformações disruptivas a um ritmo incrível. Aquilo que hoje parece novo será considerado antigo ou obsoleto dentro de um ou dois anos, pelo que este cenário dinâmico exige continuamente uma monitorização e uma investigação crítica.

Este número especial, sob a temática *Language, Law and Rights: Balancing AI-Driven Technology and Equity*, aborda a intrincada relação entre estas áreas, investigando de que modo as tecnologias da linguagem baseadas em IA contribuem, em simultâneo, para reforçar e para marginalizar, sobretudo no caso de falantes de línguas minoritárias e minorizadas. Esta coletânea de artigos resulta dos debates iniciados no workshop “Voices on Language Rights: Public Perceptions and Attitudes”, realizado na Universidade de Genebra, na Suíça, nos dias 12 e 13 de fevereiro de 2024¹. Este evento, organizado no âmbito do projeto Language in the Human-Machine Era (LITHME), financiado pela European Cooperation in Science and Technology (COST), reuniu especialistas de diversos países europeus para explorar o impacto das novas tecnologias nos direitos linguísticos. O workshop teve como objetivo sensibilizar a comunidade para os potenciais efeitos das tecnologias emergentes, quer sobre as línguas maioritárias, quer sobre as línguas minoritárias, bem como sobre os seus falantes, em contextos diversos. A compreensão destes impactos é essencial numa era em que continuamos a explorar as complexidades associadas à utilização de tecnologias inovadoras e emergentes à escala global.

Este número especial desenvolve algum do trabalho seminal apresentado no capítulo sobre “Language Rights” (Soltan et al., in press), incluído na publicação LITHME, *Language in the Human-Machine Era* (MIT Press). Porém, enquanto o livro oferece uma panorâmica abrangente da linguagem na era da interação entre o ser humano e a máquina, este número especial analisa minuciosamente determinadas implicações éticas e legais destes avanços tecnológicos. No cerne da nossa reflexão encontra-se a seguinte questão: servirá a IA genuinamente a humanidade, ou, pelo contrário, contribuirá, inadvertidamente, para exacerbar as desigualdades existentes?

¹<https://lithme.eu/2024/04/17/voices-on-language-rights-public-perceptions-and-attitudes/>

Este número investiga o impacto multifacetado da IA sobre a linguagem, o direito, os direitos linguísticos, os direitos humanos e a justiça linguística/social, com especial enfoque nas experiências de falantes de línguas minoritárias ou minorizadas, procurando oferecer uma avaliação crítica destes desenvolvimentos a partir de diferentes perspetivas, tanto no que se refere aos seus benefícios e às suas vantagens, como no respeitante às suas fragilidades, desafios e elementos de melhoria.

Existem diversos indicadores que sugerem que as reflexões apresentadas neste volume podem ser relevantes, não só para o contexto académico, mas também para a sociedade em geral, sobretudo num período em que a interação humana com a tecnologia suscita desafios significativos (Zanatto, Bifani, & Noyes, 2023). Vivemos atualmente num período marcado por mudanças profundas e por incerteza, impulsionado por novas tecnologias cujas implicações permanecem largamente ambíguas. O panorama tecnológico está a sofrer mudanças transformadoras que vão além das aplicações digitais comuns, abrangendo tecnologias apoiadas por inteligência artificial. A IA encontra-se cada vez mais integrada em todas as dimensões da atividade humana. Encontramo-nos numa “corrida à IA” que espelha a corrida espacial do século XX. Consequentemente, testemunhamos mudanças profundas na forma como interagimos, colaboramos e cocriamos com recurso a esta tecnologia. Essencialmente, a IA funciona como uma ferramenta para processar e gerar linguagem semelhante à humana, com uma influência e relevância moldadas, em última instância, pelas escolhas humanas em matéria de conceção e de implementação.

Ao descurar as considerações de equidade e de ética no desenvolvimento da IA, a tecnologia poderá aprofundar desigualdades sociais já existentes e marginalizar ainda mais comunidades vulneráveis. Claramente, os grandes modelos de linguagem (LLM) poderão perpetuar essas desigualdades devido a enviesamentos presentes nos seus dados de treino e à representação insuficiente da diversidade linguística (ver De Meulder, 2021; Krausneker & Schügerl, 2022 e Quandt et al., 2022 para uma discussão sobre o desenvolvimento ético e responsável dessas tecnologias, sobretudo relativamente a avatares de línguas de sinais, bem como De Meulder 2025, Birnie 2025, e Skorupa-Wulczyńska & Zannina 2025, neste Número Especial). Além disso, os avanços tecnológicos mais recentes poderão aprofundar desigualdades entre comunidades vulneráveis, possivelmente violando os seus direitos. Ao longo da última década, agentes humanitários e comunidades afetadas por estes desenvolvimentos têm recorrido cada vez mais às tecnologias de informação e da comunicação (TIC) e aos dados digitais delas resultantes para dar resposta a sucessivas crises. Todavia, estes progressos operacionais e tecnológicos têm decorrido sem uma abordagem assente em direitos (RBA) aceite universalmente para implementar atividades de informação humanitária no contexto atual (Greenwood et al., 2017).

Os LLM tornaram-se o foco da maioria das discussões, uma vez que essa tecnologia permite a geração de conteúdo que, anteriormente, exigia um esforço humano significativo. Além disso, os LLM operacionalizam a linguagem de uma forma que capacita os utilizadores não especializados para aceder, interagir e controlar diversos outros tipos de tecnologias comuns, como chatbots, avatares, robôs e dispositivos da Internet das Coisas (IoT). Contudo, existe uma ressalva importante a fazer: os LLM herdaram preconceitos dos dados com os quais são treinados, perpetuando uma representação desigual das diferentes línguas. Esses preconceitos decorrem, quer da

qualidade dos dados e da representação nos dados de treino, quer das línguas específicas selecionadas na recolha desses dados.

Neste cenário, nem todas as línguas são tratadas de forma igual, apesar de diferentes organizações internacionais terem desenvolvido instrumentos destinados a promover a implementação de princípios de direitos linguísticos. Embora o direito internacional ofereça um quadro de referência para a proteção dos direitos linguísticos, estas garantias raramente assumem um carácter prioritário, antes pelo contrário; normalmente, surgem inseridas em normas mais amplas sobre direitos das minorias. Exemplo claro desta prática é o tratamento dado às línguas gestuais na Convenção das Nações Unidas sobre os Direitos das Pessoas com Deficiência (UNCRPD 2006), onde os direitos linguísticos são reconhecidos, mas enquadrados numa agenda mais vasta de direitos das pessoas com deficiência, e não como direitos linguísticos plenos. Documentos fundamentais, como a Declaração sobre os Direitos das Pessoas Pertencentes a Minorias Nacionais ou Étnicas, Religiosas e Linguísticas (UNGA, 1992) e a Convenção-Quadro para a Proteção das Minorias Nacionais do Conselho da Europa (Council of Europe, 1995), determinam o direito das minorias a utilizar a sua própria língua, mas tratam-no sobretudo como um elemento de identidade cultural, e não como um direito autónomo e central. O resultado é um conjunto de princípios normalmente difíceis de aplicar diretamente para assegurar justiça linguística.

Embora ferramentas mais recentes, como o Guia Prático das Nações Unidas para a Implementação dos Direitos Linguísticos das Minorias Linguísticas (2017) ou o Guia do Conselho da Europa sobre a facilitação da implementação da Carta Europeia para as Línguas Regionais ou Minoritárias através da inteligência artificial (Gerken, 2022, Council of Europe, 1992; ver também Birnie 2025 neste Número Especial), procurem transpor esses princípios para práticas concretas, o desafio de fundo mantém-se. Alguns mecanismos de regulamentação emergentes, como a legislação de IA da União Europeia (AI Act) (Parlamento Europeu e Conselho 2024, ver também Skorupa-Wulczyńska 2025 neste Número), foram concebidas para governar a tecnologia, mas as dúvidas quanto à sua eficácia para enfrentar as profundas desigualdades linguísticas que os enquadramentos de direitos humanos mais antigos não conseguiram resolver permanecem. Esta questão é aprofundada ainda mais pela continuação do fosso digital. Desde 2024 (International Telecommunication Union, ITU), cerca de um terço da população mundial permaneceu desconectada, sobretudo em regiões de baixo rendimento e em áreas rurais. O facto é que muitas línguas minoritárias e minorizadas se encontram numa posição desfavorecida na era da IA. Consequentemente, diversas comunidades linguísticas, com as suas ricas variações, não têm acesso equitativo às inovações tecnológicas que definem a nossa era. Esta disparidade levanta sérias preocupações quanto à participação equitativa e às oportunidades à disposição dos cidadãos. Os LLM são um exemplo de como a tecnologia da informação pode influenciar a verdadeira essência do nosso ser, moldando a linguagem, a cultura, a identidade e o nosso sentido coletivo de pertença. Uma dependência crescente da tecnologia pode também criar pressões sociais para acompanhar o ritmo destas mudanças, marginalizando quem não tem acesso ou competências necessárias, o que evidencia a necessidade de garantir um acesso equitativo à informação e aos serviços, assegurando que todas as comunidades possam contribuir e prosperar na sociedade. É essencial prestar especial atenção aos direitos e ao bem-estar das pessoas em situação

de vulnerabilidade. Em suma, é necessário ponderar de que forma os desenvolvimentos tecnológicos afetam línguas, culturas, identidades e o sentimento de pertença nas comunidades.

Ao investigar as complexas interconexões entre linguagem, direito e direitos na era das tecnologias de IA, esperamos contribuir para quatro debates em curso. Em primeiro lugar, procuramos fomentar o estudo das dinâmicas da colaboração entre o ser humano e a máquina, analisando de que modo seres humanos e sistemas de IA cocriam significados, os impactos da IA na agência humana e as questões éticas associadas à interação entre seres humanos e máquinas. Em segundo lugar, este Número Especial procura aprofundar a compreensão das implicações da IA para a equidade e para a justiça, em particular no que respeita o acesso a informação e a serviços, o fosso digital e os direitos das comunidades linguísticas marginalizadas, bem como os desafios específicos com que se confrontam populações vulneráveis. Em terceiro lugar, procuramos contribuir para o debate contemporâneo sobre políticas linguísticas e planeamento linguístico. Finalmente, este Número Especial inclui estudos sobre o impacto social mais amplo da tecnologia linguística, incluindo o papel da IA na configuração da identidade cultural, a sua utilização nos serviços públicos e as implicações éticas da comunicação mediada por IA.

Os seis artigos criteriosamente selecionados para este número especial da *Language Law / Linguagem e Direito* reúnem uma diversidade de reflexões sobre o tema central, “Balancing AI-Driven Technology and Equity”, em diálogo com os campos da linguagem e do direito. Os artigos abordam esta temática sob diferentes contextos e práticas sociais, abrangendo desde enquadramentos jurídicos institucionais para a proteção de línguas de Estados de menores dimensões, línguas regionais, minoritárias e indígenas, até questões de erros de tradução, impacto da tecnologia nos direitos de comunidades surdas, aprendizagem de línguas minoritárias, perpetuação de estereótipos de género e propriedade intelectual. Este Número Especial deixa um espaço amplo para a discussão em torno da tecnologia de IA e do seu impacto sobre a linguagem, o direito e os direitos, evidenciando, por exemplo, os desafios que as sociedades enfrentam atualmente para aproveitar ao máximo as tecnologias de IA, preservando simultaneamente os direitos dos seus falantes.

O número abre com um artigo de Inge Birnie, intitulado “No language left behind? Towards an integrated framework for linguistic rights, human rights and technology regulation”. O artigo destaca o facto de o acesso à tecnologia não ser universal, analisando diversos enquadramentos jurídicos institucionais que regulam a IA em múltiplos níveis. O artigo sublinha, em particular, as limitações desses enquadramentos no que se refere à capacidade de responder às necessidades das línguas minoritárias e minorizadas, frequentemente em situação de desvantagem comparativamente a línguas dominantes como o inglês no que concerne o acesso ao espaço digital. Este contexto contribui para formas de imperialismo e colonialismo digital, com impacto negativo sobre línguas de Estados de menores dimensões, línguas regionais, minoritárias e indígenas, que, normalmente, possuem menos falantes. Birnie realça ainda os desafios enfrentados por línguas que não são tradicionalmente classificadas como estando “em risco” (mas que não são o inglês), discutindo as implicações da exclusão da maioria das línguas dos espaços digitais e *online*, bem como dos enquadramentos regulatórios

destinados a apoiá-los e a protegê-los. Adicionalmente, o artigo oferece possíveis linhas para reflexão sobre estas questões.

O artigo “Handling multilingualism by the EU institutions at the time of AI-based language technologies”, da autoria de Aneta Skorupa-Wulczynska e Anna Zannina, explora o multilinguismo como princípio fundamental da política linguística da União Europeia e analisa os vários desafios introduzidos, tanto pela tradução, como pelos desenvolvimentos tecnológicos. Com o advento da IA e de ferramentas inovadoras, como sistemas de memória de tradução, tradução automática neuronal e tecnologias de reconhecimento de fala, surgiram novas oportunidades, acompanhadas de desafios e riscos significativos. As autoras investigam de que forma estas tecnologias melhoram o acesso dos cidadãos às instituições da UE e, conseqüentemente, reforçam o seu envolvimento nos processos democráticos, ao examinar um conjunto de ferramentas utilizadas por essas instituições. O artigo destaca os riscos associados às tecnologias da linguagem assentes em IA, sem deixar de reconhecer o seu potencial para promover a igualdade no contexto do crescente multilinguismo da UE. O artigo sugere medidas de mitigação, sublinhando que essas tecnologias devem ser operadas por humanos que observem padrões éticos, morais e legais.

Prosseguindo com a temática das línguas minoritárias, Sergi Alvarez-Vidal e Maarit Koponen, em “Can machine translation really help minority languages in Europe? An analysis with value scenarios”, refletem sobre os avanços da tradução automática (TA) e a sua ubiquidade em diversos setores da sociedade, impulsionada pela tradução automática neuronal (NMT) e, mais recentemente, pelos LLM. Os autores defendem, no entanto, que muitos modelos de TA permanecem centrados no inglês, obtendo resultados de elevada qualidade sobretudo para línguas com grandes volumes de dados disponíveis, enquanto, no caso de línguas minoritárias, o problema é frequentemente descrito como uma questão de escassez de dados. Os autores convidam-nos a ir além da narrativa técnica da “escassez de dados” e a considerar realidades sistémicas mais profundas que moldam as tecnologias da linguagem. Com base na análise de cenários para o catalão e para o carélio (duas línguas com trajetórias sociolinguísticas contrastantes), o artigo explora de que modo a implementação da TA permite reforçar ou minar relações de poder existentes entre línguas. O artigo elenca os principais desafios e possíveis desvantagens associados à TA para línguas minoritárias e propõe orientações genéricas para informar trabalhos de investigação e aplicações futuras. Este trabalho destaca a necessidade de abordagens de conceção que integrem diversidade sociolinguística, reflexão ética e contributos participativos das comunidades afetadas, exemplificando o objetivo mais amplo deste número: reconfigurar a IA, não como mera ferramenta de eficiência, mas como campo de responsabilidade.

No artigo “When machine-generated mistranslation on social media becomes misinformation: Risks to users, corporate responsibility, and legal implications”, Khetam Al Sharou centra-se no impacto da tradução automática sobre a desinformação, recorrendo a exemplos reais de como a TA permite distorcer conteúdos e afetar desproporcionalmente comunidades linguísticas marginalizadas que dela dependem para aceder à informação. A proliferação descontrolada de TA nas plataformas de redes sociais constitui uma ameaça crítica, em que erros algorítmicos poderão transformar-se rapidamente em desinformação danosa. Al Sharou mostra que, embora os utilizadores dependam cada vez mais da TA para aceder a informação, estes não possuem muitas

vezes os conhecimentos da língua de partida que lhes permitam detetar imprecisões subtis, mas significativas, que conduzem a distorções de sentido. O artigo realça a responsabilidade ética das empresas tecnológicas na mitigação destes riscos e explora de que modo quadros regulatórios como o Digital Services Act da UE ou o Online Safety Act do Reino Unido podem proteger direitos linguísticos *online*. Além disso, defende uma responsabilização acrescida dos prestadores de serviços e o uso responsável de serviços de TA, de forma a assegurar o acesso a informação fidedigna. Este argumento reformula a questão como um problema de direitos linguísticos: o direito a aceder a informação rigorosa sem indução em erro por sistemas de IA inadequados.

A discussão da complexa relação entre tecnologias da linguagem baseadas em IA, interpretação de língua gestual e acesso linguístico está no centro de “Deaf in AI: AI Language technologies and the erosion of linguistic rights”, de Maartje De Meulder. A autora destaca que estes desenvolvimentos colocam em risco direitos linguísticos arduamente conquistados por comunidades surdas. Embora as ferramentas de IA prometam inovação e resiliência, também perpetuam enviesamentos, reforçam o “techoleísmo” e aprofundam desigualdades devido a falhas sistémicas e de conceção. Com base em contributos de estudos de surdez, estudos de interpretação de línguas de sinais e da *crip technoscience*, a autora critica a perceção da IA como substituto de intérpretes, sublinhando as implicações mais vastas para os regimes de acessibilidade. O artigo preconiza abordagens lideradas por pessoas surdas como forma de garantir que a IA promove práticas de acessibilidade equitativas, éticas e fiáveis, que protegem os direitos linguísticos e sociais das comunidades surdas. A autora termina com uma mensagem forte: o compromisso crítico com as mudanças tecnológicas à medida que estas emergem é essencial, de forma a promover sistemas de IA que sejam, não só inovadores, mas também equitativos, inclusivos e fiáveis, que reforcem práticas linguísticas em vez de as subordinarem às exigências tecnológicas.

Os estereótipos e a terminologia estão intimamente ligados à linguagem e à forma como o mundo tecnocultural é socialmente configurado. Os LLM correm o risco de amplificar enviesamentos sociais, sobretudo aqueles que se encontram enraizados em estereótipos de género, e que representam uma ameaça significativa à equidade no nosso mundo automatizado. O artigo de Dimitra Anastasiou, Christian Moll, Marie Gallais, Laurence Johannsen e Carole Blond-Hanten, intitulado “The complexity of gender and language: Digitization of a physical board game deconstructing gender stereotypes”, discute uma ferramenta híbrida digital-física centrada no ser humano como intervenção destinada a inverter esta tendência. Baseado nas ciências sociais, o GG foi concebido para desconstruir esses estereótipos, incentivando o público a interagir com investigação científica sobre igualdade de género. Um componente central do jogo confronta o público diretamente com linguagem genderizada, funcionando como instrumento prático para promover direitos linguísticos. Para testar a prevalência desses enviesamentos na própria IA, as autoras realizaram um estudo exploratório com recurso a IA generativa. O artigo discute os resultados, revelando como os LLM atuais lidam com questões complexas relativas a género e linguagem. Ao contrastar o conteúdo gerado por IA com as discussões realizadas por humanos no contexto do jogo, o estudo sublinha a importância de ferramentas educativas proativas, posicionando o GG como intervenção necessária para enfrentar as causas na génese dos enviesamentos herdados pelos sistemas de IA. Finalmente, o artigo defende o papel crucial dessas ferramentas

centradas no ser humano para a construção de bases mais equitativas para o futuro digital.

A ascensão da IA generativa criou um conflito fundamental no cerne do direito da propriedade intelectual, questionando as próprias definições de autoria e de criação. Este dilema jurídico opera em duas frentes: o uso controverso de obras protegidas por direitos de autor para treino de modelos e a questão, ainda em aberto, de quem detém os direitos sobre conteúdos gerados por IA. No artigo “Professor: Who holds the copyright for AI-assisted and AI-generated contents?”, Katerina Zdravkova discute estas questões, fornecendo uma análise jurídica crítica deste novo cenário tecnológico. Através de uma análise detalhada de nove estudos de caso, a autora investiga a distinção fundamental entre obras de produção assistida por IA e obras totalmente geradas por IA. As suas conclusões apontam para a emergência de um consenso jurídico segundo o qual a criatividade humana permanece o alicerce da proteção de direitos de autor. A autora sustenta que obras criadas com recurso à IA como ferramenta podem ser protegidas, desde que o contributo intelectual da pessoa autora seja determinante, e que, em contrapartida, obras inteiramente autónomas, geradas sem criatividade humana direta, não podem atualmente beneficiar da mesma proteção de propriedade intelectual. A análise de Zdravkova oferece um enquadramento essencial para compreender os direitos de criadores humanos num mundo cada vez mais povoado por conteúdos automatizados.

A leitura atenta destes artigos levou-nos a refletir mais aprofundadamente sobre o livro branco que emergiu como resultado tangível do workshop referido anteriormente. No final do workshop, os membros participantes consideraram vários princípios pertinentes para diferentes partes interessadas, como a reflexão coletiva sobre os princípios éticos e práticos necessários para orientar a governança da IA e da linguagem. Os membros participantes concordaram que o progresso real depende de uma responsabilidade partilhada entre responsáveis pelo desenvolvimento de tecnologia, investigadores, decisores políticos e comunidades implicadas.

Para os responsáveis pelo desenvolvimento de tecnologia, isso implica ir além do mero cumprimento regulamentar, concebendo sistemas que protejam ativamente direitos humanos e direitos linguísticos. Rentabilidade e valor público podem coexistir quando a inovação parte de uma compreensão clara de quem é afetado, de como o acesso é distribuído e de que novas desigualdades podem emergir. O desenvolvimento deverá estar ancorado em necessidades humanas genuínas, com contributos contínuos das comunidades que a tecnologia procura servir, o que exige a identificação precoce das comunidades afetadas, a compreensão das suas realidades linguísticas, a avaliação das desigualdades de acesso e a antecipação de novas disparidades que as tecnologias emergentes possam introduzir. Para o efeito, uma conceção inclusiva requer atenção aos enviesamentos incorporados, relacionados com género, idade, etnia, posicionamento político e outros marcadores identitários, bem como o envolvimento proativo das comunidades ao longo de todo o processo de conceção e avaliação. Esta abordagem participativa deve ser acompanhada por processos de educação e aprendizagem mútua, segundo os quais os responsáveis pelo desenvolvimento não devem apenas ensinar os utilizadores a utilizar novas tecnologias, mas também aprender, de forma contínua, com as experiências e expectativas, em constante mutação, desses utilizadores.

Os investigadores dispõem, por seu lado, de múltiplas linhas para explorar, especialmente no que concerne a equidade e igualdade. Entre os aspetos essenciais, incluem-se as questões de posicionamento e reflexividade, bem como a importância do autoconhecimento para entender as necessidades reais das pessoas. A ênfase nestes aspetos fomenta a humildade epistémica (Medina, 2013, p.43; ver também Peled, 2018) — o reconhecimento daquilo que não sabemos e o início de um processo de descoberta — e incentiva a adoção de uma perspetiva crítica que reforça o compromisso com a diversidade, não apenas de línguas, mas também de variedades linguísticas.

Além disso, é fundamental promover a consciência metalinguística, que possibilita uma reflexão mais vasta e ajuda a prevenir o chauvinismo de modalidade, isto é, a tendência para tratar as línguas apenas como sistemas de regras e protocolos, ignorando a forma como realmente funcionam em contextos sociais. Entre os temas de investigação, poderá incluir-se também a interatividade e a comunicação intercultural. O progresso significativo depende igualmente de colaboração interdisciplinar que reúna investigadores, responsáveis pelo desenvolvimento e comunidades. Este diálogo não só reforça a capacidade coletiva de enfrentar desafios emergentes, como também suscita uma questão mais profunda de legitimidade: quem autoriza e beneficia da investigação sobre linguagem e tecnologia. Sustentar o princípio “*nada sobre nós, sem nós*” assegura que a reflexão ética e a responsabilidade moral prevalecem sobre o mero cumprimento processual.

Os decisores políticos desempenham, por sua vez, um papel central na criação de condições para uma inovação responsável: promover a transparência, apoiar a diversidade linguística, incorporar auditorias de equidade na legislação e garantir a sustentabilidade financeira de ferramentas desenvolvidas para línguas com poucos recursos. É igualmente fundamental estabelecer mecanismos de fiscalização que responsabilizem programadores, empresas tecnológicas e até proprietários de sistemas pelo cumprimento de padrões de equidade e de justiça linguística. É essencial que os decisores políticos mantenham um diálogo permanente com os programadores, com as comunidades e com os investigadores. A integração desta abordagem de cocriação no enquadramento jurídico que abrange direitos humanos e direitos linguísticos constitui a base do compromisso com estes princípios. Finalmente, é fundamental lembrar que a tecnologia nem sempre é a melhor – ou a única – solução. Cada produto deve ser comparado com a sua alternativa humana: melhora realmente os direitos linguísticos, a equidade e a agência, ou limita-se a automatizar aquilo que deveria permanecer um ato humano de ligação?

As comunidades desempenham um papel determinante na forma como as tecnologias da linguagem evoluem. A sua participação é essencial, não só na expressão de necessidades e aspirações, mas também para garantir que estas perspetivas se refletem – técnica e juridicamente – ao longo de todo o ciclo de desenvolvimento, desde a recolha de dados até à sua implementação. Ao participar na avaliação de ferramentas existentes e na análise de enquadramentos políticos, as comunidades podem contribuir para identificar os pontos onde o acesso permanece desigual e de que modo as tecnologias linguísticas podem melhor servir a equidade e a inclusão.

Esperamos que este número incentive novas investigações sobre os temas aqui explorados e sobre outros domínios em que a IA molda cada vez mais direitos e práticas linguísticas. Uma reflexão contínua e uma investigação colaborativa permitem

contribuir para revelar novos desafios, aprofundar a compreensão e abrir caminho a perspectivas renovadas no debate mais amplo sobre a forma como as tecnologias emergentes estão a transformar a linguagem, a equidade e a sociedade.

Editoras Convidadas

Angela Soltan

Rebekah Rousi

Lucía Ruiz Rosendo

Referências

- Council of Europe. (1992, November). *European Charter for Regional or Minority Languages* (Tech. Rep.). Retrieved from <https://www.refworld.org/legal/agreements/coe/1992/en/41001>
- Council of Europe. (1995, February). *Framework Convention for the Protection of National Minorities. ETS 157*. Retrieved from <https://rm.coe.int/16800c10cf>
- De Meulder, M. (2021). Is “good enough” good enough? Ethical and responsible development of sign language technologies. In *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*.
- European Parliament and Council. (2024, July). *Artificial Intelligence Act. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June. Official Journal of the European Union, L 1689*. Retrieved from <https://artificialintelligenceact.eu/>.
- Gerken, M. (2022). *Facilitating the implementation of the European Charter for Regional and Minority Languages through artificial intelligence*. Council of Europe.
- Greenwood, F., Howarth, C., Poole, D. E., Raymond, N. A., & Scarnecchia, D. P. (2017). “The Signal Code: A Human Rights Approach to Information During Crisis.”. *Harvard Humanitarian Initiative*. Retrieved from <https://hhi.harvard.edu/publications/signal-code-human-rights-approach-information-during-crisis>
- Krausneker, V., & Schügerl, S. (2022). Avatars for sign languages: Best practice from the perspective of deaf users. *ICCHP-AAATE 2022 Open Access Compendium "Assistive Technology, Accessibility and (e)Inclusion, Part I*. Retrieved from <https://epub.jku.at/obvulioa/content/titleinfo/7945395> doi: 10.35011/ICCHP-AAATE22-P1-21
- Medina, J. (2013). Active Ignorance, Epistemic Others, and Epistemic Friction. In *The Epistemology of Resistance* (1st ed., pp. 27–55). Oxford University Press. Retrieved 2025-12-06, from <https://academic.oup.com/book/9202/chapter/155839040> doi: 10.1093/acprof:oso/9780199929023.003.0001
- Peled, Y. (2018). Language barriers and epistemic injustice in healthcare settings. *Bioethics*, 32(6), 360–367.
- Quandt, L., Lamberton, J., Leannah, C., Willis, A., & Malzkuhn, M. (2022, June). Signing Avatars in a New Dimension: Challenges and Opportunities in Virtual Reality. In E. Efthimiou, S.-E. Fotinea, T. Hanke, J. C. McDonald, D. Shterionov, & R. Wolfe (Eds.), *Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology: The Junction of the Visual and the Textual: Challenges and Perspectives* (pp. 85–90). Marseille, France: European Language Resources Association. Retrieved from <https://aclanthology.org/2022.sltat-1.13/>
- Soltan, A., Rousi, R., Prys, C., O’Brien, S., Ruiz Rosendo, L., & Rainò, P. (in press). Language rights in the human-machine era: A multifaceted inquiry. In D. Sayers, M. Glass, H. Kelly-Holmes, & R. Fuchs (Eds.), *Language in the Human-Machine Era: New technologies and the coming transformation of language*. Boston, MA: MIT University Press.
- United Nations. (1992, December). *Declaration on the Rights of Persons Belonging to National or Ethnic, Religious and Linguistic Minorities. General Assembly Resolution 47/135*. Retrieved from <https://www.ohchr.org/en/instruments-mechanisms/instruments/declaration-rights-persons-belonging-national-or-ethnic>.
- United Nations. (2006, December). *Convention on the Rights of Persons with Disabilities. UN Treaty Series, 2515*. Retrieved from <https://social.desa.un.org/issues/disability/crpd/convention-on-the-rights-of-persons-with-disabilities-crpd>.

- United Nations. (2017). *Language Rights of Linguistic Minorities: A Practical Guide for Implementation. Special Rapporteur on minority issues: Geneva.* Retrieved from http://www.ohchr.org/Documents/Issues/Minorities/SR/LanguageRightsLinguisticMinorities_EN.pdf.
- Zanatto, D., Bifani, S., & Noyes, J. (2023). Constraining the Sense of Agency in Human-Machine Interaction. *International Journal of Human-Computer Interaction*, 40(13), 3482–3493. Retrieved from <https://www.tandfonline.com/doi/full/10.1080/10447318.2023.2189815> doi: 10.1080/10447318.2023.2189815

No language left behind? Towards an integrated framework for linguistic rights, human rights and technology regulation

Ingeborg Birnie ¹

¹University Of Strathclyde, Glasgow, Ingeborg.birnie@strath.ac.uk

Abstract

This paper explores the intersection of artificial intelligence (AI) governance and linguistic diversity, focussing on the digital challenges minority and endangered languages face. It presents a novel framework that applies the life-cycle of digital systems—input, process, and output—to analyse the barriers faced by these languages in AI technologies. Through this lens, it critically examines existing international and European legal instruments, revealing a significant policy gap in addressing the explicit inclusion of minority languages in digital and AI domains. The paper’s key contribution lies in its proposal for an integrated, multi-framework approach that combines human rights, minority language protection, and AI governance to ensure equitable linguistic representation. It argues that without urgent and coordinated action, many languages risk digital extinction, with profound implications for cultural identity and access to information. This work calls for comprehensive regulatory reform to secure a digitally inclusive future for all linguistic communities.

Keywords: Language Rights; AI regulation; minority languages; digital inclusion; linguistic justice.

Resumo

Este artigo explora a interseção entre a regulamentação da inteligência artificial (IA) e a diversidade linguística, tendo como foco principal os desafios digitais que as línguas minoritárias e as ameaçadas pelo risco de extinção enfrentam. Apresentamos uma estrutura inovadora que aplica o ciclo de vida dos sistemas digitais — entrada, processamento e saída — para analisar as barreiras enfrentadas por essas línguas nas tecnologias de IA. Nesta perspectiva, examinamos criticamente os instrumentos jurídicos internacionais e europeus existentes, onde se revela uma lacuna significativa nas políticas para abordar a inclusão explícita das línguas minoritárias nos domínios digital e de IA.

A principal contribuição deste artigo reside na proposta de uma abordagem integrada e multifacetada que combina direitos humanos, proteção das línguas minoritárias e regulamentação da IA para garantir uma representação linguística equitativa. Discutimos ainda que, sem uma ação urgente e coordenada, muitas línguas correm o risco de extinção digital, com profundas implicações para a identidade cultural e o acesso à informação. Concluímos apelando a uma reforma regulatória abrangente para garantir um futuro digitalmente inclusivo para todas as comunidades linguísticas.

Palavras-chave: *Direitos linguísticos, regulamentação da IA, línguas minoritárias, inclusão digital, justiça linguística.*

1. Introduction

The last few years have seen a significant rise in the public awareness, availability, and use of generative artificial intelligence (AI) tools for a range of different purposes and across different domains. These tools and technologies are designed to interact with the user and are creating seemingly new and meaningful content (Feuerriegel, Hartmann, Janiesch, & Zschech, 2024) which is increasingly difficult to distinguish from responses that might have been created by humans (Mijwil et al., 2023) and therefore blurring the lines between content created in the real and digitally worlds (Ferrara, 2024). These developments have been described by Krishna (2024) as being part of the ‘fourth Industrial Revolution’ and have the potential to impact different aspects of society, communities and lives of individuals (Farina, Zhdanov, Karimov, & Lavazza, 2024). This has been recognised at an international level by the United Nations who identified that AI technologies and developments can ‘improve access to information, health, education, and public services’, whilst at the same time acknowledging that these developments in AI have the potential to ‘dramatically intensify online harms’ (2023).

This recognition of the potential duality of AI technologies has resulted in the United Nations Advisory Body on Artificial Intelligence, who produced their final report in 2024, calling for “a holistic vision for a globally networked, agile and flexible approach to governing AI for humanity ... [to] address the multifaceted and evolving challenges and opportunities AI presents ... promoting international stability and equitable development” (2024, p. 9) through a flexible framework that balances innovation with safeguards to protect public interest. This recommendation resulted in the adoption of a resolution by the United Nations General Assembly in 2024 that called on Member States to ensure that AI systems operate in compliance with international rights legislation (Mishra, 2024). This resolution, which was universally supported, has come at a time where there have been a variety of initiatives to regulate AI at national, but especially supra-national level. These have typically focussed on the risk that these technologies pose to either individual or state values (Fink, 2021) or are linked to (data) governance (United Nations AI Advisory Body, 2023) but have not explicitly considered the way these AI developments impact on the way individuals use technology to communicate with each other, the world around them, but also the impact on the linguistic and cultural diversity globally.

Linguistic and cultural diversity is under threat globally, with only 1% of the approximately 7,000 languages used in the world today considered to be ‘safe’ (UNESCO, 2023). The remainder, to a greater or smaller extent, is at risk of disappearing as a community language as a result of both external and internal pressures on the communities that use these languages (UNESCO Ad Hoc Expert Group on Endangered languages, 2003). The extent to which languages are endangered and at risk of disappearing varies. Typically, this is evaluated based on a series of different factors, including the size of the speaker population (both in absolute terms as well as the proportion of speakers vis-à-vis other language communities), the levels of intergenerational transmission, availability of the language in the education system, and also the domains in which the language can be used. One of these domains is the response to new media, including digital spaces, with the recognition that the availability of a language in online domains affect the (wider) perceptions around the functionality and possible uses of the language in contemporary society (Cunliffe, 2007). Additionally, there is an acknowledgement that the absence of languages in new technologies affects the way individuals can interact with these tools, with further recognition that developments in the digital era will significantly impact on how languages are represented in AI applications.

Initiatives to regulate AI at a supra-national level, for example, the EU AI Act (Council of Europe, 2024b) have focussed on the risk that these new technological developments pose to either individual or state values (Fink, 2021). They have also focussed on governance (United Nations AI Advisory Body, 2023), or the wider ethical consideration around the use of AI – including the environmental impact. There has also been some recognition that any measure needs to consider the way individuals use these technologies to communicate with each other, the world around them, including being cognisant of the linguistic and cultural diversity, and that ‘no human being or human community should be harmed or subordinated ... during the life cycle of the AI system’ and that any such technologies should “protect, promote, and respect human rights, fundamental freedoms, human dignity” (UNESCO, 2021, p. 18; paragraph 14).

Furthermore, this UNESCO recommendation on the Ethics of AI recognises that that ‘local knowledge, cultural pluralism, value systems and the demands of global fairness to deal with the positive and negative impacts of AI technologies’ (UNESCO, 2021, p. 6). However, to date, this recognition has not resulted in specific measures to actively develop the inclusion of linguistic diversity within these AI tools and applications. This article will provide an overview of the current position and role of minority languages in the digital domains: a precursor to their availability and presence in AI tools and technologies. Through an analysis of the life cycles of digital systems, this article looks at the current European policy and legislative frameworks, which aim to support regional and minority languages. It also observes how these can be used to strengthen the position of these languages and ensure equitable access and opportunities in current and future AI developments for the users of these languages. The issues discussed in this article, although framed within the regional and minority language paradigm, act as a proxy for the issues small(er) state languages are facing, and will continue to face, if decisive action to promote inclusive linguistic and cultural practices are not implemented as a matter of priority.

2. Linguistic diversity and digital domains

The presence of languages in digital spaces and tools is an important indicator of their current availability in AI applications and tools. The United Nations resolution *Seizing the Opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development* recognised that AI systems need to “advance, protect and preserve linguistic and cultural diversity, taking into account multilingualism in their training data and throughout the life cycle of the artificial intelligence system” (United Nations, 2024). This call to include multilingual training data is particularly pertinent in the context of the current linguistic representation in online space, which forms the basis of the datasets used to underpin the AI technologies.

The most widely used languages in online spaces are mainly those with European origins and spoken (either as a first or, increasingly, additional language) across the world as a result of widespread colonisation. Initially this colonisation involved the physical dominance of some states over others, followed by the subsequent direct or indirect imposition of the language of the colonisers in all or some domains (Liu, 2024; Phillipson & Skutnabb-Kangas, 2017). However, as described by Lukianenko (2024), this process of colonisation is no longer associated with the appropriation of land and infrastructure, but instead, through more subtle means, including through the management of, and access to, digital tools, resources and information. The growth of the internet and digital tools create a new global power dynamic as well as exacerbate the already existing socio-economic divide between different communities.

This digital imperialism exerts power over political, economic, and cultural spheres (Akıner, 2024), and has a limited number of actors. These actors are, unlike more traditional forms of colonialism, not state actors, but internationally operating for-profit organisations, which “at first glance [...] seem to be the providers of unlimited access to information and entertainment and the free circulation of information” (Akıner, 2024, p. 135). They also control the type of information, tools and technologies available through their global reach and dominance. This has been recognised by the United Nations AI Advisory Body (2023) who have acknowledged that access to technology is not universal with “developments and rewards [...] currently concentrated among a small number of private sector actors in an even smaller number of states” (p. 5). The subsequent dominance of a small number of languages in digital spaces becomes clear when the availability and presence of different languages in online domains is analysed: English is the (main) language used in 63% of all websites, with a further 9 languages (beyond English) making up 75% of all the internet content (Spanish, Russian, German, French, Japanese, Portuguese, Turkish, Italian, and Persian) (The Centre for Internet and Society, Oxford Internet Institute, & Whose Knowledge, 2022).

These languages, according to the Centre for Internet and Society et al (2022), “all have either a European colonial history ... or are dominant in specific regions where other languages struggle to remain relevant” (n.p), a view further supported by Jin (2013) who identified that although the mechanism by which this colonisation is happening has changed, the languages that are involved have remained the same. The limited use of languages in online spaces means that to access a wide variety of different types of content, tools and information, individuals will need to learn and then use these in a language in which these materials are available (Pimienta, 2022). This digital imperial-

ism disproportionately affects languages that are considered to be endangered, as the users of these languages have typically already faced assimilation into larger or state languages to access goods, services and information. This includes those provided by the state, and entails challenges to maintain their linguistic and cultural practices, while facing with the further additional barriers in having to learn another (international) language to interact with digital tools and technologies. Research in the European context has shown that lack of availability of resources and tools in a language is a significant barrier to their use. Fewer than 10% of all people in the European Union are “willing or able to use online services in English” (Rehm & Uszkoreit, 2013, p. 22) despite generally high levels of linguistic competences in the language. Not only does this limit individuals’ (full) access to digital spaces and technologies, it also further contributes to the digital divide as those with higher levels of education will be more likely to be able to access and use these tools (Cruz-Jesus, Vicente, Bacao, & Oliveira, 2016).

Even though the impact of digital imperialism is, perhaps, most obvious for languages that are currently considered endangered or at risk of disappearing as community languages, this process is beginning to impact languages which, measured by conventional language vitality assessments, would be considered ‘safe’. This includes, in the European context, many of the official state languages (Rehm & Uszkoreit, 2013). These state languages are typically being used in less than 1% of all available internet content (the Centre for Internet and Society et al, 2022) with research by Rehm (2018) suggesting that digital support for many of these state languages are non-existent. Unlike minoritised or endangered languages, these state languages might not currently face the pressures of language shift in non-digital domains. However, as Shohamy (2006) identified, the language used online impacts and influences off-line communications. The dominance of a small number of languages in digital spaces is likely to influence long-term linguistic practices.

As already stated, the limited availability of digital tools across different languages is, in part, due to the commercial nature of these technologies: tools and resources are typically developed and promoted by (large) for-profit organisations that operate at an international level (Flew, Martin, & Suzor, 2019), aiming to attract a large user-base (Gurbanova, 2022). Small(er) state, regional, and minority languages typically have a smaller user population, and this means that technological developments in these digitally disadvantaged languages is not as economically attractive. This means that these languages, unless supported at state level or through active requests by the community, will not be (automatically) considered for inclusion. This results in fewer digital resources, tools, and provisions influencing the ability of individuals and communities to “share in scientific advancement and its benefits”, as set out in Article 27 of the Universal Declaration of Human Rights (United Nations, 1948).

3. European policy and legislative frameworks

The commercial nature and international reach of digital platforms means that their regulation is complex. Digital platforms typically have their own mechanisms for regulation, although according to Cunningham and Craig (2019) and further supported by Obia (2023), this mainly serves the corporate interests of the platform rather than offering explicit support and protection to its users. These monitoring and moderation procedures can include some mechanisms to support and protect vulnerable groups

(Bruning, Alge, & Lin, 2020). However, their implementation relies mostly on the policies and wider governance instruments and their application by the platforms themselves.

At the state level, there have been requests for content or information to be removed from these platforms (Leerssen, 2015). X (previously known as Twitter) for example, received 97,006 government requests for content to be removed, including 3,831 from the European Union countries in the period July to December 2024 (X, 2025). Furthermore, there have been instances of states implementing restrictions and technical restraints to block or limit access to certain types of applications or content, for instance, in China or in Brazil where social media platforms have been blocked on either a permanent or temporary basis. De Souza Abreu (2018) suggests that this allows states to exert control over the content available to individuals living within (or as a part of) their jurisdiction. These decisions can be based on (alleged) violations of legal and constitutional provisions.

Elements of the digital applications, platforms, and technologies, will be subject to local, regional or state level regulation and support – for example through providing the infrastructure required to access digital technology or through limiting certain types of content and response (AlAshry & Al-Saqaf, 2024; Bao, Sun, & Teplitskiy, 2025). Yet, the nature of the advancement in technology – including the governance of characters and scripts and how these are rendered through input systems (for example keyboards) or displayed on an output system (e.g., a display), means that many aspects require a multi-agency and international approach towards regulation and governance. Many of these regulations - as will be discussed below – do not consider the nature of the language except as being a ‘system’. Although, provisions may be made for input to enable the possibility to use minority languages that do not use a standard script or have diacritics to accurately represent the language.

At the European level, support for digital domains and online spaces for small(er) languages and their user community is implicitly included in the two current frameworks that aim to protect linguistic and cultural diversity of (indigenous) minorities: the European Charter for Regional and Minority Languages (ECRML); and the Framework Convention for the Protection of National Minorities (FCNM). It is important to recognise that not all countries, which are part of the Council of Europe, have ratified and / or are signatories to either or both the ECRML with and the FCNM. Furthermore, there is a degree of flexibility in the application of their provisions in a particular state, with both elements affecting the extent of the overall implementation. The ECRML and FCNM fall under the Democracy and Human Dignity Directorate that aims to protect human rights and dignity, strengthen democratic governance, foster innovation, and promote participation and diversity. It also aims to fulfil a separate yet complimentary role in supporting minority communities in the Council of Europe area.

The aim of the ECRML is to protect the historic regional and minority languages (RMLs) – defined in the charter as those languages “traditionally used within a given territory of a State by nationals of that State who form a group numerically smaller than the rest of the State’s population and which are different from the official language(s) of that State” (Council of Europe, 1992). The remit of this Framework Convention goes beyond that of ECRML to cover wider issues associated with the rights of persons belonging to national minorities to enable them to “express, preserve and develop [the ethnic, cultural, linguistic and religious] identity [of each person belonging to a na-

tional minority]” (Council of Europe, 1995, p. 1). However, there are also significant areas of overlap between the ECRML and the FCNM in terms of support for the use of minority languages, including its provisions in the media. However, as recognised by Oeter (2013), the ECRML and the FCNM approach minority languages from different angles, with the ECRML setting the standards for positive action, while the FCNM takes a “classical human rights protection” approach (p. 220). Both the ECRML and FCNM also include languages that are used (as main or state languages) within ‘kinstates’ and thus might receive support through different measures – including at the state level – although this might not necessarily result in a high level of digital inclusion (see Rehm, 2018).

These frameworks aim at protecting small(er) languages and their communities in general, including both implicitly and explicitly supporting their (increased) use in a variety of different domains, for example in interactions with the authorities, the education system and the media. This also (indirectly) includes digital spaces. The aspects of both these policy and legislative frameworks, which most obviously link to the developments in AI and technologies, are the provisions made regarding the inclusion of these languages in the mainstream media. Both the ECRML and the FCNM reference this domain explicitly (Article 11 and Article 9 respectively). However, both the ECRML (and the FCNM) predate the rise of the internet and the increasing influence of online and digital media on the lives of individuals. This includes those affiliating with (national) minorities and how individuals and communities might (potentially) be using their respective languages in these domains to support their overall maintenance and revitalisation. Thus, the ECRML and FCNM focus more explicitly on traditional media (radio and television) which are also (more readily) regulated through traditional measures at state level.

This has resulted in McMonagle (2012) suggesting that technological advancements have been (largely) overlooked “either by states that are party to the ECRML or by the Committee of Experts that conducts three-yearly monitoring exercises of those states and their regional and minority languages” (p. 7). These concerns were brought to the attention of the Committee of Experts of ECRML (COMEX) in December 2024 (Council of Europe, 2024a), following the report produced by Jones, Lainio, Moring, and Resit (2019). With this report called for a mechanism for assessing the use of new media *im* and *for* RMLs in the wider context of the monitoring of this charter, whilst at the same time, acknowledging the challenges of explicitly monitoring the inclusion of regional and minority languages in digital and online domains within the parameters of the current framework of the ECRML.

Although not receiving as much (public) commentary within the academic discourse, similar issues have arisen in the State Opinions of the Advisory Committee of the FCNM that have considered and commented on the availability in online spaces and digital domains of languages used by national minorities as part of the wider media-environment. This exists typically through the availability and use of social media platforms which allows community members to exchange and share information and communicate with each other. This has also included online communication with the authorities at local, regional and national level, for example through the provision of information or service portals in different languages. To date, there have not, as yet, been any active recommendations through the Opinions to States to support the inclusion of

these languages in online spaces. That is, there has not been support beyond measures to make information available to all national minorities and ensure that individuals affiliating with national minorities can participate in the “cultural, social, and economic life and in public affairs” (Article 15).

This does not mean that new tools and applications, including those using AI, have not been considered in the implementation of the various aspects of these frameworks, and in particular the ECRML. Gerken (2022), in her report written with the Secretariat of the ECRML, identified how different tools and technologies could benefit RMLs users and enhance the extent to which these languages can be used to access services (including with the public and judicial authorities), support language acquisition, and increase their use in private domains, a view also supported by Glass, Inge, and Ross (forthcoming). This report also recognised that, although these developments would be highly desirable, ‘resolute action’ is required to ensure that these technologies can be made available in these languages (p. 21) but without addressing what this action might consist of and how any such recommendations might be enacted in practice.

As identified by Grützner-Zahn and Rehm (2022) any such ‘resolute action’ to support regional and minority languages (RMLs) in digital spaces should consider the wider socio-economic ecosystem of the language, including the existing support (both financial and in terms of the availability of digital technology). Moreover, these should also aid the capacity from within the language community to support such developments. This is important as technology is increasingly complex, relying on the interplay of many different components that make up the lifecycle of a digital system. These digital systems consist, at their most basic, of three interrelated elements: the input, process, and output. This means that any action to support inclusive practices for minority languages needs to consider these various aspects of the lifecycle (Birnie, Ross, & Glass, 2025): each of these aspects contribute to the overall availability, quality and quantity of digital provisions, and, as will be seen in the following discussion, rely on different policy and regulatory frameworks for their regulation.

4. Life-cycles of digital systems

AI technologies generally require a large training corpus to ensure an accurate response, in terms of the extent of content these tools can produce, as well as the linguistic range and accuracy within a given linguistic context (Le, Bigi, Besacier, & Castelli, 2003). The training data for these technologies is typically based on the publicly available texts and other online resources developed in the digital era. This means that (within the current technological frameworks) the languages that currently have the greatest digital representation will also have the most accurate and advanced outputs, further increasing the already existing digital divide between languages. To establish the challenges and barriers that digitally disadvantaged languages face, including those that are covered by the ECRML and the FCNM, there has to be an evaluation of the current challenges and barriers these languages are experiencing. It also needs to include considerations for the provisions that are required to ensure that they can be represented in AI technologies. This supports the goal, in the words of UNESCO (2021) that “no language is left behind” (n.p.). This analysis needs to consider the basic building blocks that make up online technology cycles: the input, the process, and the output.

Any digital system starts with the input. The input can be defined as the initial interaction by the user of the application that activates the communication between an individual and the technology. This interaction needs, at its most basic, access to technology, incorporating both the physical hardware (through internet enabled devices) and reliable connectivity. Barriers to access include physical connectivity issues (for example, access to broadband or reliable mobile internet access), as well as costs associated with accessing technology hardware (including devices and equipment). Additionally, the digital literacy of individuals within a community also presents barriers. These various factors affect the overall use of digital technologies for all individuals. In particular, minority language users and their communities are more likely to be economically and socially disadvantaged (UNESCO, 2003), and this is also reflected in their overall access to digital tools and access (Duarte, 2017; Steinhauer-Mozejko, 2024).

Minority language users are more likely to live in rural or remote areas where connectivity might be more limited, with a higher reliance on mobile signals or satellite connections (Soylu & Şahin, 2024). Providing this infrastructure in the first instance requires a high level of investment, both at the state as well as international level, and typically relies on partnerships with for-profit organisations. These partnerships make their availability vulnerable to changes in socio-political circumstances and the power relationships between states and for-profit organisations (Abels, 2024), with Gertz and Evers (2020), recognising that “businesses have become key actors in contemporary politics” (p. 199). Furthermore, although these provisions might contribute to reducing the digital divide, their relatively high costs (Oughton, Amaglobeli, & Moszoro, 2023) contribute to, or even increase, the socio-economic pressures these communities might already be facing (Pinhanez, Cavalin, Vasconcelos, & Nogima, 2023; Tepper, 2023).

Even where access to these technologies is available, a further barrier is the interaction with the system to be able to provide an input. Currently, the most common modality used in digital spaces is written text. To allow a language to be used as an input to a digital system the characters that make up the language (including any diacritics or additional symbols) need to be recognised. This requires the language and all its characters to be represented in Unicode – the underlying internationally agreed standards that govern the use of texts in digital systems. It also needs to be recognised and included by input systems (such as keyboards) (Diki-Kidiri, 2009; Yacob, 2006) and can be displayed on a screen (Hossain, 2024). Both the inclusion of characters in the Unicode database and the creation of tools that allow for the creation of inputs, can be initiated by the language community. This can be a time-consuming process, requiring technical expertise that might not be available within the language community itself.

Unicode encoding and support for input tools requires the languages to have an (agreed) orthography that the users are familiar with. This familiarity will, in turn, depend on the literacy levels of the language users. This is particularly pertinent in terms of the input, as this determines the quality and accuracy of the next stages of the digital life-cycle (the process and the output). Minority language users face additional challenges here, as to acquire a high level of literacy in their language there has to be some educational provision to support the development of these skills. Chiaráin et al. (2022) suggest that the level of literacy in a language determines the confidence and willingness of individuals to use their language in digital spaces.

This is very much an under-rated and under-reported factor in discourse around the equitable provision of digital tools for minority languages. Focus is placed typically on the availability of suitable input mechanisms, support for the process and the accuracy and inclusion of these languages in the output. Although education is recognised by both the ECRML and the FCNM (mirroring the Universal Declaration of Human Rights (United Nations, 1948)) as being an important contributor to the knowledge of the language (and culture), especially in communities where intergenerational transmission might be limited, it also plays a fundamental role in creating the conditions for current and future use of languages, across all domains including digital spaces. This is especially the case in contexts where the language might have a largely oral tradition, and therefore literacy is not necessarily part of the community linguistic practices.

With this said, although the majority of inputs will be based on written texts, and thus orthography and literacy skills, increasingly there are different input modalities, which include spoken interactions. Perhaps even more so than the development of input text-based tools, these developments require significant community support and involvement to meet the minimum required input (de Wet et al., 2023). This firmly places the emphasis on individuals and communities themselves rather than being supported through policy or other regulatory frameworks (Armentano-Oller, Marimon, & Villegas, 2024). Different modalities, especially where these contain audio or video materials that can identify the user of the language, are particularly sensitive to data breaches - an issue that has already received attention from international regulatory frameworks (as will be discussed in the next session).

The second element of these digital systems life cycles is the process: the way in which the initial input is analysed and a response formulated, which is then presented as the output. Current AI technologies are founded on the presence of publicly available corpora (typically from internet or other open-sources) to create a large language model (LLM): an AI model that is trained on a large data set (Ozdemir, 2023), which then uses statistical modelling to produce an output (Aydin & Karaarslan, 2023). The output is designed to approximate human-like responses (Feuerriegel et al., 2024). The mechanism by which this typically happens is hidden from the user (a “black box” (Card, 2017)) and involves the use of algorithms, frequency analysis, and statistical inferences. The operation of the black box is not language specific: the processes can be trained on any data, assuming there is a sufficiently large data sample to allow the input to be processed into a meaningful output (Aydin & Karaarslan, 2023). As with all statistical modelling, the larger the corpus, or training data, the more accurate the output produced will be. This is true in terms of the content it can produce, as well as the linguistic range and accuracy within a given language context (Le et al., 2003).

Even where a digitally disadvantaged language (including all the languages covered under either the ECRML and FCNM) is recognised as an input to the system and can be displayed, it might not have a significant online presence. This impacts the size of the corpora available, and thus, the quality of the responses that the system is able to provide – resulting in inaccurate, incorrect, or unreliable outputs. This means that within the current technological frameworks and paradigms languages with the greatest digital representation also have the most reliable and advanced output, contributing to the digital divide (Cahyawijaya, 2024).

Furthermore, in smaller corpora there is a significant risk of data protection legislation breaches, including that of data, which can be directly traced to individuals, locations, or events within the community. As Pinhanez et al. (2023) discussed, digital technologies have played, and are continuing to play, an important role in language documentation initiatives as part of revitalisation efforts. This has, in some instances, included the digitisation of materials recorded and collected in the pre-digital era— including those collected from individuals who are no longer alive (Cahyati & Madya, 2019). Data governance, particularly that which relates to AI technologies, has been a significant concern of the Council of Europe. This has been reflected in their Convention on Artificial Intelligence, the first legally binding international treaty in this field, which opened for signatures in September 2024 (Council of Europe Committee on Artificial Intelligence, 2024). This Act (under Article 11), recognises the need for personal data and privacy rights of individuals to be protected through “applicable domestic and international laws, standards and frameworks” with elements of this covered under the European General Data Protection Regulation (GDPR) framework. The GDPR framework allows individuals to control their personal data in the face of technological advances (Torre, Alferez, Soltana, Sabetzadeh, & Briand, 2021). However, the complexity of the statistical analysis in the “black box” makes it complex to understand and evaluate how personal data is used and then rendered in the output. The likelihood that any data could potentially be used to ‘train’ these databases might make individuals more reluctant to engage with digital spaces – resulting in an even smaller corpus.

Although these concerns might be related to personal data governance, issues around literacy might also affect the willingness of individuals to create content, but also the quality of the data. In many instances this will relate to written texts, and can also consist of audio or video content and / or other modalities in which languages are used. The quality of the data determines how this can be used in applications – data can be poor in quality because of the input mechanism (for example, where materials are transcribed or transferred from physical sources), or due to their age or limited range of domains. This can also be as a result of errors in the input itself, for instance through spelling mistakes or incorrect grammatical constructions. The quality of the input is significant as any biases in the system will be introduced within the dataset that is used to train the system. In a small dataset, as is typically the case for minority languages, any such errors or limitations are more prone to amplification as there will not be enough additional input to counter these. This will also result in users of these systems and applications becoming more biased and prone to promoting stereotypes (Kotek, Sun, Xiu, Bowler, & Klein, 2024), as they assume that the output generated is accurate (Glickman & Sharot, 2024).

The final element of a digital systems lifecycle is the output or ‘end-stage’: the combined product of the input and the process. The output is the response that the user receives as a consequence of their interaction with the system. The quality of the output content will be very dependent on the earlier stages – the input as well as the way in which the data has been processed. As explained by Bender, Gebru, McMillan-Major, and Shmitchell (2021), AI systems act as a “stochastic parrot” where the output is created by the “haphazardly stitching together [of] sequences of linguistic forms it (= the system) has observed in its vast training data, according to probabilistic information about how they combine, but without reference to meaning” (p. 617). The quality of

the output is significant as any biases in the system, which in turn, will be introduced within the dataset that is used to train the system, and which will be (negatively) affected by the size, will be amplified. This will also result in the users of these systems and applications becoming more biased, as they assume that the output generated is accurate (Glickman & Sharot, 2024).

Furthermore, the combination of limited data that is used to create a corpus, coupled with a small(er) user base, results in more limited output and thus availability of tools and technologies using these languages. Many projects claiming to be inclusive, for example the “No Language Left Behind” project (META, 2025), supported by UNESCO, only include a limited range of languages, in this case 200 (representing around 3% of all languages), and Wikipedia available in 340 languages. Both examples include minority or endangered languages, and aim to allow individuals to share information and communicate regardless of their language preferences, but have large discrepancies in training materials supporting the creation of the LLM underpinning these tools. They are thus not yet able to create an output that is equally accurate and detailed in all the languages.

Outputs that are not ‘fit for purpose’ or contain significant errors not only risk being used (in turn) to train the corpora further (through a feedback loop) – but also can result in the wrong information being shared, causing potential harm to the community. The quality of the output is significant, not only in terms of the user experience, but also recognising that “everyone has the right to freedom of opinion and expression ... and to seek, receive and impart information and ideas through any media” (United Nations, 1948 Article 19) – and that this needs to be accurate and reliable to ensure equitable access to quality information. However, where these digital outputs are not available, or where they are inaccurate or have a more limited functionality, digitally disadvantaged language users and their communities are ‘pushed’ into using these applications in languages in which they are available to ensure an equitable end-user experience.

5. Conclusions

The breakdown of the technology life cycle into the input, process, and output allows for an identification and breakdown of the challenges that many languages face, those that are already, to a greater or lesser extent, endangered but also, as shown by Rehm (2018), languages which are currently considered to be ‘safe’ by conventional language assessment measures. Yet, they are at risk of digital extinction. As discussed in this article, each of these stages introduces and compounds the challenges that users of digitally disadvantaged languages face in accessing technologies in these languages.

While the existing international and European minority language frameworks - such as the ECRML and the FCNM - aim to protect linguistic and cultural diversity, the interpretation of the provisions of these instruments will need to be re-adjusted to consider the current complexities of the provision for a digital presence of the languages. Any further regulations need to ensure that representation of individuals, communities, cultures and languages is accurate and a reflection of current values and lived experiences of the communities. This requires concerted efforts on the part of the authorities to ensure that communities are supported in contributing accurate data to any dataset,

whilst also ensuring that there is an appropriate recourse for removing data that does meet current privacy legislation standards.

Furthermore, the issues affecting digitally disadvantaged languages neither fully sit within the human rights framework, nor solely within the wider frameworks for the protection of minority language nor does the AI governance regulation cover all aspects of the life cycle for these languages. Each of the current policy and legislative frameworks in these areas, both at a global level (for example through the United Nations and UNESCO) and the wider European context (through the Council of Europe), approaches these aspects from a different angle, although with clear overlaps between them. Although, in theory, there are a range of recommendations and protections in place to support the inclusion of digitally disadvantaged languages (and specifically indigenous, regional, and minority languages respectively) and protect the individuals and communities that use these languages, all frameworks are implemented and overseen through different mechanisms. This means that the oversight is currently limited. This therefore requires an overarching (regulatory) framework or body that can consider and incorporate these various elements: one that is cognisant of the current lack of digital inclusion of languages, and which recognises that any language can become digitally endangered if the various aspects that make up the lifecycle of technological systems are not respected and supported.

Acknowledgements

The author wishes to extend her thanks to those involved in Language in the Human Machine Era COST Action, especially those participating in the events organised by Working Group 4 (Language Diversity, Vitality and Endangerment) and Working Group 3 (Language Rights), with the discussions held very much informing the contents of this article with a particular thanks to Dr Maggie Glass (TU Dortmund) and Dr Melody Ross (Universität Duisburg-Essen).

References

- Abels, J. (2024). Private infrastructure in geopolitical conflicts: The case of Starlink and the war in Ukraine. *European Journal of International Relations*, 30(4), 842–866. Retrieved from <https://journals.sagepub.com/doi/10.1177/13540661241260653> doi: 10.1177/13540661241260653
- Akner, N. (2024). A theoretical approach to the digital colonialism in the context of media imperialism. *ICONSR 2024*, 131.
- AlAshry, M. S., & Al-Saqaf, W. (2024). Constraints on AI: Arab Journalists' experiences and perceptions of governmental restrictions on ChatGPT. *Journal of Information Technology & Politics*, 1–21. Retrieved from <https://www.tandfonline.com/doi/full/10.1080/19331681.2024.2421388> doi: 10.1080/19331681.2024.2421388
- Armentano-Oller, C., Marimon, M., & Villegas, M. (2024). Becoming a high-resource language in speech: The catalan case in the common voice corpus. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*.
- Artificial Intelligence Act. (2024). Retrieved from <https://data.consilium.europa.eu/doc/document/PE-24-2024-INIT/en/pdf>

- Aydin, , & Karaarslan, E. (2023). Is ChatGPT leading generative AI? What is beyond expectations? *Academic Platform Journal of Engineering and Smart Systems*, 11(3), 118–134. Retrieved from <http://dergipark.org.tr/en/doi/10.21541/apjess.1293702> doi: 10.21541/apjess.1293702
- Bao, H., Sun, M., & Teplitskiy, M. (2025). Where there's a will there's a way: ChatGPT is used more for science in countries where it is prohibited. *Quantitative Science Studies*, 1–23. (Version Number: 4)
- 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). Virtual Event Canada: ACM. Retrieved from <https://dl.acm.org/doi/10.1145/3442188.3445922> doi: 10.1145/3442188.3445922
- Birnie, I., Ross, M., & Glass, M. (2025). *Addressing Inequalities Faced by Regional and Minority Languages of Europe in the Human Machine Era. Policy recommendations to mitigate harms and facilitate improved access to artificial intelligence for regional and minority languages users across Europe. LITHME Working Group 4.*
- Bruning, P. F., Alge, B. J., & Lin, H.-C. (2020). Social networks and social media: Understanding and managing influence vulnerability in a connected society. *Business Horizons*, 63(6), 749–761. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S0007681320300999> doi: 10.1016/j.bushor.2020.07.007
- Cahyati, P., & Madya, S. (2019). Teaching English in primary schools: Benefits and challenges. In *3rd International Conference on Current issues in Education (ICCIE 2018)*.
- Cahyawijaya, S. (2024). *LLM for Everyone: Representing the Underrepresented in Large Language Models* (Doctoral dissertation). University of Science and Technology, Hong Kong. (Version Number: 1)
- Card, D. (2017). The “black box” metaphor in machine learning. *Medium*. Retrieved from <https://dallascard.medium.com/the-black-box-metaphor-in-machine-learning-4e57a3a1d2b0>
- Chiaráin, N. N., Nolan, O., Comtois, M., Robinson-Gunning, N., Berthelsen, H., & Chasaide, A. (2022). Using speech and NLP resources to build an iCALL platform for a minority language, the story of An Scéalaí, the Irish experience to date. In *Proceedings of the Fifth Workshop on the Use of Computational Methods in the Study of Endangered Languages*.
- Council of Europe. (1992). *European Charter for Regional or Minority Languages*. Retrieved from <https://rm.coe.int/1680695175>
- Council of Europe. (1995). *Framework Convention for the Protection of National Minorities*. Retrieved from <https://rm.coe.int/168007cdac>
- Council of Europe. (2024a). *COMEX 80th plenary meeting*. Retrieved from <https://www.coe.int/en/web/european-charter-regional-or-minority-languages/-/comex-80th-plenary-meeting>
- Council of Europe. (2024b). *Framework Convention on artificial intelligence, human rights, democracy and the rule of law*. Retrieved from <https://rm.coe.int/1680afae3c>
- Cruz-Jesus, F., Vicente, M., Bacao, F., & Oliveira, T. (2016). The education-related digital divide: An analysis for the EU-28. *Computers in Human Behavior*, 56, 72–82. doi: 10.1016/j.chb.2015.11.027

- Cunliffe, D. (2007). Minority languages and the Internet: New threats, new opportunities. In *Multilingual Matters* (Vol. 138, p. 133).
- Cunningham, S., & Craig, D. (2019). Creator governance in social media entertainment. *Social Media + Society*, 5(4), 2056305119883428. Retrieved from <https://journals.sagepub.com/doi/10.1177/2056305119883428> doi: 10.1177/2056305119883428
- De Souza Abreu, J. (2018). Disrupting the disruptive: Making sense of app blocking in Brazil. *Internet Policy Review*, 7(3). Retrieved from <https://policyreview.info/node/928> doi: 10.14763/2018.3.928
- de Wet, F., Bukula, A., Karsten, W., Puttkammer, M., Schillack, E., Wierenga, R., & Eisen, R. (2023, January). Localising the Mozilla Common Voice platform for South Africa's official languages. *Journal of the Digital Humanities Association of Southern Africa (DHASA)*, 4(01). doi: 10.55492/dhasa.v4i01.4437
- Diki-Kidiri, M. (2009). *Securing a place for a language in cyberspace*.
- Duarte, M. E. (2017). *Network Sovereignty Building the Internet across Indian Country*. University of Washington Press. Retrieved from <http://www.jstor.org/stable/j.ctvcwn92r>
- Farina, M., Zhdanov, P., Karimov, A., & Lavazza, A. (2024). AI and society: A virtue ethics approach. *AI & SOCIETY*, 39(3), 1127–1140. Retrieved 2025-10-26, from <https://link.springer.com/10.1007/s00146-022-01545-5> doi: 10.1007/s00146-022-01545-5
- Ferrara, E. (2024). GenAI against humanity: Nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science*, 7(1), 549–569. doi: 10.1007/s42001-024-00250-1
- Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, 66(1), 111–126. Retrieved from <https://link.springer.com/10.1007/s12599-023-00834-7> doi: 10.1007/s12599-023-00834-7
- Fink, M. (2021). The EU Artificial Intelligence Act and access to justice. *EU Law live*, 1–4.
- Flew, T., Martin, F., & Suzor, N. (2019, March). Internet regulation as media policy: Rethinking the question of digital communication platform governance. *Journal of Digital Media & Policy*, 10(1), 33–50. Retrieved from https://intellectdiscover.com/content/journals/10.1386/jdmp.10.1.33_1 doi: 10.1386/jdmp.10.1.33_1
- Gertz, G., & Evers, M. M. (2020). Geoeconomic Competition: Will State Capitalism Win? *The Washington Quarterly*, 43(2), 117–136. doi: 10.1080/0163660X.2020.1770962
- Glass, M., Inge, B., & Ross, A. R. (2026). Working Group Four: Language diversity, vitality and endangerment. In D. Sayers, M. Glass, H. Kelly-Holmes, & R. Fuchs (Eds.), *Language in the Human Machine Era: new technologies and the coming transformation of language*. MIT University Press.
- Glickman, M., & Sharot, T. (2024). How human–AI feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour*, 9(2), 345–359. Retrieved 2025-10-26, from <https://www.nature.com/articles/s41562-024-02077-2> doi: 10.1038/s41562-024-02077-2
- Grützner-Zahn, A., & Rehm, G. (2022). Introducing the digital language equality metric: Contextual factors. In *Proceedings of the Workshop Towards Digital Language Equality within the 13th Language Resources and Evaluation Conference*.
- Gurbanova, A. (2022). Problems and prospects for minority languages in the age of industry 4.0. In *The International Symposium on Computer Science, Digital Economy*

and Intelligent Systems.

- Hossain, A. (2024). Text standards for the “rest of world”: The making of the unicode standard and the OpenType format. *IEEE Annals of the History of Computing*, 46(1), 20–33. Retrieved from <https://ieeexplore.ieee.org/document/10384703/> doi: 10.1109/MAHC.2024.3351948
- Jin, D. Y. (2013). The construction of platform imperialism in the globalization era. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, 11(1), 145–172. doi: 10.31269/triplec.v11i1.458
- Jones, E. H. G., Lainio, J., Moring, T., & Resit, F. (2019). *New technologies, new social media and the European Charter for Regional or Minority Languages*. Strasbourg: Council of Europe. Retrieved from <https://edoc.coe.int/en/minority-languages/8265-new-technologies-new-social-media-and-the-european-charter-for-regional-or-minority-languages.html>
- Kotek, H., Sun, D. Q., Xiu, Z., Bowler, M., & Klein, C. (2024). *Protected group bias and stereotypes in Large Language Models*. arXiv. Retrieved from <http://arxiv.org/abs/2403.14727> (arXiv:2403.14727 [cs]) doi: 10.48550/arXiv.2403.14727
- Krishna, V. V. (2024). AI and contemporary challenges: The good, bad and the scary. *Journal of Open Innovation: Technology, Market, and Complexity*, 10(1). doi: 10.1016/j.joitmc.2023.100178
- Le, V. B., Bigi, B., Besacier, L., & Castelli, E. (2003). *Using the Web for fast language model construction in minority languages*. Eurospeech.
- Leerssen, P. (2015). Cut out by the middle man: The free speech implications of social network blocking and banning in the EU. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, 6(2), 99–119. Retrieved from <https://www.jipitec.eu/issues/jipitec6-2-2015/4271>
- Liu, Y. (2024). Linguistic imperialism as a tool in cultural hegemony: Language decline and revitalization of indigenous communities in Canada. *Lecture Notes in Education Psychology and Public Media*, 47(1), 136–141. doi: 10.54254/2753-7048/47/20240899
- Lukianenko, N. (2024). Language and power: Linguistic imperialism. *International Science Journal of Education & Linguistics*, 3(5), 41–49. doi: 10.46299/j.isjel.20240305.06
- McMonagle, S. (2012). The european charter for regional or minority languages: Still relevant in the information age. *JEMIE*, 11(1).
- META. (2025, February). *No Language Left Behind - Driving Inclusion through the power of AI translation*. Retrieved from <https://ai.meta.com/research/no-language-left-behind/#>
- Mijwil, M. M., Hiran, K. K., Doshi, R., Dadhich, M., Al-Mistarehi, A.-H., & Bala, I. (2023). ChatGPT and the future of academic integrity in the artificial intelligence era: A new frontier. *Al-Salam Journal for Engineering and Technology*, 2(2), 116–127. doi: 10.55145/ajest.2023.02.02.015
- Mishra, V. (2024). *General Assembly adopts landmark resolution on artificial intelligence*. Retrieved from <https://news.un.org/en/story/2024/03/1147831>
- Obia, V. (2023). Regulatory Annexation: Extending Broadcast Media Regulation to Social Media and Internet Content. *Communication Law and Policy*, 28(2), 99–123. Retrieved from <https://www.tandfonline.com/doi/full/10.1080/10811680.2023>

- .2206382 doi: 10.1080/10811680.2023.2206382
- Oeter, S. (2013). Working with the Language Charter Committee of Experts. In T. H. Malloy & U. Caruso (Eds.), *Minorities, their rights, and the monitoring of the European framework convention for the protection of national minorities: essays in honour of Rainer Hofmann* (pp. 205–227). Martinus Nijhoff Publishers.
- Oughton, E. J., Amaglobeli, D., & Moszoro, M. (2023). What would it cost to connect the unconnected? Estimating global universal broadband infrastructure investment. *Telecommunications Policy*, 47(10). doi: 10.1016/j.telpol.2023.102670
- Ozdemir, S. (2023). *Quick start guide to large language models: Strategies and best practices for using ChatGPT and other LLMs*. Addison-Wesley Professional.
- Phillipson, R., & Skutnabb-Kangas, T. (2017). Linguistic imperialism and the consequences for language ecology. In *The Routledge handbook of ecolinguistics* (pp. 121–134). Routledge.
- Pimienta, D. (2022, June). Resource: Indicators on the presence of languages in internet. In M. Melero, S. Sakti, & C. Soria (Eds.), *Proceedings of the 1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages* (pp. 83–91). Marseille, France: European Language Resources Association. Retrieved from <https://aclanthology.org/2022.sigul-1.11/>
- Pinhanez, C. S., Cavalin, P., Vasconcelos, M., & Nogima, J. (2023). Balancing social impact, opportunities, and ethical constraints of using AI in the documentation and vitalization of indigenous languages. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence* (pp. 6174–6182). Macau, SAR China: International Joint Conferences on Artificial Intelligence Organization. Retrieved from <https://www.ijcai.org/proceedings/2023/685> doi: 10.24963/ijcai.2023/685
- Rehm, G. (2018). The META-NET strategic research agenda for language technology in Europe: An extended summary. *Language technologies for a multilingual Europe*, 4(19).
- Rehm, G., & Uszkoreit, H. (2013). *Strategic research agenda for multilingual Europe 2020*.
- Soylu, D., & Şahin, A. (2024). The role of AI in supporting indigenous languages. *Sciences*, 2(4), 11–18.
- Steinhauer-Mozejko, P. (2024). *Nêhiyawak Networks: Native Perspectives of Digital Connectivity*.
- Tepper, E. (2023). Space Commercialization is Closing the Digital Divide, but Expanding Global Economic Inequality. *Georgetown Journal of International Affairs*, 24(1), 55–64. Retrieved from <https://muse.jhu.edu/article/897701> doi: 10.1353/gia.2023.a897701
- The Centre for Internet and Society, Oxford Internet Institute, & Whose Knowledge. (2022). *State of the Internet's Languages Summary Report*. Retrieved from <https://internetlanguages.org/en/>
- Torre, D., Alferez, M., Soltana, G., Sabetzadeh, M., & Briand, L. (2021). Modeling data protection and privacy: Application and experience with GDPR. *Software and Systems Modeling*, 20(6), 2071–2087. Retrieved from <https://link.springer.com/10.1007/s10270-021-00935-5> doi: 10.1007/s10270-021-00935-5
- UNESCO. (2003). *Digital Initiatives for Indigenous Languages*. Retrieved from <https://www.unesco.org/en/articles/digital-initiatives-indigenous-languages>

- UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- UNESCO. (2023). *World Atlas of Languages (Beta Version)*. Retrieved from <https://en.wal.unesco.org>
- UNESCO Ad Hoc Expert Group on Endangered languages. (2003). *Language Vitality and Endangerment*. Retrieved from <https://ich.unesco.org/doc/src/00120-EN.pdf>
- United Nations. (1948). *Universal Declaration of Human Rights*. Retrieved from <https://www.un.org/en/about-us/universal-declaration-of-human-rights>
- United Nations. (2023). *UN Global Communications Chief urges AI developers to “put people before profit”*. Retrieved from <https://www.un.org/en/hate-speech/ai-concerns>
- United Nations. (2024). *Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development. 87th Session of the United Nations General Assembly*. Retrieved from <https://docs.un.org/en/A/78/L.49>
- United Nations Advisory Body on Artificial Intelligence. (2024). *Governing AI for Humanity: Final report* (Tech. Rep.). Retrieved from https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf
- United Nations AI Advisory Body. (2023). *Interim report: Governing AI for humanity* (Tech. Rep.). Retrieved from https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf
- X. (2025). *Global Transparency Report. H2 2024*. (Tech. Rep.). Retrieved from <https://transparency.x.com/en/reports/global-reports/2025-transparency-report#government-legal-andlaw-enforcement-requests>
- Yacob, D. (2006). Unicode for under-resourced languages. *Strategies for developing machine translation for minority languages*, 33.

Handling multilingualism by the EU institutions at the time of AI-driven language technologies

Aneta Skorupa-Wulczynska ¹ and Anna Zannina ²

¹Chancellery of the Sejm, aiskorupa@gmail.com

²University St. Gallen, anna.zanina@unisg.ch

Abstract

This article investigates how artificial intelligence (AI)-driven language technologies employed by European Union (EU) institutions affect the relationship between AI and the EU's respect for linguistic diversity. It emphasises the necessity for better protection and representation of under-resourced languages, in particular minority or minoritised languages, in the face of rapid technological advancement. Systemic disadvantages may be faced by regional, minority, and migrant languages and even official EU languages despite their cultural significance. The study identifies key challenges of AI-powered language technologies, such as scarce data resources, tendencies towards linguistic homogenisation, unequal development of language technologies and reinforcement of cultural stereotypes. The findings reveal the pressing need for responsible AI practices within the EU institutions that promote inclusivity, targeted policies, community engagement, interdisciplinary collaboration, and strict human oversight to mitigate bias and ensure fairness. The analysis underlines the critical role of policy initiatives and active involvement of linguistic communities in shaping such AI tools that genuinely promote and uphold linguistic diversity. Only by embracing these principles, the EU can strive for cultivating more inclusive digital spaces and thereby preserve Europe's rich multilingual legacy.

Keywords: AI-driven language technologies, EU institutions, linguistic diversity, minority languages, minoritised languages, under-resourced languages.

Resumo

Este artigo procura analisar como as tecnologias linguísticas baseadas em Inteligência Artificial (IA) utilizadas pelas instituições da União Europeia (UE) influenciam a relação entre a IA e o respeito da UE pela diversidade linguística.

O estudo sublinha a necessidade de uma melhor proteção e representação das línguas com poucos recursos, em particular as línguas minoritárias ou minorizadas, face ao rápido avanço tecnológico. As línguas regionais, minoritárias e migrantes, e mesmo as línguas oficiais da UE, podem enfrentar desvantagens sistémicas, apesar da sua importância cultural. O estudo identifica os principais desafios das tecnologias linguísticas baseadas em IA, tais como a escassez de recursos de dados, as tendências para a homogeneização linguística, o desenvolvimento desigual das tecnologias linguísticas e o reforço dos estereótipos culturais. As conclusões revelam a necessidade premente de práticas responsáveis de IA nas instituições da UE que promovam a inclusão, políticas específicas, o envolvimento da comunidade, a colaboração interdisciplinar e uma supervisão humana rigorosa para mitigar os preconceitos e garantir a equidade. A análise sublinha o papel crítico das iniciativas políticas e do envolvimento ativo das comunidades linguísticas na definição de ferramentas de IA que promovam e defendam genuinamente a diversidade linguística. Só adotando estes princípios é que a UE poderá fazer um esforço para cultivar espaços digitais mais inclusivos e, assim, preservar o rico legado multilíngue da Europa.

Palavras-chave: *Tecnologias linguísticas baseadas em IA, instituições da UE, diversidade linguística, línguas minoritárias, línguas minorizadas, línguas com poucos recursos.*

1. Introduction

Multilingualism has been the cornerstone of European integration since its outset. It is one of the fundamental European Union (EU) principles enshrined in the founding Treaties and the Charter of Fundamental Rights, which aims to express the organisation's respect for Member States' national identities. EU multilingualism is based on the equal status of all Member States' official languages, in compliance with Regulation no. 1/1958. With 27 Member States, the EU encompasses a rich tapestry of languages that mirror the cultural heritage of its Member States and serves as an essential aspect of national identities. Over the years, EU multilingualism has developed into a complex, multilayered concept and a cross-cutting policy that includes practices in EU institutions, multilingual legislation, and the European Commission's strategy on language learning. Managing linguistic diversity in the EU has become ever more complicated and costly due to the growing number of Member States, and as a consequence, official languages of the community. Despite this, the New Framework Strategy for Multilingualism - adopted by the European Commission (EC) in 2005 - reaffirmed the Union's commitment to multilingualism, outlining various action fields aimed at promoting it not only in the economic environment, but also in society and in the Commission's interactions with EU citizens. The Strategy sets out that multilingualism is essential for the proper functioning of the EU and contributes positively to exploiting Europe's potential and achieving European goals (Communication COM(2005) 596 final) (Skorupa-Wulczyńska, 2021).

The approach presented by the EC has been endorsed by EU citizens. According to the Public Hearing on Multilingualism held by the EC in 2008, most respondents supported the view that European linguistic diversity is an asset and should be safeguarded. The Hearing clearly highlighted citizens' desire for the policy to extend beyond mere economic functionality, emphasising the importance of recognising the identities and cultures represented by various languages within the EU. As illustrated by the Eurobarometer of 2024 addressing the language knowledge of Europeans, citizens' attitudes and expectations towards multilingualism remain considerably positive, with 86% agreeing that everyone should speak at least one other language than their native language, and 69% stating that everyone should speak more than one foreign language. A noteworthy 59% of Europeans declared to be able to hold a conversation in at least one other foreign language - a slight increase when compared to 55% reported in the Eurobarometer of 2012. Moreover, approximately 28% of Europeans reported skills to converse in at least two additional languages. This 28% is a subset of 59% representing those who speak two or more languages. Meanwhile, the data also reveals that nearly half of the population (49%) is unable to converse in any foreign language, which appears contradictory to the value of 59%, as adding both values does not equal 100%. However, this discrepancy arises from differences in the formulation of survey questions and rounding, rather than a direct numerical inconsistency. Interestingly, 84% of Europeans are of the opinion that regional and minority languages should be protected in the EU (European Commission, 2024).

Despite the EU's commitment to preserving and fostering linguistic diversity, its language policy continues to face significant challenges. Although noticeable progress has been made in promoting multilingualism, problems persist in implementation, unequal resource allocation, inconsistent policy execution, and the treatment of minority languages. An inclusive language policy should not only guarantee access, but also empower citizens to fully participate in democratic processes. Enshrined in Article 10(3) of the Treaty on European Union, participatory democracy gives citizens the right to participate in the democratic life of the EU and emphasises that decisions should be made "as openly and as closely as possible to the citizen". However, with approximately 60 regional and minority languages spoken by 40 to 50 million people across the EU, many face barriers such as lack of recognition, limited educational resources, and declining numbers of speakers (European Parliament, 2016). Without more effective and inclusive measures, linguistic inequalities could limit democratic participation, as language barriers exclude individuals from meaningful civic engagement. A more adaptive policy is needed - one that balances respect for cultural heritage with citizens' practical communication needs and ensures that all citizens can fully access and contribute to democratic life. AI offers promising tools to bridge these linguistic gaps. Language technologies driven by AI have the potential to enhance accessibility and enable citizens to address EU institutions in their preferred languages. This can help foster wider participation in democratic processes and reduce representational disparities. However, leveraging AI responsibly requires a thoughtful approach that respects linguistic diversity not merely as a functional asset, but as a core element of Europe's cultural identity.

This article explores how AI-driven language technologies employed by EU institutions influence citizens' participation in decision-making democratic processes across a multitude of languages. It examines whether such technologies improve institutional

accessibility and facilitate greater engagement among linguistically diverse populations. To address this question, the analysis focuses on the technologies currently used by the EU institutions for multilingual communication, assessing both their benefits and associated risks. Special attention is given to the implications for under-resourced languages, in particular minority and minoritised languages.¹ In the AI context, these languages require a tailored approach that acknowledges both their significance in cultural preservation, representation, and equitable access to the digital sphere. Unlike mainstream AI development, which often centres on technical performance and market viability, the inclusion of minority and minoritised languages entails distinct challenges, such as safeguarding linguistic survival, addressing algorithmic bias, and ensuring fair participation in Europe's evolving digital democracy.

The article primarily employs a qualitative and analytical approach, drawing from recent literature, policy documents, and case studies relevant to AI-driven language technologies within the EU context. It synthesises interdisciplinary sources, including scholarly research on AI bias, linguistic diversity, and policy analysis, to identify key challenges and opportunities associated with the use of AI in multilingual settings. The approach involves critical examination of technological trends, policy initiatives, and ethical considerations, emphasising community engagement, interdisciplinary collaboration, and human oversight as vital components for responsible AI development. This methodology aims to provide a comprehensive understanding of the systemic issues and strategic solutions necessary for aligning AI practices with the EU's commitment to linguistic diversity and inclusivity.

2. Language technologies at EU institutions – historical background

Recognising the growing importance of managing multilingualism effectively, European Community institutions began exploring language technologies as early as the 1980s, paving the way for today's sophisticated AI-powered services. The EC's commitment to technological innovation was exemplified by the EUROTRA machine translation project (1982–1994), guided by the principle that all citizens should have the right to access EC proceedings in their own language. Whereas EUROTRA failed to deliver a 'working' machine translation system, the project made a far-reaching long-term impact on the language industries in Community Member States (Maegaard, 1995).

In 1995, the EC launched the *European Advanced Multilingual Information System* (EURAMIS), a shared repository of clearly labelled equivalent phrases contributed by various EU institutions jointly. That same year, the EU also founded the European Language Resources Association (ELRA), a non-profit organisation whose main mission was to make Language Resources for Human Language Technologies available to the public. As a response to the multilingual communication needs of the EU institutions and bodies faced by growing translation volumes of varied and specialised content to be pro-

¹The terms 'minority language' and 'minoritised language' both refer to languages spoken by groups that are not dominant in a given society, but they emphasise different aspects. Whereas 'minority' is a demographic term, 'minoritised' is a sociopolitical one. A minority language is defined primarily by numbers, as it is spoken by a smaller segment of the population, which does not necessarily reflect the social or political status of the language. A minoritised language highlights the process of marginalisation, where a language is actively suppressed or excluded from public life, education, or governance, often due to historical or systemic power imbalances.

cessed with limited budgetary resources, the European Commission Translation Centre supported by the European Union Intellectual Property Office implemented an ambitious programme focused mainly on using AI for translation purposes. A key outcome of this initiative was the launch of *eTranslation* (November 2017), a cutting-edge neural machine translation service. *eTranslation* is able to deliver raw machine translations in 24 official EU languages as well as in Arabic, Chinese, Icelandic, Japanese, Norwegian, Russian, Turkish and Ukrainian, and could be used to get the gist of a text or as the starting point for a human-quality translation. Additional outputs of the programme include a suite of AI multilingual services collectively known as *eLangTech*, such as *Speech Transcription* (a speech-to-text-tool), *eSummary* and *eBriefing*, and *Website Translation Tool (WEB-T)*. These services have been made available under the Digital Europe Programme for use by EU institutions, public administrations, academia, SMEs, or NGOs.

The publication of the Regulation establishing the Digital Europe Programme on 11 May 2021 was a breakthrough moment in fostering Europe's digital transformation by assuring EU funding for AI, including AI-based multilingual technologies. The EU institutions have become the major beneficiaries of the Digital Europe Programme's AI-powered tools and has served as its testbed at the same time. By integrating AI multilingual services into their workflows, the EU institutions aimed not only at improving their internal efficiency, but also at increasing access to information for EU citizens, regardless of the language they speak. For instance, *eTranslation* was widely used during the COVID-19 pandemic by enabling dissemination of up-to-date health information to EU citizens in 24 official EU languages, via the EU's official social network channels. These platforms are aligned with the aforementioned service of *eTranslation*, which constitutes the cornerstone of General Data Protection Regulation (GDPR)-compliant EU neural machine translation services.

Beyond language technologies, the EU has also focussed on the expansion of the European digital infrastructure, which plays a crucial role in fostering linguistic diversity and protecting minority rights. Initiatives such as the Common European Language Data Space (a three-year project initiated in 2023) and the Alliance for Language Technologies (ALT-EDIC) (established under Commission Implementing Decision (EU) 2024/458) aim to build a robust language data ecosystem. These projects seek to improve AI-driven translation and NLP models while ensuring accessibility for all linguistic communities. With collaboration across multiple EU Member States, they strive to boost AI competitiveness, enhance data-sharing mechanisms, and reinforce strategic autonomy. The OpenEuroLLM initiative, awarded in 2025 under the Digital Europe Programme, represents another promising step towards the reinforcement of the EU digital infrastructure. It focuses on training AI models that uphold European values of transparency, openness, and accessibility, in full compliance with the AI Act. For these efforts to be truly inclusive, dedicated resources must be allocated to minority languages, ensuring that digital tools bridge communication gaps rather than deepen them.

3. AI strategies of EU institutions

The EU has embraced AI as a transformative force in multilingual governance, with its core institutions - the EC, the Court of Justice of the European Union (CJEU, Court) and the European Parliament (EP) - developing tailored strategies to integrate AI tools into their operations. While each institution clearly approaches AI from a distinct functional

perspective, together they seem to reflect a shared commitment to safe, transparent, and inclusive digital transformation.

Firstly, the EC is the primary initiator of AI-based translation tools within the EU institutions under the Digital Europe Programme. Its strategic vision, outlined in the *AI@EC* Communication (C(2024) 380 final), outlines a roadmap for fostering lawful, safe, and trustworthy AI systems. This document details actions to build institutional and operational capacity, ensuring that AI use remains human-centred and transparent. To support this vision, the EC has introduced a range of internal enablers, including a prompt library, collaborative platforms, and knowledge-sharing mechanisms. The *AI@EC* policy also addresses change management, communication strategies, and IT governance through tools such as the digital innovation dashboard and digital landscape mapping. The oversight is provided by the European AI Office, which is a central authority for AI governance and enforcement of the AI Act, the Interservice Steering Group on AI - responsible for coordination across EC departments and the *AI@EC* Network facilitating knowledge exchange and alignment across services.

Complementing the EC's institutional focus, the Court of Justice of the European Union released its official AI strategy in January 2024. The document details a clear methodology that encompasses governance models, risk assessment protocols, and specific mitigation strategies for the integration of AI into the Court's mechanism for managing the multilingual legal translations. The Strategy envisions AI as a tool to enhance efficiency, accessibility, and transparency in judicial processes, yet at the same time it emphasises that human expertise remains indispensable. It advocates a synergistic model where AI supports, rather than replaces legal professionals. The Strategy outlines five interconnected AI capabilities relevant to multilingual language processing: natural language processing (NLP), advanced data analytics, chatbot systems, speech recognition, and, to a certain extent (through multimodal applications), computer vision. These operational goals are framed within the document as practical applications of core EU values, such as inclusion and democratic engagement. It emphasises the importance of neural translation and interpretation technologies in bridging language gaps within legal communication. However, the Strategy also notes that the use of AI for legal translation presents distinct challenges of complex legal terminology, contextual sensitivity, and the need for high precision. NLP-based translation supports both internal communication at the Court and external interactions with national courts and EU citizens. Tools like SIGA - an integrated and adaptive case management solution - are being developed to increase institutional efficiency, offering multimodal, contextual, and conversational AI-driven translation support. Looking ahead, the potential for cross-lingual database search - where users retrieve information in a language different from their query - is under exploration. Yet, semantic drift and domain-specific ambiguities pose ongoing challenges to such capabilities in legal contexts. It remains to be seen whether such tools can handle the nuanced legal terminology across all 24 EU languages with sufficient accuracy and contextual awareness.

The Strategy rightly acknowledges several embedded risks in deploying AI-based language technologies in multilingual legal environments - ranging from algorithmic bias that skews translation of minority languages, to data privacy concerns when handling sensitive legal communications, to the challenge of managing AI 'hallucinations' in high-stakes judicial reasoning. It further warns against insufficient oversight, which

may lead to systemic misuse, such as exposure to security vulnerabilities or unintentional infringement of intellectual property rights. To mitigate these risks, the Strategy introduces a high-level capability map, ensuring that AI functions can be reused efficiently. For instance, speech-to-text technology could aid in transcription, meeting minutes, and document drafting. Despite the recurring Strategy's assurances that the language and type of media will no longer pose barriers, one cannot overlook the fact that minority languages have been almost completely neglected. This gap raises concerns about inclusivity as well as ensuring equitable access and representation to all language communities (Soria et al., 2016).

While the EC and CJEU focus on governance and legal precision, the EP emphasises accessibility and public engagement in its AI strategy. The governance of AI in the EP is structured around a dedicated AI Governance Board, which is responsible for establishing and overseeing the rules and principles guiding the use of AI within the institution. The Board plays a central role in ensuring that AI deployment aligns with both ethical standards and regulatory obligations. The Board is chaired by the Secretary General of the Parliament, underscoring its strategic importance and high-level oversight function in shaping AI governance across the organisation. The EP's translation system reflects its broad commitment to linguistic inclusivity and institutional transparency within the EU. The integration of AI tools into this system signals a strategic shift towards efficiency and scalability. However, the Parliament's emphasis on deploying AI that is safe, transparent, traceable, non-discriminatory, and environmentally sustainable suggests a deliberate effort to balance innovation with ethical governance. This approach acknowledges the potential risks of automation in sensitive domains like multilingual communication, where nuance and cultural context are paramount. The system's reliance on historical translation data and reference materials - channeled through EURAMIS, a shared multilingual database - underscores the importance of institutional memory and consistency. By centralising outputs, EURAMIS not only enhances translation quality but also fosters coherence across EU bodies. Moreover, the use of the customised SDL-Trados Studio and Cat4Trad, optimised for multilingual documents with dual-column layouts, reflects a nuanced understanding of document structure and user experience.

In view of the above, the AI strategies of the EC, CJEU, and EP collectively illustrate a complex and strategic approach to digital multilingualism. Rather than treating language as a logistical hurdle, these institutions seem to leverage AI to reinforce core democratic values while optimising operational performance. At the institutional level, AI tools are deployed to streamline translation workflows, reduce turnaround times, and ensure terminological consistency - enhancing efficiency without compromising quality. Nevertheless, despite the promise of AI-driven multilingualism, a notable gap persists across institutional strategies: the limited integration of minority and under-resourced languages. Although technological advancements offer powerful tools to overcome many linguistic barriers, the absence of explicit commitments to minority language inclusion risks reinforcing existing digital divides.

4. Potential of AI-based language technologies

Looking broadly, the integration of technology into political organisations encompasses a wide array of dimensions, ranging from internal organisational practices through process management to external communication strategies and citizen engagement in

democratic systems. AI-driven technologies increasingly influence not only the internal workflows of these organisations, but also their public communication policies, offering new possibilities for multilingual interaction with citizens and enhanced accessibility to particular institutions. AI tools streamline engagement with citizens in their native languages, fostering more inclusive and responsive governance. A notable example of an AI tool commonly used at the EU institutions is the AI Legislative Assistant, which helps manage legislative documents across 24 official languages by automatic translation, summarising and comparing policy proposals and legal texts in multiple languages. The tool aims to improve transparency, detect legal conflicts and enhance accessibility for policymakers and citizens. As evidenced by the EC, the AI Legislative Assistant has streamlined the work of the EP policy teams and associated agencies which reported a reduction of over 60% in the time spent reviewing and comparing legislative texts (European Commission, 2024).

As AI technologies continue to evolve, scholars increasingly examine how digital and algorithmic technologies facilitate coordination among individuals, stakeholders, and their interests (Paul, Carmel, & Cobbe, 2024, p. 33). From a policymaking perspective, this issue necessitates a thorough evaluation of the effects of digital technologies on public trust and institutional governance (e.g., Rabari & Storper, 2015). As Faraj, Pachidi, and Sayegh (2018) argue, algorithmic tools shift the locus of expertise from human judgment to machine-generated insights, challenging traditional authority structures.

Moreover, as noted by Láncoš (2021, pp. 77–79), AI language technologies' advancements have the potential of strengthening the standing of minority languages. They can improve the socio-economic status of minority groups and increase their political participation. By enabling speakers of minority languages to interact with public institutions and essential services in their own languages, these technologies can reinforce the principles of equal opportunity and active citizenship. AI technologies may also positively contribute to minority language preservation by documenting oral traditions, transcribing speech, and creating written records. Tools like speech-to-text and text-to-speech enhance accessibility by providing content in multiple formats. NLP techniques enable the integration and structuring of multilingual data from audio, textual, and visual sources - laying a vital foundation for equitable digital representation. Practical AI solutions - such as intra-language translation (i.e. adapting the text within the same language e.g. by change of the register, complexity or style), machine translation platforms (e.g., DeepL, Google Translate, Amazon Translate), screen readers (e.g., NVDA, Orca), and multilingual chatbots - can also play a key role in facilitating cross-institutional and interpersonal communication.

5. Concerns about AI-based language technologies

Whereas AI-driven translation tools have the potential to provide efficiency gains, enhanced communication, and broader accessibility, their implementation also poses significant risks and operational challenges. The major risks are related to data control, unequal distribution of technological power, and influencing public opinion in potentially unethical ways by disseminating misinformation and manipulation. As highlighted by Kemper and Kolkman (2019, p. 2090), a critical gap in algorithmic literacy persists, and it may inadvertently enable technology experts and infrastructure owners to wield disproportionate influence, potentially undermining democratic processes

(e.g., Williamson, 2018). Moreover, there are also issues surrounding the accuracy and quality of AI-generated output, legal and ethical considerations, transparency of algorithms, and just representation of small and under-resourced languages raise significant challenges. These concerns are of particular importance in the EU multilingual context, where the effective governance of AI language tools is essential to upholding the EU's commitment to democratic principles and linguistic equality.

5.1. Lack of accuracy and inadequate quality

The accuracy and reliability of AI translation tools in official capacities, including EU, have emerged as a primary concern. The reason is that language intricacies play a pivotal role in shaping human thoughts and interactions (Kamaluddin, Abqoriyyah, Khoerul Rasyid, & Saehu, 2024). As people increasingly rely on technology, the nuances of linguistic structures become even more critical. LLMs, such as ChatGPT, alter text generation and reveal specific patterns that differ from human writing, including a tendency towards vagueness and repetition that can hinder critical discussions. This issue is relevant in scenarios where EU institutions communicate with their citizens, when clarity and emotional engagement are necessary. The consequences of these opaque linguistic practices questions the efficiency of communication and its alignment with the EU fundamental democratic principles based on inclusiveness and transparency. The accuracy of AI translation tools is highly relevant for the EU multilingual law framework. LLMs often struggle with precision of legal terminology and specialised language. Misinterpretations in legal or policy documents can have far-reaching consequences, potentially undermining compliance with regulatory standards such as Directive (EU) 2019/1024 on Open Data and the Re-use of Public Sector Information, mandating accuracy and reliability in publicly accessible data. Still, even if syntactic ambiguities have been eliminated, translations generated by AI tools remain inferior to authorised translations due to persistent lexical inaccuracies that may significantly alter meaning.

The quality of AI generated translations raises another serious concern. Research findings show that the quality of AI-powered translation varies considerably across language pairs. As noted by Koponen (2010), translation quality can be evaluated through different aspects such as accuracy, fluency and fitness for purpose. As the process of translation has undergone a significant technological makeover, quality control requires adequate skills of human translators, who face the challenge of adapting to new technologies, roles and processes. In this sense, translators not only translate, but also revise and post-edit texts in the same technological environment (Carmo & Koponen, 2024). The role of human translators in the quality control process is of utmost importance in legal translations. As emphasised by Carmo and Koponen (2024), AI tools' efficiency, speed and ability to handle large volumes of data are highly beneficial capabilities in legal translation. However, careful human review and expert correction is still required, despite technological advancement - especially in fields such as science, administration, and law (Zanina-Seck & Groener, 2025, p. 256). Human translators offer contextual understanding, cultural nuance and ethical judgement, the qualities that AI cannot currently replicate. Last but not least, human translators bear responsibility for the final translation, and therefore they are obliged to monitor its quality.

The issue of quality may be analysed while referring to the EU flagship AI-powered translation tools *eLangTech*, as already mentioned above. Although Varga and Wetzel

(2024) claim that the *eTranslation* service performs well in EU-official language legal, finance and public health domains, it is also evident that the general text option of the service does not handle highly technical, marketing, culturally sensitive content or other specialised content. Moreover, it must be noted that *eTranslation* supports all EU official languages, yet not all language pairs are supported equally well in quality. None of the *eLangTech* tools extend their capabilities to regional and minority languages, leaving many linguistic communities without access to automated translation services. Furthermore, neither *eTranslation* nor *Speech Transcription* and *eSummary* are benchmarked by independent quality standards, which clearly leaves the issue of quality testing to human highly-skilled translators. Additionally, rapid commercial advancements in AI and language technology mean that tools like *eTranslation* and *Speech Transcription* need continuous updates to meet the needs of growing multilingualism. Although the EU declares its intention to develop AI language technologies in order to improve the capabilities of its multilingual services, the genuine assessment of the accuracy and reliability of AI translations in EU institutions has not been made available to date. The EU simply offers *eLangTech* not implying any clear path for future developments, updates or bug fixing, which is not attractive for the commercial use in a rapidly developing language technology arena (Varga & Wetzel, 2024).

5.2. Legal and ethical considerations

As AI language technologies advance, they raise proportionally greater ethical and legal concerns. Zaki (2024) identifies a core tension: these technologies offer significant benefits, but also pose risks of surveillance and misuse. This makes clear the need for strong ethical frameworks and effective regulatory oversight. Key challenges relate to fairness, linguistic integrity, and data protection, including the risk of exposing sensitive information when training data contains personal identifiers or confidential communications. This is not a problem that improves with scale or sophistication. On the contrary, as these tools become more integrated into public services and legal institutions, the potential for harm - especially to already minoritised language communities - only increases.

While compliance with frameworks such as the GDPR, AI Act, and Digital Services Act is a legal requirement, it also serves as a baseline for promoting data security and transparency in the deployment of AI-driven language technologies across the EU. However, questions remain around how effectively these frameworks address the specific risks posed by multilingual AI systems, particularly in legal and administrative contexts. Since May 2018, the GDPR has established strict requirements on data processing and storage, ensuring that personal data is handled securely. Article 32 of the GDPR mandates that institutions implement robust safeguards to protect personal data during AI-assisted processing, reinforcing the need for strict oversight when adopting AI-based language tools. Building on this foundation, the AI Act, launched in July 2024, categorises AI applications into four risk levels - ranging from minimum risk to unacceptable risk - based on their potential societal impact. These obligations are particularly relevant to AI-driven language technologies, where large-scale data inputs, including potentially sensitive or multilingual datasets, heighten the risk of privacy breaches and regulatory non-compliance. Although both the GDPR and the AI Act are regional regulations, they exert global implications, particularly in shaping international debates around trans-

parency and accountability in AI systems. As such, they are designed to set standards for responsible AI governance, emphasising the importance of clear, understandable, and accountable decision-making. This is particularly relevant for language models such as ChatGPT, often perceived as ‘black boxes’, because of their opaque, difficult-to-audit decision-making processes. ChatGPT and other AI models operate with limited interpretability, rendering them difficult for users to understand how linguistic outputs are generated. This lack of transparency can hinder error detection, bias mitigation, and overall quality assurance, as a result complicating institutional oversight. Therefore, EU institutions may struggle to ensure accountability for translation errors, partiality and unintended consequences. In response to this challenge, the Digital Services Act (2022) sets forth that service providers must ensure that users are aware of how AI-based services function, reinforcing the necessity for educating EU personnel on the operational mechanics and risks of AI translation tools.

Algorithmic transparency remains the major ethical consideration associated with AI-driven language tools. Its deficiencies increase the probability of algorithmic bias and discrimination within AI models. Although the EU Charter of Fundamental Rights (2016 C 202/2) explicitly underlines the importance of non-discrimination and equal treatment (Article 21), AI models are trained on extensive datasets that contain biases, which may give rise to serious legal and societal implications, such as discriminatory translations, legal misrepresentations, and unequal access to law or policy implementation (Rodrigues, 2020). The use of AI translation tools risks perpetuating systemic inaccuracies, potentially distorting the effective communication of EU laws and policies. Moreover, as indicated in the European Union Agency for Fundamental Rights Report (2022), algorithmic bias can adversely affect fundamental rights, especially for vulnerable groups such as immigrants and minorities. As the Report states algorithmic bias in AI systems may result in both direct discrimination - where reliance on protected characteristics leads to less favourable treatment - and indirect discrimination, where seemingly neutral processes disproportionately disadvantage certain groups.

Yet, another concern in the field of legal and ethical perspectives of AI language technology implementation is the lack of awareness of LLM developers, owners and operators as to the direct implications of algorithms on perpetuating embedded societal biases. They become particularly evident in feedback loops, where biased outputs from the model are fed back into the system, further entrenching existing biases (European Union Agency for Fundamental Rights, 2022). AI systems bear imprints of ingrained human biases through their data-based foundations, generated within historically inequitable social systems. Historical patterns of discrimination become entrenched in social structures, assuming a facade of objectivity and perpetuating harm.

To effectively tackle AI bias, a more nuanced understanding that recognises the overlapping and interdependent nature of marginalised identities and the unique challenges they face must be adopted. Without deliberate and focused efforts to mitigate these biases, there is the risk that they become entrenched in the fabric of AI systems, as a result amplifying existing social inequalities. This is particularly evident in the treatment of minority and under-resourced languages. A 2024 study titled “Do Llamas Work in English? On the Latent Language of Multilingual Transformers” by Wendler, Veselovsky, Monea, and West (2024), investigates how the Llama-2 family of models processes multilingual input. The researchers found evidence that Llama-2 models rely on

English as a latent conceptual space, even when handling prompts in other languages. The researchers indicated three-phase processing “input space”, “concept space” and “output space”, with bias occurring in the “concept space” where semantic understanding takes place. When non-English prompts are given, this intermediate layer often assigns higher probability to the English equivalent of the correct token before finally settling on the correct non-English token in the output space. This phenomenon is especially relevant for tasks such as translation, sentiment analysis, or cultural interpretation, where subtle linguistic and cultural cues matter deeply. Moreover, Lewis’s analysis of AI language learning tools reveals a 30% drop in engagement among minority-language students when cultural bias is present in educational platforms. His work underscores how systemic prejudice in training corpora - often dominated by Western-centric content - can alienate learners and distort linguistic diversity (Lewis, 2025).

Abeba Birhane (2021), founder of the AI Accountability Lab at Trinity College Dublin, has emphasised the ethical risks posed by opaque AI systems. Her audits of large-scale datasets have revealed the presence of misogynistic, racist, and otherwise harmful content, emphasising the dangers of training AI models on uncurated data (Birhane et al. (2022)). Birhane’s research reaffirms the importance of rigorous evaluation, dataset audits, and justice-oriented AI governance frameworks to prevent the reinforcement of societal inequalities. Her studies underscore the critical need for fairness and transparency in development and deployment of AI-driven language models (Birhane, 2021). Birhane criticises the blind trust often placed in algorithmic systems and calls for a relational ethics approach - one that puts in the spotlight the lived experience, historical context, and structural power asymmetries.

5.3. Unequal representation of languages

Although the EU officially upholds language equality among its Member States, the practical deployment of these systems often favours dominant, high-resource languages with extensive digital infrastructure. The reliance on AI tools trained predominantly on data-rich languages not only undermines the EU’s commitment to multilingualism and linguistic diversity, but also perpetuates systemic exclusion of minoritised languages within the digital ecosystem. Moreover, models trained on outdated or culturally insensitive datasets may reinforce stereotypes or misrepresent linguistic identities, contributing to digital homogenisation and erasure (Lee, Montgomery, & Lai, 2024; ?). This imbalance risks marginalising lesser-spoken languages, which typically lack the linguistic corpora, annotated datasets, and technological support necessary for effective integration into AI systems. As a result, speakers of these languages face restricted access to digital discourse and diminished opportunities for socio-political engagement (Thomas, King, & Jones, 2001), as exemplified further in the section.

AI-based language technologies pose unique risks to minority language communities in terms of cultural preservation, representation, and linguistic evolution. These risks are amplified by the data-driven nature of modern neural machine translation (NMT) systems, which require vast amounts of high-quality training data - far more than traditional rule-based or statistical approaches (Bosque-Gil et al., 2022). Designated as under-resourced, many minority languages lack foundational digital tools such as linguistic corpora, electronic dictionaries, and speech recognition systems. In contrast, dominant languages benefit from robust digital ecosystems and sustained in-

vestment, making them more adaptable to AI applications. This disparity exacerbates the digital divide, leaving minority language users vulnerable to exclusion from essential services and technological innovation. Recent research highlights that NMT models struggle with low-resource languages due to data scarcity and modeling limitations (Her & Kruschwitz, 2024).

The issue of underrepresentation extends beyond minority and regional languages and includes ‘smaller’ majority languages - those that are official within a Member State, but receive insufficient technological investment and hence have less digital representation. Recent evaluations of Large Language Models (LLMs) by Thellmann et al. (2024) across European languages reveal notable performance disparities among linguistic families. The study categorises languages into three primary families - Germanic, Romance, and Slavic - and examines their impact on model effectiveness. The findings indicate that LLMs generally perform better on Germanic and Romance languages compared to Slavic ones. This suggests that AI models benefit from stronger linguistic similarities and higher representation in training datasets. Furthermore, the analysis proves that LLM performance depends on both data availability and linguistic complexity. For example, despite Swedish being less represented in training datasets than Polish, it outperformed Polish in all evaluated tasks. Similarly, Romanian - regardless of its existence as one of the least represented languages - still demonstrated competitive performance. These benchmarks highlight the need for AI models to consider linguistic properties and resource distribution when assessing multilingual capabilities.

The above analysis demonstrates that if Polish, a relatively well-represented European language, struggles with AI performance compared to Swedish, the implications are even more concerning for regional and minority languages. These languages face limited training data, fewer digital resources, and lower AI prioritisation, and as a result they are structurally sidelined in the digital ecosystem. As findings of the European Language Equality (ELE) project show, many regional and minority languages such as Scottish Gaelic, Breton, Sámi, and Sardinian rank low in digital readiness, highlighting a severe imbalance in technological support. The example of Scottish Gaelic is particularly telling. As a language with a small digital footprint, Gaelic is often absent from the linguistic soundscape and is rarely heard or seen in public or digital spaces. This lack of visibility contributes to low sociolinguistic vitality, making it harder for learners to see Gaelic as a living, usable language (Birnie, 2024). Similarly, Breton, despite having a significant number of around 200 000 speakers, has historically lacked state support, and its presence in the digital environment is almost non-existent. The Sámi languages spoken across several Nordic countries, remain severely under-resourced in the digital and AI space. Sardinian, one of Italy’s largest minority languages, has a substantial speaker base, but suffers from a lack of standardisation and is almost invisible in the world of AI. The ELE project findings reaffirm that minority languages risk digital extinction, their speakers may be excluded from digital spaces, education and civic participation (“European language equality: a strategic agenda for digital language equality”, 2023).

The threat of digital extinction also concerns sign languages, remaining largely absent from AI-driven translation models. They constitute a distinct challenge as they differ fundamentally from spoken languages in terms of visual-spatial modality. This difference makes data collection and machine training significantly more complex. Unlike spoken languages, where large text corpora are readily available, sign language datasets

are limited, fragmented, and often not designed for machine learning purposes. As highlighted by De Sisto et al. (2022), the development of sign language recognition and translation tools is hindered by data sparsity, non-standardised annotation formats, and limited corpus availability, all of which obstruct the training of effective neural models. De Meulder (2021) and more recently Desai, De Meulder, Hochgesang, Kocab, and Lu (2024) have emphasised that such biases can marginalise Deaf communities by embedding hearing-centric assumptions into AI systems, particularly when Deaf stakeholders are excluded from the design and evaluation processes.

As a positive sign, one should perceive several initiatives that have been launched to integrate artificial intelligence with minority languages in ways that promote ethical and inclusive development. These efforts reflect a growing commitment to linguistic inclusion within the AI sphere. Notably, a number of projects led by Google Research and Microsoft have developed models capable of recognising and transcribing indigenous and minority languages. Google's *Endangered Languages Project* exemplifies this trend by collecting and digitising audio recordings of at-risk languages, thereby contributing to the preservation of linguistic heritage for future generations. In the academic sphere, Bangor University has underscored AI's transformative potential for minority language use, particularly through NLP tools that enhance communication in languages such as Welsh (Prys, 2023). At the regional level, the Nordic Cooperation on AI and Minority Languages, supported by the Nordic Council of Ministers, aims to safeguard linguistic diversity across the Nordic region. This is exemplified by the development of platforms like Borealium, which offers accessible, open-source AI tools for small Nordic languages. On a global scale, the Digital Transformation Initiative supported by the World Bank also explores AI's role in fostering inclusive language technologies. It focuses on the importance of diverse datasets and local language support to bridge the digital divide and ensure equitable access to AI-driven services.

Next to the language divide of majority-minority language, a growing concern is whether LLMs might alter the structural features of underrepresented languages, affecting their grammar, vocabulary, and syntactic evolution. As AI-driven translations rely on pattern recognition and statistical correlations, they may introduce standardised linguistic forms that homogenise distinct expressions (Farina & Lavazza, 2025). Over time, this could erode the organic evolution of minority languages, potentially diminishing regional dialects and linguistic diversity. This concern echoes a historical pattern of linguistic centralisation rooted in the rise of a nation-state concept. In both contexts - nation-state time and AI revolution - standardisation is prioritised over diversity. However, the mechanisms, incentives, and impacts diverge in significant and more insidious ways today. Historically, nation-states suppressed regional dialects and minority languages to forge a unified national identity. The tools of control were state-driven. They included mandatory education in the national language, censorship or outright bans on minority languages, as well as a systemic effort to label diverse languages as "dialects" of a dominant standard, which made minority languages inferior. These acts were politically motivated, designed to facilitate governance, consolidate power, and inflict a sense of national unity. This led to minority languages being neglected and made it harder for languages to be transmitted between generations (Grucza, 2012; Grucza, Kita, & Dębska, 2015).

By contrast, the AI era introduces a global and scalable form of linguistic homogenisation that echoes the historical model of nation-state language control. While nation-states employed overt mechanisms in the pursuit of national unity, AI systems operate invisibly across borders, embedded in everyday tools like search engines, voice assistants, and translation apps. These systems favour standardised grammar and vocabulary, marginalise idioms and regional variations, and reinforce prevailing norms through automated feedback loops. The authority lies with private algorithms optimising for efficiency and commercial scalability. What makes AI-driven tools particularly problematic is their invisibility, as users interact with LLMs daily without recognising the linguistic distortions embedded within them. There is no central authority enforcing change. These are systems that scale silently and globally. As Farina and Lavazza (2025) argue, the threat is not merely exclusion but algorithmic invisibility, where languages disappear not by force, but by omission, making the erosion of minority languages harder to detect and resist.

6. Possible mitigation measures

As AI-driven language technologies become increasingly central to communication, governance, and public services, EU institutions must undertake targeted measures to foster multilingualism as a cornerstone of digital inclusion. Firstly, addressing the challenges posed by NLP tools requires a comprehensive strategy that mitigates algorithmic bias, and promotes equitable access across all EU languages. This includes sustained funding for NLP research in all EU languages, introducing well-defined and transparent assessment mechanisms, and support for the development of a shared, high-quality multilingual training data repository. As existing models remain limited in their capacity to capture the full linguistic and contextual complexity required for equitable NLP, future algorithm development should be systematically paired with bias auditing protocols and context-sensitive evaluation frameworks. Secondly, coping with bias in AI requires not only technical solutions, but also a fundamental re-evaluation of the socio-technical systems within which these technologies are developed and deployed. Embedding ethical considerations and diverse perspectives throughout the AI development process is crucial to create a more equitable and inclusive technological future (Wendehorst, 2022). This commitment must include dedicated support for underrepresented and marginalised languages through ethical design practices, improved data collection methodologies, and sustained institutional investment in minority language resources.

Next, an effective strategy for addressing AI requires a nuanced understanding of the complex and interconnected nature of marginalised identities is presented, this is coupled by mention of the specific challenges faced by minoritised communities. The strategy begins with sourcing information directly from marginalised groups. Their life experiences, languages, and cultural variations must be adequately represented in training data, helping to prevent distortions rooted in outsider perspectives. This shift calls for participatory and culturally conscious AI development, where inclusive data collection and community engagement are central. As Bella, Helm, Koch, and Giunchiglia (2024) argue, linguistic diversity must be understood not merely as the inclusion of additional languages, but as a deeper recognition of the socio-cultural and epistemic frameworks those languages embody. Engaging communities in the design and over-

sight of AI systems fosters a participatory process that aligns technological development with the needs, values, and realities of those most affected. Moreover, responsible AI development depends on collaboration among researchers, linguistic experts, cultural specialists, and community members to ensure that models are built and deployed in a culturally sensitive and ethically sound manner (Soltan et al., in press). This approach not only helps rectify existing inequalities embedded in technology, but also contributes to a more inclusive digital future. Nevertheless, it must be noted that the shortage of experts fluent in both minority languages and AI presents a major obstacle. This skills gap slows innovation and deepens digital exclusion.

Furthermore, to support structural diversity in language, AI technologies applied by EU institutions must be grounded on culturally sensitive data curation. This entails building models on diverse datasets that reflect historical, regional, and evolving language forms. As noted by Gerken (2022), building robust text corpora is essential. Moreover, even if this increases financial investment and technical complexity. These corpora should include slang, dialects, and colloquialisms, even if this raises costs and technical complexity. Without such careful data curation, AI outputs risk reinforcing stereotypes or alienating speakers through culturally inappropriate and reductive language use.

By implementing these mitigation measures, EU institutions can uphold their commitment to multilingualism while ensuring that AI technologies serve all communities equitably. This approach not only addresses current limitations in NLP systems, but also lays the foundation for a more inclusive and culturally aware digital future.

7. Conclusions

The EU's commitment to clear and inclusive communication is fundamental to protecting democratic values and ensuring that all voices are heard and understood in policymaking. Where AI-powered language tools offer transformative potential to improve access to information and facilitate communication across the EU's diverse linguistic landscape, their integration must be approached with careful ethical consideration, rigorous testing, and a steadfast focus on linguistic accuracy to build trust and ensure no one is left behind.

To achieve this, AI translation systems must undergo continuous refinement - particularly in handling complex terminology, regional dialects, and emotional nuance. As noted in the EU institutions' AI strategies, human oversight remains indispensable for verifying translations, maintaining quality, and preserving the rich cultural and linguistic diversity that defines the EU. The goal should not be to prioritise speed or efficiency at the expense of linguistic variety. Instead, multilingual AI must be developed to actively support minority languages, safeguarding cultural heritage, sustaining linguistic identities, and promoting inclusivity in digital spaces. Addressing bias in AI requires a nuanced, intersectional approach that acknowledges the complex realities of marginalized communities. This includes collecting diverse and representative data through direct engagement with these communities to ensure their languages, experiences, and cultural contexts are accurately reflected.

Moreover, actively involving communities in the design and oversight of AI tools is crucial to ensure their needs are properly understood and addressed. This participatory

approach seems to be a good method to reduce biases rooted in external assumptions and fosters a broader range of perspectives. Cross-disciplinary collaboration - uniting social scientists, linguists, ethicists, and AI experts - is equally vital for identifying and mitigating bias, and for developing culturally sensitive algorithms that respect and preserve the identities of minoritized groups. The regular testing and updates of AI systems are necessary to detect and correct harmful patterns before they become deeply embedded. At the same time, establishing clear regulations and ethical guidelines that emphasise transparency, accountability, and inclusion is key to ensuring responsible AI development that actively combats discrimination.

These efforts are not only about improving technology, they are also about reinforcing the democratic principles of inclusion, equality, and participation. When AI language tools are developed without attention to bias, cultural sensitivity, and linguistic diversity, they risk undermining these principles. If AI fails to preserve minority languages and cultures within the EU context, some EU citizens may become increasingly isolated in the digital realm, with decreased ability to advocate for their own rights and shape the policies that affect them. To conclude, the future of multilingual AI tools must be created in such a way that they enhance - not undermine - linguistic diversity and democratic participation of citizens. Only by centering inclusivity, cultural respect, and ethical responsibility, the EU can lead the way in building a digital future where all languages, and all voices, are truly valued.

References

- Bella, G., Helm, P., Koch, G., & Giunchiglia, F. (2024). Tackling language modelling bias in support of linguistic diversity. In *The 2024 ACM Conference on Fairness Accountability and Transparency* (pp. 562–572). Rio de Janeiro Brazil: ACM. Retrieved from <https://dl.acm.org/doi/10.1145/3630106.3658925> doi: 10.1145/3630106.3658925
- Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns*, 2(2), 100205. Retrieved from [https://www.cell.com/patterns/fulltext/S2666-3899\(21\)00015-5](https://www.cell.com/patterns/fulltext/S2666-3899(21)00015-5)
- Birhane, A., Ruane, E., Laurent, T., S. Brown, M., Flowers, J., Ventresque, A., & L. Dancy, C. (2022). The forgotten margins of AI ethics. In *2022 ACM Conference on Fairness Accountability and Transparency* (pp. 948–958). Seoul Republic of Korea: ACM. Retrieved from <https://dl.acm.org/doi/10.1145/3531146.3533157> doi: 10.1145/3531146.3533157
- Birnie, I. (2024). *In search of Gaelic: the social linguistic soundscape as an indicator of minority language use in a bilingual English/ Gaelic Island community*. Retrieved from <https://pure.strath.ac.uk/ws/portalfiles/portal/187932183/Birnei-LL-2024-In-search-of-Gaelic-the-social-linguistic-soundscape.pdf>
- Bosque-Gil, J., Barbu Mititelu, V., Gonçalo-Oliveira, H., Ionov, M., Gracia, J., Rychkova, L., ... Dojchinovski, M. (2022). Balancing the digital presence of languages in and for technological development. A Policy Brief on the Inclusion of Data of Under-resourced Languages into the Linked Data Cloud. *Nexus Linguarum*. Retrieved from https://nexuslinguarum.eu/wp-content/uploads/2022/10/02_Policy-Briefs.pdf
- Carmo, F. D., & Koponen, M. (2024). Chapter 9 Revisers and post-editors: The guardians of quality. In G. Massey, M. Ehrensberger-Dow, & E. Angelone (Eds.),

- Handbook of the Language Industry* (pp. 203–224). De Gruyter. Retrieved from <https://www.degruyter.com/document/doi/10.1515/9783110716047-010/html> doi: 10.1515/9783110716047-010
- De Meulder, M. (2021). Is “good enough” dood enough? Ethical and responsible development of sign language technologies. In *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*.
- Desai, A., De Meulder, M., Hochgesang, J. A., Kocab, A., & Lu, A. X. (2024). *Systemic biases in sign language AI research: A deaf-led call to reevaluate research agendas*. arXiv. Retrieved from <https://arxiv.org/abs/2403.02563> (Version Number: 1) doi: 10.48550/ARXIV.2403.02563
- De Sisto, M., Vandeghinste, V., Egea Gómez, S., De Coster, M., Shterionov, D., & Saggion, H. (2022). Challenges with Sign Language Datasets for Sign Language Recognition and Translation. In N. Calzolari et al. (Eds.), *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (pp. 2478–2487). Marseille, France: European Language Resources Association. Retrieved from <https://aclanthology.org/2022.lrec-1.264/>
- European language equality: a strategic agenda for digital language equality. (2023). In G. Rehm & A. Way (Eds.), *Cognitive Technologies*. Cham: Springer.
- Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, 28(1), 62–70. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S1471772718300277> doi: 10.1016/j.infoandorg.2018.02.005
- Farina, M., & Lavazza, A. (2025). English in LLMs: The role of AI in avoiding cultural homogenization. In P. Hacker (Ed.), *Oxford Intersections: AI in Society* (1st ed.). Oxford University Press Oxford. Retrieved from <https://academic.oup.com/edited-volume/59762/chapter/511416518> doi: 10.1093/9780198945215.003.0140
- Gerken, M. (2022). *Facilitating the implementation of the European Charter for Regional and Minority Languages through artificial intelligence*. Council of Europe.
- Grucza, F. (2012). Język jako narzędzie ideologii państwowej. In J. Porayski-Pomsta (Ed.), *Język w polityce i mediach [Language in Politics and Media]* (pp. 45–58). Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.
- Grucza, F., Kita, M., & Dębska, A. (2015). Zanik języków mniejszościowych jako efekt polityki asymilacyjnej. In *Problemy językowe społeczeństw wielokulturowych [Linguistic Problems of Multicultural Societies]* (pp. 112–126). Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Her, W. H., & Kruschwitz, U. (2024). Investigating Neural Machine Translation for Low-Resource Languages: Using Bavarian as a Case Study. In M. Melero, S. Sakti, & C. Soria (Eds.), (pp. 155–167). Torino, Italia: ELRA and ICCL. Retrieved from <https://aclanthology.org/2024.sigul-1.20/>
- Kamaluddin, M. J., Abqoriyyah, F. H., Khoerul Rasyid, M. W., & Saehu, A. (2024). Accuracy analysis of DeepL: Breakthroughs in machine translation technology. *Journal of English Education Forum*, 4(2), 122–126.
- Kemper, J., & Kolkman, D. (2019). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*, 22(14), 2081–2096. Retrieved from <https://www.tandfonline.com/doi/full/10.1080/1369118X.2018.1477967> doi: 10.1080/1369118X.2018.1477967

- Koponen, M. (2010). Assessing Machine Translation Quality with Error Analysis. *Mikael: Kääntämisen ja tulkkauksen tutkimuksen aikakauslehti*, 4. Retrieved 2025-12-05, from <https://journal.fi/mikael/article/view/129675> doi: 10.61200/mikael.129675
- Lee, M. H., Montgomery, J. M., & Lai, C. K. (2024). Large language models portray socially subordinate groups as more homogeneous, consistent with a bias observed in humans. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1321–1340). Rio de Janeiro, Brazil: ACM. Retrieved from <https://dl.acm.org/doi/10.1145/3630106.3658975> doi: 10.1145/3630106.3658975
- Lewis, A. A. (2025). Unpacking Cultural Bias in AI Language Learning Tools: An analysis of impacts and strategies for inclusion in diverse educational settings. *International Journal of Research and Innovation in Social Science*, IX(I), 1878–1892. Retrieved from <https://rsisinternational.org/journals/ijriss/articles/unpacking-cultural-bias-in-ai-language-learning-tools-an-analysis-of-impacts-and-strategies-for-inclusion-in-diverse-educational-settings/>
- Láncos, P. L. (2021). The role of language technologies in promoting the participation of linguistic minorities in social, political and economic life. *Foreign Policy Review*, 14(2), 73–87. Retrieved from https://kki.hu/wp-content/uploads/2021/12/FPR_2021_02_5_73-87-1.pdf doi: 10.47706/KKIFPR.2021.2.73-87
- Maegaard, B. (1995). Eurotra, history and results. In *Proceedings of Machine Translation Summit V*. Luxembourg, Luxembourg. Retrieved from <https://aclanthology.org/1995.mtsummit-1.5/>
- Paul, R., Carmel, E., & Cobbe, J. (2024). *Handbook on Public Policy and Artificial Intelligence*. Cheltenham, UK.: EE Publishing.
- Prys, D. (2023). Language Report Welsh. In G. Rehm & A. Way (Eds.), *European Language Equality* (pp. 223–226). Cham: Springer.
- Rabari, C., & Storper, M. (2015). The digital skin of cities: Urban theory and research in the age of the sensor and metered city, ubiquitous computing and big data. *Cambridge Journal of Regions, Economy and Society*, 8(1), 27–42. Retrieved from <https://academic.oup.com/cjres/article-lookup/doi/10.1093/cjres/rsu021>
- Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, 4, 100005. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S2666659620300056> doi: 10.1016/j.jrt.2020.100005
- Skorupa-Wulczyńska, A. (2021). *Language rights of the citizens of the European Union*. Berlin: Peter Lang.
- Soltan, A., Rousi, R., Prys, C., O'Brien, S., Ruiz Rosendo, L., & Rainò, P. (in press). Language rights in the human-machine era: A multifaceted inquiry. In D. Sayers, M. Glass, H. Kelly-Holmes, & R. Fuchs (Eds.), *Language in the Human-Machine Era: New technologies and the coming transformation of language*. Boston, MA: MIT University Press.
- Soria, C., Russo, I., Quochi, V., Hicks, D., Gurrutxaga, A., Sarhimaa, A., & Tuomisto, M. (2016). Fostering digital representation of EU regional and minority languages: the Digital Language Diversity Project. In N. Calzolari et al. (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 3256–3260). Portorož, Slovenia: European Language Resources Association (ELRA). Retrieved from <https://aclanthology.org/L16-1518/>

- Theilmann, K., Stadler, B., Fromm, M., Buschhoff, J. S., Jude, A., Barth, F., ... Ali, M. (2024). *Towards cross-lingual LLM evaluation for European languages*. arXiv. Retrieved 2025-12-05, from <https://arxiv.org/abs/2410.08928> (Version Number: 2) doi: 10.48550/ARXIV.2410.08928
- Thomas, N., King, A., & Jones, E. (2001). *Machine Translation and the Internet*. Retrieved from <https://research.aber.ac.uk/en/publications/machine-translation-and-the-internet>
- Varga, L. K., & Wetzel, A. (2024). *Assessing eTranslation and company, the multilingual tools of the EU's digital Europe programme*. Retrieved from <https://www.nimdzi.com/etranslation-elangtech-eu-multilingual-toolset/#elangtech-the-key-multilingual-technologies-of-the-digital-europe-programme>
- Wendehorst, C. (2022). *Bias in Algorithms: Artificial Intelligence and Discrimination* (1st ed.). Retrieved from https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf
- Wendler, C., Veselovsky, V., Monea, G., & West, R. (2024). *Do Llamas work in English? On the latent language of multilingual transformers*. arXiv. Retrieved from <https://arxiv.org/abs/2402.10588> (Version Number: 4) doi: 10.48550/ARXIV.2402.10588
- Williamson, B. (2018). The hidden architecture of higher education: Building a big data infrastructure for the 'smarter university'. *International Journal of Educational Technology in Higher Education*, 15(1), 12. Retrieved from <https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-018-0094-1> doi: 10.1186/s41239-018-0094-1
- Zaki, M. Z. (2024). Bridging linguistic divides: the impact of AI-powered translation systems on communication equity and inclusion. *Journal of Translation and Language Studies*, 5(2), 20–30.
- Zanina-Seck, A., & Groener, C. U. (2025). The Secret Power of Syntax: Improving ChatGPT Translation Quality Through Sentence Constituent Analysis? In H. Degen & S. Ntoa (Eds.), *Artificial Intelligence in HCI* (Vol. 15821, pp. 242–260). Cham: Springer Nature Switzerland. Retrieved from https://link.springer.com/10.1007/978-3-031-93418-6_17 (Series Title: Lecture Notes in Computer Science) doi: 10.1007/978-3-031-93418-6_17

EU legal acts and publications

- Artificial Intelligence Strategy of the Court of Justice of the European Union, Directorate-General for Information, January 2024.
- Charter of Fundamental Rights of the European Union. Official Journal 2016 C 202/2, 7 June 2016.
- Commission Implementing Decision (EU) 2024/458 of 1 February 2024 on setting up the European Digital Infrastructure Consortium for the Alliance for Language Technologies (ALT-EDIC), OJ L, 2024/458.
- Communication from the Commission on the Artificial Intelligence in the European Commission (AI@EC). A strategic vision to foster the development and use of lawful, safe and trustworthy Artificial Intelligence systems in the European Commission, C(2024) 380 final.



- Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions: A New Framework Strategy for Multilingualism (COM(2005) 596).
- Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast), OJ L 172, 26 June 2019.
- European Commission. (2012) *Special Eurobarometer 386. Report. Europeans and their languages*. Fieldwork: February-March 2012. http://data.europa.eu/88u/dataset/s1049_77_1_ebs386.
- European Commission. (2021). *Special Eurobarometer 517. Report. Future of Europe*. Fieldwork: September-October 2021. https://data.europa.eu/data/datasets/s2554_96_1_517_eng?locale=en.
- European Commission (2024). Directorate-General for Digital Services, Fitsilis, F. and Mikros, G., AI-based solutions for legislative drafting in the EU – Summary report, Publications Office of the European Union. <https://data.europa.eu/doi/10.2799/6892931>.
- European Commission. (2024). *Special Eurobarometer 540. Report. Europeans and their languages (2024)*. Fieldwork: September-October 2023. https://data.europa.eu/data/datasets/s2979_100_1_sp540_eng?locale=en.
- European Parliament Resolution of 11 September 2018 on language equality in the digital age, OJ C 433.
- European Parliament (2016). Briefing. Regional and minority languages in the European Union. https://www.europarl.europa.eu/RegData/etudes/BRIE/2016/589794/EPRS_BRI%282016%29589794_EN.pdf.
- European Union Agency for Fundamental Rights (FRA) (2022). *Bias in Algorithms. Artificial Intelligence and Discrimination*. Report. Vienna.
- Proposal for a Regulation on a European Approach to Artificial Intelligence and Amending Certain Union Legislative Acts COM/2021/206, 21 April 2021.
- Regulation No. 1 determining the languages to be used by the European Economic Community, OJ 17, 6 October 1958. Current consolidated version of 1 July 2013.
- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119, 4 May 2016.
- Regulation (EU) 2021/694 of the European Parliament and of the Council of 29 April 2021 establishing the Digital Europe Programme and repealing Decision (EU) 2015/2240, OJ L 166, 11 May 2021.
- Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), OJ L 277, 27 October 2022.
- Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), OJ L 2024/1689, 12 July 2024.

Internet sources

- <https://www.euoparl.europa.eu/translation/en/translation-at-the-european-parliament/technology-to-support-translation>
- <https://www.elra.info/>
- <https://www.ntnu.edu/norwai/nordic-cooperation-to-protect-minority-languages-in-the-age-of-ai>
- <https://research.google/pubs/multimodal-language-identification/>
- https://language-data-space.ec.europa.eu/index_en
- https://language-data-space.ec.europa.eu/related-initiatives/alt-edic_en
- <https://language-tools.ec.europa.eu/>
- <https://iate.europa.eu/>

Can machine translation really help minority languages in Europe?

An analysis with value scenarios

Sergi Alvarez-Vidal ¹ and Maarit Koponen ²

¹Universitat Autònoma de Barcelona, sergi.alvarez@uab.cat

²University of Eastern Finland, maarit.koponen@uef.fi

Abstract

Machine translation (MT) has greatly improved its quality in the last decade and has become nearly omnipresent in all aspects of society. Neural MT (NMT) and, more recently, large language models (LLMs) such as the generative pretrained transformer (GPT) have made translations to many languages easily accessible to all users from any phone or computer. However, most MT models are English-centric and only produce good quality results for those languages with great amounts of data. For minority languages, the challenge is often understood as the scarcity of data. Systemic differences between language communities should be taken into account for MT systems to actually be useful. In this paper, we use value scenarios to imagine the systemic impacts for two languages with differentiated sociolinguistic realities: Catalan and Karelian. The goal is to outline the main challenges and potential harms when considering MT for minority languages and to suggest some general guidelines that can be followed in future research and applications.

Keywords: machine translation; minority languages; low-resource languages; value scenarios.

Resumo

Na última década, a tradução automática (TA) melhorou significativamente a sua qualidade e tornou-se quase omnipresente em todos os aspetos da sociedade. A tradução automática neuronal (TAN) e, mais recentemente, os grandes modelos de linguagem (LLMs), como o transformador pré-treinado generativo (GPT), tornaram as traduções para muitas línguas facilmente acessíveis a todos os utilizadores a partir de qualquer telefone ou computador. No entanto, a maioria dos modelos de TA são centrados no inglês e só produzem resultados de boa qualidade para as línguas com grandes quantidades de dados. Para as línguas minoritárias, o desafio é frequentemente entendido como a escassez de dados.

As diferenças sistêmicas entre as comunidades linguísticas devem ser levadas em consideração para que os sistemas de TA sejam realmente úteis. Neste artigo, recorreremos a uma análise de cenários para compreender os impactos sistêmicos para duas línguas com realidades sociolinguísticas diferenciadas: o catalão e o careliano. O objetivo é delinear os principais desafios e danos potenciais ao considerar a TA para línguas minoritárias e sugerir algumas diretrizes gerais que podem ser seguidas em investigações e aplicações futuras.
Palavras-chave: Tradução Automática, línguas minoritárias, línguas com recursos limitados, análise de cenários.

1. Introduction

For many years, especially after World War I, great efforts were invested into “imposing an ideology of state-centered monolingualism” (Gal, 2017, p. 225), especially for those citizens of new nation-states who, until then, had identified with another language. In the mid-1980s, many political movements advocated for the protection of the “indigenous” languages, especially in the framework of international organizations such as the United Nations (UN) (Muehlebach, 2001). More recently, and in response to international pressure, many of the states changed their policies to fund the conservation and recuperation of these now minoritized languages (Laasko et al., 2016). This policy change has been largely propelled by the European Union (EU), which has linguistic identity as one of its constituent characteristics (Arzoz, 2008).

Europe has a multiplicity of languages, cultures, and peoples. Encouraging the use of these languages means promoting and encouraging diversity and different ways of understanding the world, regardless of the status and number of speakers of each language (Rehm & Way, 2023). According to the UN (2023), 40% of the approximately 7,000 languages that are spoken in the world are endangered to some degree. The *Atlas of the World's Languages in Danger* (Moseley, 2010) defines four levels of vulnerability: vulnerable, definitely endangered, severely endangered and critically endangered.

Language serves an instrumental purpose for accessing information and communication, but it also has an intrinsic function as part of minority identity and culture (Mowbray, 2017, p. 40). Recognizing and supporting minority languages is therefore important to support the autonomy, self-realization and equal dignity of speakers of all languages (De Schutter, 2017, pp. 22-24). However, language diversity has been enormously impacted by globalization, which has facilitated the growth of international languages such as English in the political, social and economic domains (Khan, 2023). Technology has also helped English become a language of prestige and status (Crystal, 2003), and has influenced its spread in media and business. This has largely contributed to its status as a lingua franca.

Language technologies, in particular MT, have also had a notable impact on society, especially since MT quality has been dramatically improved with the emergence of neural MT (NMT) (Wang et al., 2022). While these technologies offer potential benefits, their impact on linguistic diversity and equality should be considered carefully

(Rehm & Way, 2023). Even though communication has never been so easy and accessible to users, technological development risks concentrating powerful resources in a limited number of languages, while marginalising smaller ones. The uses of MT should be planned with consideration for the needs and sociolinguistic situation of the different minority languages. The emerging technological initiatives that develop MT models for these languages (Costa-jussà et al., 2022) should also focus on the target communities and users, in order to understand their needs.

We use value scenarios to incorporate the views of minority language stakeholders to imagine the systemic impacts of MT for two European minority languages. These languages possess different sociolinguistic realities and illustrate some of the potential harms both communities might face when implementing new or improved MT systems. The paper is structured as follows: in Section 2, we contextualize and define minority languages; in Section 3, we describe the most common MT systems and their use with minority languages; in Section 4, we introduce value scenarios and present a use case for two minority languages in Europe with differentiated sociolinguistic situations – Catalan and Karelian. We conclude the paper with Section 5, where we present some general guidelines that should be taken into account when developing responsible and useful MT systems for minority languages.

2. Minority, minoritized, low-resourced languages

The monolingual identification of nation-states, and thus, the negative stigmatization of multilingualism, was propelled after World War I, consolidated after World War II, and lived on for most part of the 20th century. This ideology stressed the importance of having one sole language to guarantee the well-being of all citizens of the same nation-state (Kamusella, 2009), ignoring the multilingual societies in which ideologies were implemented. However, the fall of the Berlin Wall in 1989 escalated tensions and conflicts between ethnic groups in Central and Eastern Europe, leading to growing concern across Europe about ethnic instability. As a consequence, in the beginning of the 1990s, there was a considerable increase in laws and treaties meant to protect minorities, such as the European Charter for Regional or Minority Languages (1992), the Framework Convention for the Protection of National Minorities (1995) and the adoption by the General Assembly of the United Nations of the Declaration on the Rights of Persons Belonging to National or Ethnic, Religious and Linguistic Minorities (1992) (Vieyetz, 2014).

Diverse labels are used to refer to languages from minority groups, mainly stressing some of their characteristics: minoritized, low-resourced. In this paper, we use the term “minority language”, following the definition stated by Cenoz and Gorter (2023). First, a minority is a group that is smaller in relation to another group. It is not a question of absolute numbers, but rather the numerical strength of the group in relation to other groups. Second, the use of “minority language” implies that there is a “majority language”, that is, there is an asymmetric relation between both languages. Minority languages are always minoritized languages in power and status compared to the majority language because they have been disempowered by dominant groups in society. Third, the term “minority language” is widely used in many international organizations, such as the UN and the EU, as well as in international law.

According to the European Charter for Regional or Minority Languages, “regional or minority languages” means languages that are: (a) traditionally used within a given territory of a state by nationals of that state, who form a group numerically smaller than the rest of the state’s population; and (b) different from the official language(s) of that state – this does not include either dialects of the official languages of the state or languages of migrants (Council of Europe, 1992). In Europe, there are a minimum of 69 minority languages that represent 89% of the linguistic diversity within the EU (Videsott, 2023). These include languages such as the Sámi languages, Aragonese, Karelian, Basque, Breton, Catalan, Corsican, Frisian, Friulian, Galician, Ladin, Occitan, Sardinian, and Welsh, among others.

Even though minority languages include the aforementioned characteristics, they also experience different situations, depending on their specific geographical location and social status. We can differentiate between minority languages that are “unique” and those that are also spoken in a kin-state (Extra & Gorter, 2008). Unique minority languages refer to languages spoken in one or more states, but which are not the dominant language in any state. This is the case of Basque, which is spoken in Spain and France, and also of Catalan, which is spoken in France, Spain, Andorra and the island of Sardinia (Italy). However, neither of these two languages has a dominant status in any of these countries. This also includes languages with very few speakers or that are on the verge of extinction, such as Ume Sámi in Sweden, where the youngest speakers are over 60 years of age (Gorter & Cenoz, 2011).

Kin-state minority languages are languages that are minority languages in one state, but are also the official dominant language in another state. This is the case, for example, of Swedish in Finland, where it is a minority language, even though it is the dominant language in Sweden (Potinkara, 2024). Another example is German, the official language in Germany and Austria, yet it is a minority language in countries such as Belgium, the Czech Republic, Denmark, France and Italy (Gorter and Cenoz, 2011). There are also a few minority languages that are official state languages, even though they function to great lengths in the same way as other minority languages. This is the case of Luxembourgish, which is official in Luxembourg even though the dominant languages in this country are French and German. The same happens with Catalan in Andorra, where it is the only official language. However, much of the linguistic space in this country is occupied by Spanish (Pla, 2000).

During the 20th century, there has also been a standardization process for many minority languages, so they could be used in education, the media, and various institutions (Costa, Korne, & Lane, 2017). In this process, one variety of a language becomes widely accepted throughout the linguistic community as a supradialectal norm. Even though this can facilitate the use of this language among all language users and contribute to the social prestige of the language. Recognizing the spoken and written communication practices of certain speakers as standard has placed many other speakers in a position of inferiority (Gal, 2006). These standards for minority languages can be perceived by the speakers of these languages as lacking authority as well as authenticity, and not being able to represent the speakers implied (Woolard, 2008).

3. Machine translation and minority languages

In recent years, MT quality has been greatly improved with the emergence of neural machine translation (NMT) and large language models (LLMs). This has changed the way people perceive multilingual communications. Users interact with MT daily and use it in very different contexts. When it is used for making sense of a text in a language that users do not understand, we talk about assimilation, which is obviously the most frequent use. When MT is used as part of a more complex process to produce a translation that will be later published, we talk about dissemination purposes (Way, 2018). This is usually used by professional translators, but can also be employed by other professionals like journalists (Fité Labaila, 2001, 2007) as one step in the whole translation process.

Early approaches to MT used handcrafted grammatical and lexical rules to translate into the target language (España-Bonet et al., 2011). However, there is a high human cost in formalizing all the linguistic rules. Moreover, rules are difficult to transfer from one domain to another, and from one language to another (Wang et al., 2022). With the availability of bilingual corpora, methods using these corpora as the base for MT training became the dominant approach from the mid-2000s onward (Kenny, 2018). These statistical MT approaches often improved translation quality, but lacked fluency, and it had important reordering problems for distant language pairs (Wang et al., 2022). However, rule-based MT (RBMT) systems continue to be used for minority languages that have less available training data and produce poor results when using corpus-based approaches (Bayatli et al., 2018; Islam, Anik, & Islam, 2022; Kasthuri & Kumar, 2014; F. Pirinen & Wiecheteck, 2022; T. A. Pirinen, 2019; Sghaier & Zrigui, 2020). Well-known examples of RBMT are Lucy LT1 (Alonso & Thurmair, 2003) and the open-source Aperi-tium (Forcada et al., 2011; Khanna et al., 2021).

NMT is a corpus-based MT, which is trained on huge amounts of corpora, usually formed by pairs of source and target language segments (Forcada, 2017). NMT needs large quantities of data and uses deep neural networks. The first NMT models consist of an encoder and a decoder which are jointly trained (Sutskever, Vinyals, & Le, 2014). The encoder encodes the source language sentences into a sequence of vectors, which are the hidden representations of the source tokens, and also a meaning vector. Then the decoder generates the sequence of tokens for the target language, considering the meaning vector from the encoder. The transformer architecture, which is the one mostly used in current systems, follows mainly the encoder-decoder model with attention passed from encoder to decoder (Vaswani et al., 2017). This employs a self-attention mechanism that allows the encoder and decoder to account for every word included in the entire input sequence. As it yields high quality results, both in automatic and human evaluations (Bentivogli et al., 2016; Castillo et al., 2017; Klubička, Toral, & Sánchez-Cartagena, 2017), with more fluent outputs (Bentivogli et al., 2018), it has been quickly adopted by all industry sectors.

In recent years, the improvements in LLMs have brought new developments in MT and many other natural language processing tasks (Kim et al., 2021). Generative pre-trained transformers (GPTs) are decoder-only LLMs pretrained on large-scale unsupervised text data and are fine-tuned for different natural language processing tasks, which include MT (Brown et al., 2020). GPT models are also based on transformer architec-

ture (as NMTs), but with some differences. They are decoder-only models that use the same parameters to process the context and source as a single input for generating the next output. GPT models are basically trained on large quantities of monolingual data, mostly obtained by scraping texts from the Internet. Since the main language of online digital texts is English, this establishes a bias towards English in the dataset. These high data requirements are an important barrier for most of the smaller language communities, because many of these state-of-the-art tools cannot be applied to them due to the lack of data (Joshi et al., 2019). In fact, even though there is linguistic diversity around the world that includes more than 7,000 languages, few of them are represented in the current language technologies (Haddow et al., 2022; Joshi et al., 2020).

4. Value scenarios

The impact of new technologies is shaped not only by technical design, but also by how individuals and communities engage with them (Friedman, Kahn, Borning, & Huldgren, 2013). In the case of minority languages, this engagement is deeply embedded in sociolinguistic, cultural, and political realities. MT is not merely a translation tool; it is part of a broader socio-technical system that can reinforce or disrupt language ideologies, power relations, and community practices.

To explore this complexity, we apply the value scenarios method, a speculative and ethical design approach that combines narrative-based scenario building with an explicit focus on human values (Nathan, Klasnja, & Friedman, 2007). Rather than predicting the future, value scenarios offer a structured way to explore how technologies might interact with complex social realities, including unintended consequences and ethical tensions. They help uncover latent assumptions in design and policy choices, as well as encourage reflection on how values such as fairness, representation, and autonomy are affected across different communities. They build on scenario-based design (Carroll, 1999), incorporate imaginative and subversive elements from design noir (Dunne & Raby, 2001), and anchor values such as linguistic justice, identity, and agency. Value scenarios help structure reflections along five key dimensions: stakeholders (both direct users and those indirectly affected), pervasiveness (how widespread and embedded the technology becomes), time (typically near- to mid-term), systemic effects (cascading consequences on language practices and social structures), and value implications (tensions between community needs and the goals or assumptions of technological design).

This approach has been previously applied by Haroutunian (2022), who constructed value scenarios based on stakeholder interviews for the Armenian-English MT context. His work highlighted the importance of participatory development and the need to assess which language technologies might best serve the interests of low-resourced communities beyond MT alone. Inspired by this precedent, our study expands the approach to a comparative perspective by focusing on two minority languages in Europe with contrasting sociolinguistic contexts: Catalan and Karelian.

To ground the analysis, we designed a speculative example of MT technology called MIGHTYMT: a fictional, advanced language technology envisioned to support real-time translation for minority languages. MIGHTYMT is a LLM-based assistant integrated across personal and public devices. It is capable of translating both written and spoken interactions in real time, using voice, text, or multimodal inputs. It operates seamlessly

across messaging apps, search engines, browsers, and voice assistants. Users can query it conversationally or use it passively to translate content displayed on screens, or spoken aloud in physical spaces.

MIGHTYMT supports high-quality, bidirectional translation between minority languages (such as Catalan and Karelian) and major world languages. It uses context-aware translation modules adapted to dialectal variation and allows for user feedback to train personal language profiles. It is designed to be intuitive, customizable, and accessible, with offline functionalities for rural and under-connected regions. Unlike traditional MT systems, it includes modules to explain translation decisions and detect sociolinguistic sensitivities (for example, formality, regional variation, gendered language), making it more transparent and adaptable for speakers of diverse linguistic backgrounds.

This imagined tool serves as the basis for our two value scenarios—one for Catalan and one for Karelian—which explore both the promises and challenges associated with integrating such a system into daily life. These scenarios examine functional, social, and ethical dimensions to better understand how advanced language technologies may interact with complex language ecologies and community values.

4.1. The Catalan Value Scenario

Catalan is a standardized language with around 10 million speakers (Generalitat de Catalunya, 2024), yet it continues to face marginalization in key domains. It has a marginal status in the Balearic Islands and Aragon, is co-official in Catalonia, and is the sole official language in Andorra. It is also spoken in the Sardinian city of Alghero, in some villages in the South of France, and in frontier areas of Murcia. Despite this wide geographical spread, Catalan remains a minoritized language, particularly in the justice and health systems as well as in state-level services, where Spanish is often the dominant language of operation (Pla, 2000).

Since the early 1980s, the education system and public media have played a crucial role in maintaining and promoting Catalan, while civic organizations such as Softcatalà, Òmnium Cultural, and Plataforma per la Llengua have been instrumental in advocating for language rights. In the field of MT, various systems now support Catalan. Google Translate incorporated it in 2008, and Softcatalà maintains several MT engines covering different language combinations to and from Catalan (Briva Iglesias, 2022). Catalan is also increasingly included in LLMs by default, primarily due to its structural similarity to other Romance languages such as Spanish, French, and Italian. However, the resulting outputs are not always genuine or contextually appropriate, often reflecting limited training data or linguistic bias.

In addition to these tools, several research projects are dedicated to the development of language technologies for Catalan. One example is TAN-IBE, which focuses on building NMT models and resources for the Iberian Romance languages (Oliver et al., 2023). Another is Projecte AINA (Montserrat, 2023), a publicly funded initiative coordinated by the Barcelona Supercomputing Center (BSC) in collaboration with the Catalan government. AINA develops large-scale language resources and open-source AI tools—including curated speech and text corpora and Catalan-adapted LLMs—with the overarching goal of ensuring the language's digital viability.

In our value scenario, we consider the deployment of MIGHTYMT and its potential impact on different stakeholder groups, particularly young speakers and healthcare professionals. For young Catalan speakers who are increasingly shifting toward Spanish in digital spaces (Idescat, 2023), MIGHTYMT could lower linguistic barriers and promote the use of Catalan in informal online communication. This may reinforce linguistic confidence and help normalize Catalan in everyday digital practices. However, overreliance on automated output could contribute to linguistic deskilling, limiting opportunities for active language production. If the system consistently reproduces standardized forms, speakers from regions such as the Balearic Islands or Alghero might feel alienated, reinforcing perceptions of exclusion.

For healthcare professionals, MIGHTYMT could support communication with Catalan-speaking patients, particularly older adults or individuals with cognitive or speech-related impairments, who may find it easier to express themselves in their first language. This could promote more inclusive and patient-centred care. However, the use of automated tools in clinical contexts raises legal and ethical concerns. If MIGHTYMT processes sensitive health data without adequate safeguards, it may violate patients' privacy rights as defined under the General Data Protection Regulation (GDPR), designating health data as a special category requiring explicit consent and enhanced protection (European Union, 2016, Art. 9). Article 32 further mandates that data controllers implement technical and organisational measures to ensure the confidentiality, integrity, and security of such data, especially in high-risk domains like healthcare.

Systemic effects of deploying MIGHTYMT could include increased visibility and usability of Catalan across sectors, enhanced digital literacy, and a stronger presence in multilingual public communication, reinforcing existing efforts such as the parallel publication of press content in Catalan and Spanish (Fité Labaila, 2007). However, the tool could also entrench dominant language norms, reduce active language engagement, and obscure the need for structural reforms. Public institutions may interpret the presence of MT systems as a substitute for guaranteeing linguistic rights through human-mediated services, thus shifting responsibility for language access onto users and technology.

If developed with community input, MIGHTYMT could also address the longstanding invisibilization of dialectal varieties—such as those spoken in the Balearic Islands or the Valencian Country—which are often flattened in mainstream MT systems. Allowing user-controlled variation and transparent dialect tagging could promote a more inclusive representation of Catalan's internal diversity.

From a legal perspective, integrating MIGHTYMT into public services entails obligations regarding linguistic equality as established by regional autonomy statutes. Inaccurate or biased outputs—particularly in sensitive domains like healthcare or justice—could constitute violations of linguistic rights and require formal accountability mechanisms. When third-party providers are involved, public administrations must ensure compliance with data protection, transparency, and quality standards. In the absence of a clear regulatory framework for AI-based multilingual communication, these risks may go unaddressed. This scenario underscores the tension between enhanced access and the preservation of linguistic agency. While tools like MIGHTYMT may contribute to the normalization of Catalan in digital environments, their design and

governance must be inclusive and community-driven to avoid reproducing structural inequalities.

4.2. The Karelian Value Scenario

Karelian is a critically endangered minority language spoken in Finland and Russia. It is not recognised as an official language in either country, although some laws defining its status exist in the Republic of Karelia in the Russian Federation. The Finnish government included provisions in its language policy programme in 2022 to strengthen the legal status of Karelian in Finland (Valtioneuvosto, 2022). The number of speakers is currently decreasing even though there have been different attempts to revitalize the language, and a planned language revitalization has been in place since the end of the 1980s (Moshnikov, 2022).

Historically, Karelian was a territorial language. As a result of World War II, the majority of speakers of Karelian lost their traditional homelands and were resettled in various parts of Finland. Linguistically, Karelian belongs to the Finnic branch of the Uralic language family, or more precisely, to its eastern Finnic subgroup, which can be distinguished geographically from the southern and western groups, and on the basis of their common history. Karelian does not have a standard variety and is divided into two main dialects: Olonets (or Livvi) Karelian, and Karelian Proper. The latter consists of Viena (North) Karelian and South Karelian (Uusitupa, Koivisto, & Palander, 2017). Based on census data, the number of Karelian speakers in Russia was about 25,000 in 2010 (Moshnikov, 2022, p. 278). In Finland, the number of fluent speakers is estimated to be around 11,000, with a further 20,000 who speak or understand Karelian at least to some extent (Sarhimaa, 2016, p. 3). Some schools in Russian Karelia teach Karelian, and university-level courses are offered at Petrozavodsk State University in Russia and at the University of Eastern Finland, in Finland (Moshnikov, 2022). According to a recent survey conducted by Moshnikov (2022), only 17% of the 161 respondents reported they speak Karelian daily, 18% at least once a week, and 16% more rarely. Additionally, 40% of the surveyed users reported that while they do not speak Karelian, they can understand written or spoken Karelian (34%), or write in Karelian (6%). In the digital sphere, 40% of the respondents search for information in Karelian daily, and an additional 46% at least once a week (Moshnikov, 2022, p. 285). An earlier survey by Salonen (2017, p. 9) found that 73.7% of the nearly 2,000 respondents used Karelian online, mainly for reading, and to a lesser extent also writing content on websites, blogs or discussion forums.

None of the Karelian varieties are currently covered by commercial MT systems such as Google Translate or large-scale initiatives such as Meta's 'No Language Left Behind' (Costa-jussà et al., 2022). Some work has been done on Livvi Karelian as part of the SMUGRI-MT NMT system for low-resource Finnic languages (Purason et al., 2024; Yankovskaya et al., 2023), and a rule-based approach (T. A. Pirinen, 2019). OpenAI's ChatGPT and Anthropic's Claude appear to generate translations that the models claim to be to and from Karelian. However, establishing the output quality these tools produce is outside of the scope of this article.

For our MIGHTYMT value scenario, stakeholders firstly include members of the Karelian speaker community. The tool could help them search for information or answer questions in their language. Particularly for speakers who do not currently use Karelian in an active way, this could give more opportunities to engage with the language.

Another group of stakeholders would be people learning Karelian as a new language, or hoping to improve their skills. This chatbot offers the possibility of interacting in Karelian without having the time constraints of formal learning environments. Speech translation capacities could be useful particularly for those who currently indicate that they do not actively speak although they understand written or spoken Karelian. They could ask for information or talk about current issues to practice their Karelian, and even ask for corrections if they think their level of language is not good enough. It could also enable Karelian speakers to use their own language in interactions with the majority language speakers.

MIGHTYMT could also offer support for translation activities that have been identified as important for language revitalization, such as creating vocabulary and educational materials (cf. Kuusi, Kolehmainen, & Riionheimo, 2017). One important source for Karelian language content is the Finnish national broadcaster Yle, which currently produces news only in one dialect (Livvi Karelian), although some other newspapers publish articles in different variants (Moshnikov, 2022). As a tool for journalists — and others producing materials — MIGHTYMT could also help produce more content in the other varieties, if these are included in the tool.

The value scenario presumes that use of the technology would be widespread. In practice, however, the interest of the speaker community to adopt the technology may be limited. Karelian speakers tend to be bilingual, which means that they also speak a majority language, and commonly resort to that in digital environments, in a phenomenon which has been labeled as digital diglossia (Lozano Sañudo, 2023). Hence, Karelian speakers would need to be convinced this new product can accommodate their needs in Karelian before using it in many contexts. The lack of a standard variety also presents a challenge: users may not identify with the variant used by the chatbot and could respond negatively to MIGHTYMT not using their own language variety, which could lead to a low sense of reliability in the chatbot. As Moshnikov (2022) notes, however, the language community is accustomed to multidialectal content in newspapers, for example, so dialectal variation might not be considered an obstacle by Karelian speakers.

Systemic effects of the use of MIGHTYMT could in this way include support for the revitalization efforts of the community and wider use of Karelian both online and off. More widespread use would also improve the visibility of Karelian among majority language speakers. Currently, even the existence of Karelian is not well-known among Finnish people, and it is commonly perceived as a regional dialect of Finnish. By enabling wider use of Karelian in society, MIGHTYMT could raise awareness of the language and strengthen its status.

However, widespread use of the technology could also lead to adverse systemic effects. Firstly, large quantities of machine-generated outputs could start to direct standardization of the language on the terms of the technology, rather than the language community. Secondly, focusing on technological solutions could end up directing societal resources and support away from the speaker community's revitalisation efforts. Thirdly, quality and reliability of the output is naturally a major concern, particularly in high-stakes scenarios like healthcare, legal settings and other public services (see also Guerberof-Arenas & Moorkens, 2023). Similarly to the Catalan case, implementation in these scenarios could have adverse implications for the Karelian speakers' equal rights.

5. Implications for responsible development of MT for minority languages

The envisioned MIGHTYMT system highlights both the potential and pitfalls of deploying MT for minority languages. While Catalan benefits from a relatively high level of institutional support and digital infrastructure, Karelian exemplifies the challenges faced by critically endangered languages with minimal technical resources. The contrast underscores the need for differentiated strategies when developing and implementing MT for diverse language communities.

While Karelian faces a scarcity of digital resources and lacks stable language norms, Catalan is comparatively well-represented in existing language technologies. However, technological inclusion alone does not ensure linguistic empowerment. In both cases, systems like MIGHTYMT risk displacing human-driven language use. For Catalan, overreliance on automated tools may discourage active written production, particularly among young users, and reinforce a narrow standard at the expense of dialectal variation. For Karelian, the risks are amplified in the context of language endangerment: widespread MT use could impose artificial norms on a language with limited standardization and potentially low-quality outputs due to scarce training data. Moreover, mass generation of machine-produced content may saturate digital spaces with synthetic language, displacing and marginalizing authentic, community-produced discourse and expressions (see Aranberri & Iñurrieta, 2024). This risk is particularly acute for endangered languages like Karelian, where vitality depends on active use, and MT quality is likely to remain low due to data scarcity. Some scholars have therefore argued that MT should not be offered for such languages until its quality and consequences are better understood (Bowker, 2021; F. Pirinen & Wiechetek, 2022).

Community acceptance is another crucial factor shaping the role of MT in minority language contexts. Acceptability depends not only on output quality, but also on the perceived purpose and framing of the technology. Research has shown that users tend to accept MT more readily for informational tasks than for contexts involving linguistic identity or visibility (Bowker, 2009; Bowker & Buitrago-Ciro, 2015). In Catalonia, for example, some criticism of MT in public services stems from concerns that it might be used to avoid institutional commitments to Catalan, particularly in sensitive domains where language choice is symbolic of rights and recognition (Baumgarten & Cornellà-Detrell, 2019; Nurminen & Koponen, 2020).

These reflections underscore that MT is never a neutral tool. As Bird (2024) argues, language technologies should be designed to support human agency, not replace it. For minority languages, this means embedding MT within community-led strategies that prioritize speaker autonomy, linguistic diversity, and long-term sustainability. Following this principle, development of MT for minority languages should not be carried out simply for the sake of technological development. Rather, the development should be guided by the needs and priorities of the language communities (cf. Riemland, 2023), for example, through participatory research practices. Tools like MIGHTYMT may support access and visibility, but only if they are transparent, participatory, and sensitive to the values and goals of the communities they aim to serve. The central role of the language community as end-users of the technology is reflected in the discussion in our article. In future work, it would be useful to address also the other parts of the

“MT stakeholder ecosystem”: companies, governmental agencies, language professionals, researchers and developers (cf. Guerberof-Arenas & Moorkens, 2023, p. 115).

In conclusion, the value scenarios of Catalan and Karelian demonstrate that MT can contribute to the sustainability of minority languages under specific conditions, but may also deepen existing asymmetries if deployed without care. Ensuring that MT supports rather than undermines these languages requires ongoing community involvement, legal and ethical oversight, and a commitment to addressing not only technological gaps, but the structural conditions in which these languages are used.

Acknowledgment

The work of Maarit Koponen was supported by the Strategic Research Council (SRC) established within the Research Council of Finland; funding agreements 352577 (University of Eastern Finland) and 352557 (DECA consortium coordinator University of Helsinki).

References

- Alonso, J. A., & Thurmair, G. (2003). The compendium translator systems. In *Proceedings of the Ninth Machine Translation Summit*.
- Aranberri, N., & Iñurrieta, U. (2024). When minoritized languages encounter MT: Perceptions and expectations of the Basque community. *The Journal of Specialised Translation*, 41, 179–205. doi: 10.26034/cm.jostrans.2024.4718
- Arzoz, X. (2008). Preface. In X. Arzoz (Ed.), *Respecting Linguistic Diversity in the European Union* (Vol. 2, pp. vii–viii). Amsterdam: John Benjamins Publishing Company. doi: 10.1075/wlp.2.01arz
- Baumgarten, S., & Cornellà-Detrell, J. (2019). Translation and the economies of power. In *Translation and the Global Spaces of Power* (pp. 11–26). Bristol: Multilingual Matters.
- Bayatli, S., Kurnaz, S., Salimzianov, I., Washington, J. N., & Tyers, F. M. (2018). Rule-based machine translation from Kazakh to Turkish. In *Proceedings of the 21st Annual Conference of the European Association for Machine Translation: 28-30 May 2018* (pp. 49–58). Universitat d’Alacant, Alacant, Spain. Retrieved from <https://dialnet.unirioja.es/servlet/articulo?codigo=6474388>
- Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2016). Neural versus phrase-based machine translation quality: A case study. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 257–267). Austin, Texas: Association for Computational Linguistics. Retrieved from <http://aclweb.org/anthology/D16-1025> doi: 10.18653/v1/D16-1025
- Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2018). Neural versus phrase-based MT quality: An in-depth analysis on English–German and English–French. *Computer Speech & Language*, 49, 52–70. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S0885230817301079> doi: 10.1016/j.csl.2017.11.004
- Bird, S. (2024). Must NLP be extractive? In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Retrieved from <https://aclanthology.org/2024.acl-long.797/>

- Bowker, L. (2009). Can Machine Translation meet the needs of official language minority communities in Canada? A recipient evaluation. *Linguistica Antverpiensia New Series-Themes in Translation Studies*, 8, 123–155.
- Bowker, L. (2021). Translation technology and ethics. In K. Koskinen & N. K. Pokorn (Eds.), *The Routledge Handbook of Translation and Ethics* (1st ed., pp. 262–278). London/New York: Routledge. Retrieved from <https://www.taylorfrancis.com/books/9781000288988/chapters/10.4324/9781003127970-20> doi: 10.4324/9781003127970-20
- Bowker, L., & Buitrago Ciro, J. (2015). Investigating the usefulness of machine translation for newcomers at the public library. *Translation and Interpreting Studies*, 10(2), 165–186. doi: 10.1075/tis.10.2.01bow
- Briva Iglesias, V. (2022). English-Catalan neural machine translation: State-of-the-art technology, quality, and productivity. *Revista Tradumàtica: Traducció i Tecnologies de La Informació i La Comunicació*, 20, 149–176.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... Amodei, D. (2020). *Language Models are Few-Shot Learners*. arXiv. (Version Number: 4) doi: 10.48550/ARXIV.2005.14165
- Carrol, J. (1999). Five reasons for scenario-based design. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers* (p. 11). Maui, HI, USA. Retrieved from <http://ieeexplore.ieee.org/document/772890/> doi: 10.1109/HICSS.1999.772890
- Castillo, S., Moorkens, J., Gaspari, F., Sennrich, R., Sосoni, V., Georgakopoulou, Y., ... Gialama, M. (2017, September). A comparative quality evaluation of PBSMT and NMT using professional translators. In *Proceedings of Machine Translation Summit XVI: Research Track*. Retrieved from <https://aclanthology.org/2017.mtsummit-papers.10>
- Cenoz, J., & Gorter, D. (2023). Second language acquisition and minority languages: An introduction. In J. Cenoz & D. Gorter (Eds.), *The Minority Language as a Second Language* (pp. 1–15). New York: Routledge.
- Costa, J., Korne, H. D., & Lane, P. (2017). Standardising Minority Languages: Reinventing Peripheral Languages in the 21st Century. In P. Lane, J. Costa, & H. De Korne (Eds.), *Standardizing Minority Languages* (pp. 1–23). New York: Routledge.
- Costa-jussà, M. R., Cross, J., Çelebi, O., Elbayad, M., Heafield, K., Heffernan, K., ... Wang, J. (2022). *No language left behind: Scaling human-centered Machine Translation*. arXiv. (arXiv:2207.04672 [cs]) doi: 10.48550/arXiv.2207.04672
- Council of Europe. (1992). *European Charter for Regional or Minority Languages*. Retrieved from <https://rm.coe.int/1680695175>
- Crystal, D. (2003). *English as a Global Language* (2nd ed.). Cambridge University Press. doi: 10.1017/CBO9780511486999
- De Schutter, H. (2017). Translational justice: Between equality and privation. In G. G. Núñez & R. Meylaerts (Eds.), *Translation and Public Policy: Interdisciplinary Perspectives and Case Studies* (1st ed., pp. 15–31). London: Routledge. Retrieved from <https://www.taylorfrancis.com/books/9781315521763> doi: 10.4324/9781315521770
- Dunne, A., & Raby, F. (2001). *Design Noir: The Secret Life of Electronic Objects*. Springer Science & Business Media.

- España-Bonet, C., Labaka, G., Díaz de Ilarraza, A., & Màrquez, L. (2011). Hybrid Machine Translation Guided by a Rule-Based System. In *Proceedings of Machine Translation Summit XIII: Papers*. Xiamen, China. Retrieved from <https://aclanthology.org/2011.mtsummit-papers.63>
- European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Official Journal of the European Union*, L119, 1–88. Retrieved from <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
- Extra, G., & Gorter, D. (2008). The constellation of languages in Europe: An inclusive approach. In G. Extra & D. Gorter (Eds.), *Multilingual Europe: Facts and Policies* (pp. 3–61). Berlin/New York: Mouton de Gruyter.
- Fité Labaila, R. (2001). La traducció automàtica aplicada a la premsa escrita. El cas d'El Periódico en català. *Treballs de Comunicació*, 21–25.
- Fité Labaila, R. (2007). Cas d'integració de la TA: El Periódico. *Tradumàtica: Traducció i Tecnologies de La Informació i La Comunicació*, 4. Retrieved from <https://raco.cat/index.php/Tradumatica/article/view/56010>
- Forcada, M. L. (2017). Making sense of neural machine translation. *Translation Spaces*, 6(2), 291–309. doi: 10.1075/ts.6.2.06for
- Forcada, M. L., Ginestí-Rosell, M., Nordfalk, J., O'Regan, J., Ortiz-Rojas, S., Pérez-Ortiz, J. A., ... Tyers, F. M. (2011). Apertium: A free/open-source platform for rule-based machine translation. *Machine Translation*, 25(2), 127–144. doi: 10.1007/s10590-011-9090-0
- Friedman, B., Kahn, P. H., Borning, A., & Huldgtren, A. (2013). Value sensitive design and information systems. In N. Doorn, D. Schuurbiens, I. Van De Poel, & M. E. Gorman (Eds.), *Early engagement and new technologies: Opening up the laboratory* (Vol. 16, pp. 55–95). Dordrecht: Springer Netherlands. Retrieved from https://link.springer.com/10.1007/978-94-007-7844-3_4 (Series Title: Philosophy of Engineering and Technology) doi: 10.1007/978-94-007-7844-3_4
- Gal, S. (2006). Contradictions of standard language in Europe: Implications for the study of practices and publics*. *Social Anthropology*, 14(2), 163–181. doi: 10.1111/j.1469-8676.2006.tb00032.x
- Gal, S. (2017). Visions and Revisions of Minority Languages: Standardization and Its Dilemmas. In P. Lane, J. Costa, & H. De Korne (Eds.), *Standardizing Minority Languages* (pp. 222–242). New York: Routledge.
- Generalitat de Catalunya. (2024). *Situació del català, de l'occità aranès i de la llengua de signes catalana i principals línies d'intervenció en política lingüística durant la legislatura*.
- Gorter, D., & Cenoz, J. (2011). Multilingual education for European minority languages: The Basque Country and Friesland. *International Review of Education / Internationale Zeitschrift Für Erziehungswissenschaft / Revue Internationale de l'Education*, 57(5/6).
- 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). Cham: Springer.
- Haddow, B., Bawden, R., Barone, A. V. M., Helcl, J., & Birch, A. (2022). Survey of low-

- resource Machine Translation. *Computational Linguistics*, 48(3), 673–732. doi: 10.1162/coli_a_00446
- Haroutunian, L. (2022). Ethical considerations for low-resourced Machine Translation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop* (pp. 44–54). Dublin, Ireland: Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-srw.5
- Idescat. (2023). *Enquesta d'usos lingüístics de la població*. Retrieved from <https://www.idescat.cat/pub/?id=eulp&n=7195>
- Islam, M. A., Anik, M. S. H., & Islam, A. B. M. A. A. (2022). An enhanced RBMT: When RBMT outperforms modern data-driven translators. *IETE Technical Review*, 39(6), 1473–1484. doi: 10.1080/02564602.2022.2026828
- Joshi, P., Barnes, C., Santy, S., Khanuja, S., Shah, S., Srinivasan, A., ... Bali, K. (2019). Unsurpassed challenges of building and deploying language technologies for low resource language communities. In *Proceedings of the 16th International Conference on Natural Language Processing* (pp. 211–219). NLP Association of India. Retrieved from <https://aclanthology.org/2019.icon-1.25>
- Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 6282–6293). Online: Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.560
- Kamusella, T. (2009). *The Politics of Language and Nationalism in Modern Central Europe*. London: Palgrave Macmillan UK. Retrieved from <http://link.springer.com/10.1057/9780230583474> doi: 10.1057/9780230583474
- Kasthuri, M., & Kumar, S. B. R. (2014). Rule Based Machine Translation System from English to Tamil. In *2014 World Congress on Computing and Communication Technologies* (pp. 158–163). Trichirappalli, India: IEEE. doi: 10.1109/WCCCT.2014.50
- Kenny, D. (2018). Machine translation. In P. Rawling & P. Wilson (Eds.), *The Routledge Handbook of Translation and Philosophy* (pp. 428–445). Abingdon/New York: Routledge.
- Khan, A. (2023). Language and globalization: A critical study on language, culture, and power dynamics. *International Journal of English Learning & Teaching Skills*, 5(3), 3412–3423. doi: 10.15864/ijelts.5305
- Khanna, T., Washington, J. N., Tyers, F. M., Bayatlı, S., Swanson, D. G., Pirinen, T. A., ... Alòs I Font, H. (2021). Recent advances in Apertium, a free/open-source rule-based machine translation platform for low-resource languages. *Machine Translation*, 35(4), 475–502. doi: 10.1007/s10590-021-09260-6
- Kim, Y. J., Awan, A. A., Muzio, A., Salinas, A. F. C., Lu, L., Hendy, A., ... Awadalla, H. H. (2021). *Scalable and Efficient MoE Training for Multitask Multilingual Models*. arXiv. (Version Number: 1) doi: 10.48550/ARXIV.2109.10465
- Klubička, F., Toral, A., & Sánchez-Cartagena, V. M. (2017). Fine-grained human evaluation of neural versus phrase-based Machine Translation. *The Prague Bulletin of Mathematical Linguistics*, 108(1), 121–132. doi: 10.1515/pralin-2017-0014
- Kuusi, P., Kolehmainen, L., & Riionheimo, H. (2017). Introduction: Multiple roles of translation in the context of minority languages and revitalisation. *Trans-Kom*, 10(2), 138–163.

- Laakso, J., Sarhimaa, A., Åkermark, S. S., & Toivanen, R. (2016). *Towards Openly Multilingual Policies and Practices: Assessing Minority Language Maintenance Across Europe*. Bristol: Multilingual Matters.
- Lozano Sañudo, B. (2023). The role of language technologies and MT in fostering multilingualism by contributing to the revitalization of minor and endangered languages. In F. M. Sirignano & R. M. Roig (Eds.), *Educación, investigación e innovación en la red, 2023* (pp. 131–149). Aula Magna. Retrieved from <https://dialnet.unirioja.es/servlet/articulo?codigo=9527870>
- Montserrat, M. V. (2023). El projecte AINA, la IA i les tecnologies del llenguatge. *Terminàlia*, 27, 80–84.
- Moseley, C. (2010). *Atlas of the world's languages in danger*. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000187026>
- Moshnikov, I. (2022). The use of Karelian language online: Current trends and challenges. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics*, 13(2). doi: 10.12697/jeful.2022.13.2.09
- Mowbray, J. (2017). Translation as marginalisation? International law, translation and the status of linguistic minorities. In G. G. Núñez & R. Meylaerts (Eds.), *Translation and Public Policy: Interdisciplinary Perspectives and Case Studies* (1st ed., pp. 32–57). Routledge. doi: 10.4324/9781315521770
- Muehlebach, A. (2001). "Making Place" at the United Nations: Indigenous Cultural Politics at the U.N. Working Group on Indigenous Populations. *Cultural Anthropology*, 16(3), 415–448. doi: 10.1525/can.2001.16.3.415
- Nathan, L. P., Klasnja, P. V., & Friedman, B. (2007). Value scenarios: A technique for envisioning systemic effects of new technologies. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems* (pp. 2585–2590). San Jose CA USA: ACM. doi: 10.1145/1240866.1241046
- Nurminen, M., & Koponen, M. (2020). Machine translation and fair access to information. *Translation Spaces*, 9(1), 150–169. doi: 10.1075/ts.00025.nur
- Oliver, A., Vázquez, M., Coll-Florit, M., Álvarez, S., Suárez, V., Aventín-Boya, C., ... Pardo, A. (2023, June). TAN-IBE: Neural Machine Translation for the romance languages of the Iberian Peninsula. In M. Nurminen et al. (Eds.), *Proceedings of the 24th Annual Conference of the European Association for Machine Translation* (pp. 495–496). Tampere, Finland: European Association for Machine Translation. Retrieved from <https://aclanthology.org/2023.eamt-1.50/>
- Pirinen, F., & Wiecheteck, L. (2022). Building an Extremely Low Resource Language to High Resource Language Machine Translation System from Scratch. In R. Schaefer, X. Bai, M. Stede, & T. Zesch (Eds.), *Proceedings of the 18th Conference on Natural Language Processing (KONVENS 2022)* (pp. 150–155). Potsdam, Germany: KONVENS 2022 Organizers. Retrieved from <https://aclanthology.org/2022.konvens-1.18/>
- Pirinen, T. A. (2019). Workflows for kickstarting RBMT in virtually No-Resource Situation. In A. Karakanta et al. (Eds.), *Proceedings of the 2nd Workshop on Technologies for MT of Low Resource Languages* (pp. 11–16). Dublin, Ireland: European Association for Machine Translation. Retrieved from <https://aclanthology.org/W19-6803/>
- Pla, J. T. i. (2000). La Llengua catalana a Andorra: Problemes i perspective. *Treballs de*

- sociolingüística catalana*, 165–167.
- Potinkara, N. (2024). Finland-Swedes and the concept of national minorities in Sweden. *Ethnopolitics*, 23(1), 59–75. doi: 10.1080/17449057.2022.2108596
- Purason, T., Ivanov, A., Yankovskaya, L., & Fishel, M. (2024). SMUGRI-MT - Machine Translation System for Low-Resource Finno-Ugric Languages. In C. Scarton et al. (Eds.), *Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 2)* (pp. 31–32). Sheffield, UK: European Association for Machine Translation (EAMT). Retrieved from <https://aclanthology.org/2024.eamt-2.16/>
- Rehm, G., & Way, A. (Eds.). (2023). *European Language Equality: A Strategic Agenda for Digital Language Equality*. Springer.
- Salonen, T. (2017). Karelian – a digital language? In *Reports on Digital Language Diversity in Europe*. Retrieved from http://www.dldp.eu/sites/default/files/documents/DLDP_Karelian-Report.pdf
- Sarhimaa, A. (2016). *Karelian in Finland: ELDIA Case-Specific Report*. Retrieved from <https://phaidra.univie.ac.at/detail/o:471733>
- Sghaier, M. A., & Zrigui, M. (2020). Rule-based Machine Translation from Tunisian dialect to modern standard Arabic. *Procedia Computer Science*, 176, 310–319. doi: 10.1016/j.procs.2020.08.033
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). *Sequence to Sequence Learning with Neural Networks*. arXiv. (Version Number: 3) doi: 10.48550/ARXIV.1409.3215
- United Nations. (2023). *Why Indigenous languages matter: The international decade on Indigenous languages 2022–2032*. Retrieved from www.un.org/development/desa/dpad/wpcontent/uploads/sites/45/publication/PB151.pdf
- Uusitupa, M., Koivisto, V., & Palander, M. (2017). Border Karelian dialects and the terminology of Karelian borderlands' language varieties. , 121.
- Valtioneuvosto. (2022). Kielipoliittinen ohjelma: Valtioneuvoston periaatepäätös (No. 51; Valtioneuvoston julkaisu). *Valtioneuvosto*. Retrieved from <http://urn.fi/URN:ISBN:978-952-383-645-7>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). *Attention Is All You Need*. arXiv. (Version Number: 7) doi: 10.48550/ARXIV.1706.03762
- Videsott, P. (2023). Minority and minoritised languages as part of the European linguistic and cultural diversity | Think Tank | European Parliament. *European Parliament*. Retrieved from [https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU\(2023\)751273](https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2023)751273)
- Vieyetz, E. J. R. (2014). Cultural Diversities and Human Rights: History, Minorities, Pluralization. *The Age of Human Rights Journal*, 3(Article 3).
- Wang, H., Wu, H., He, Z., Huang, L., & Church, K. W. (2022). Progress in Machine Translation. *Engineering*, 18, 143–153. doi: 10.1016/j.eng.2021.03.023
- Way, A. (2018). Quality Expectations of Machine Translation. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation Quality Assessment* (Vol. 1, pp. 159–178). Cham: Springer International Publishing. Retrieved 2025-12-02, from http://link.springer.com/10.1007/978-3-319-91241-7_8 (Series Title: Machine Translation: Technologies and Applications) doi: 10.1007/978-3-319-91241-7_8
- Woolard, K. A. (2008). Language and identity choice in Catalonia: The interplay of

contrasting ideologies of linguistic authority. In K. Süselbeck, U. Mühlischlegel, & P. Masson (Eds.), *Lengua, nación e identidad: La regulación del plurilingüismo en España y América Latina*. Iberoamericana Vervuert. Retrieved from <https://dialnet.unirioja.es/servlet/articulo?codigo=5371500>

Yankovskaya, L., Tars, M., Tättar, A., & Fishel, M. (2023). Machine Translation for Low-resource Finno-Ugric Languages. In T. Alumäe & M. Fishel (Eds.), *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)* (pp. 762–771). Tórshavn, Faroe Islands: University of Tartu Library. Retrieved from <https://aclanthology.org/2023.nodalida-1.77/>

When Machine-generated Mistranslation on Social Media Becomes Misinformation: Risks to Users, Corporate Responsibility, and Legal Implications

Khetam Al Sharou ¹

¹Dublin City University & University of Damascus, khetam.alsharou@dcu.ie

Abstract

Machine-generated mistranslations on social media can result in misinformation, with potential major consequences for users, especially marginalised communities. As Machine Translation (MT) is increasingly used to access on-line content, its errors often go unnoticed by users lacking knowledge of the source language. MT inaccuracies can distort meaning, contribute to misinformation, and reinforce digital inequality. Social media has become a main source of information. The unchecked use of machine-generated content introduces vulnerabilities, especially in politically and culturally-sensitive contexts. Through real-world case studies and empirical analysis, this work shows how mistranslations can distort meaning and cause misinformation. It highlights the ethical responsibility of tech companies and service providers to ensure accuracy and transparency while mitigating the risks that arise when MT errors lead to real-world harm. It further assesses how regulatory frameworks, including the EU's Digital Services Act and other similar frameworks, can help address these challenges. This work advocates for responsible MT integration, equitable information access, and stronger corporate and regulatory accountability in combating MT-driven misinformation.

Keywords: *Machine Translation, social media, misinformation, language rights, user experience and corporate responsibility.*

Resumo

Os erros de tradução automática (TA) nas redes sociais podem resultar em desinformação, com consequências potencialmente graves para os utilizadores, especialmente para comunidades marginalizadas. À medida que a TA é cada vez mais utilizada para aceder a conteúdos online, os seus erros passam frequentemente despercebidos aos utilizadores que não têm conhecimento da língua de origem.

As imprecisões da TA podem distorcer o significado, contribuir para a desinformação e reforçar a desigualdade digital. As redes sociais tornaram-se numa das principais fontes de informação. A utilização descontrolada de conteúdos gerados por máquinas potencia algumas vulnerabilidades, especialmente em contextos politicamente e culturalmente sensíveis. Através de estudos de caso reais e análises empíricas, este trabalho mostra como os erros de tradução podem distorcer o significado e gerar desinformação. Destacamos a responsabilidade ética das empresas de tecnologia e dos prestadores de serviços em garantir a precisão e a transparência, mitigando os riscos que surgem quando os erros de TA causam danos no mundo real. Além disso, avaliamos como os quadros regulamentares, incluindo a Lei dos Serviços Digitais da UE e outros quadros semelhantes, podem ajudar a enfrentar esses desafios. Este trabalho defende a integração responsável da TA, o acesso equitativo à informação e uma maior responsabilidade corporativa e regulamentar no combate à desinformação impulsionada pela TA.

Palavras-chave: Tradução Automática, redes sociais, desinformação, direitos linguísticos, experiência do utilizador e responsabilidade corporativa.

1. Introduction

False or misleading information, including both misinformation and disinformation, has always been a societal concern, significantly exacerbated by the widespread use of social media platforms as vehicles to disseminate information and misinformation/disinformation. The term misinformation/disinformation began to increasingly appear in scholarly discussions of social media following the 2016 US election, referring to information that is untrue and can mislead people (Comito, 2023, 2024). UNESCO differentiates between misinformation as “information that is false but not created with the intention of causing harm” and disinformation as “information that is false and deliberately created to harm a person, social group, organisation or country” (Ireton & Posetti, 2018). The main concern is the ease with which information, verified or otherwise, is posted, shared and consumed on social media platforms. Statistics indicate that, as of February 2025, 5.24 billion people (63.9% of the global population) were using social media (Petrosyan, 2025). Misinformation/disinformation on social media is thought to have influenced major events worldwide, including those related to health and politics (see, Allcott & Gentzkow 2017; Baker, 2022; Martin et al., 2022)

International organisations have taken steps to raise awareness of the harm that such activities can cause. For example, UNESCO has pointed out that misinformation/disinformation can spread at a large scale in different shapes and forms, fuelled by new technology especially social media platforms (Ireton & Posetti, 2018). Governments have also introduced emergency laws to prevent the spread of false information online. Regulations to reduce online harms and ensure platform accountability, for example: the Digital Services Act (DSA) - Regulation (EU) 2022/2065 (European Union, 2022); the EU 2022 Code of Practice on Disinformation (European Commission, n.d.); the UK Online

Safety Act 2023 (Department for Science, Innovation & Technology, 2025); and Ireland's Online Safety and Media Regulation Act 2022 (Department of Culture, Communications & Sport, 2022), represent important steps in online content moderation and regulation.

The EU's DSA, fully applicable since February 2024, represents a significant effort to regulate online platforms by imposing stricter transparency and accountability measures, particularly on what it terms as Very Large Online Platforms (VLOPs), which have more than 45 million monthly users within the EU, such as Facebook, Instagram, X, YouTube and LinkedIn. Both the UK's Online Safety Act 2023 (OSA) and Ireland's Online Safety and Media Regulation Act 2022 aim to regulate online activities and address illegal and harmful content. At the EU level, the Code of Practice on Disinformation, revised in 2022, was recognised as a Code of Conduct under the DSA in 2025, transitioning from voluntary compliance to enforceable standards and reinforcing the EU's commitment to establishing a more transparent, accountable, and trustworthy digital information environment (European Commission, n.d.).

These frameworks, while important, primarily focus on the source text, overlooking Machine Translation¹(MT) and related language technologies as potential contributors of information distortion (Al Sharou & Moorkens, 2024). Access to information is recognised as a human right in both international and national legal frameworks. It is protected under Article 19 of the International Covenant on Civil and Political Rights (ICCPR), which affirms every individual's right to seek, receive, and share information (UN, 2022). This right is considered essential for societal participation and for promoting equality (Nurminen & Koponen, 2020). In today's globalised world, access to multilingual information is often supported by advanced technology. Many social media giants now have MT as an additional service on their platforms; for example, Meta has developed its own automated translation systems (models) to increase user engagement and make information available across several languages on its platforms (Costa-jussà et al., 2022; Vincent, 2022).

While MT made access to information in other languages possible, its outputs, when not reviewed by humans, can have critical errors with major impact on the end-users. This particularly holds true for speakers of non-dominant languages who rely on MT to access information or engage in online discussions (see Al Sharou & Specia, 2022; Lee & Qian, 2022; Martindale, 2024, among others). Current MT systems are capable of producing fluent content and some users seem to accept it as correct even when it is not (Lee & Qian, 2022; Martindale, 2024). Therefore, it is crucial when deploying these MT systems to consider language rights, as per the ICCPR, and the risks to these vulnerable communities. These communities, especially those who are not fluent in the original language of a text (or do not have access to translation professionals) are particularly vulnerable to the impact of such mistranslation (Martindale, 2024). This is particularly important as liability for harm caused by MT mistranslations is still an unresolved issue, leaving users in a legally ambiguous position (Canfora & Ottmann, 2020). Therefore, the ethical responsibility of MT providers should extend beyond merely fixing technical

¹In this case study, we refer to Machine Translation (MT) as an integrated feature within social media platforms, either developed by the platform providers or as a third-party service. The discussion may also apply to other automatic translation systems, including tasks performed by large language models (LLMs), and those offered as standalone tools and apps, available as free-to-use services (paid services are not included).

issues to include proactive safeguards which ensure equal and fair treatment for all users. Without these principles in practice, the goal of inclusivity on global platforms remains unfulfilled.

Translation as a human activity has been tied to issues including: ideology manipulation in translation; how ideology can affect translators' linguistic choices and the impact that it has on the receivers of the translations; and how translations can shape receivers' views of the world (Calzada-Pérez, 2014; Farhan, 2017; Wang & Feng, 2018). For example, bias in news reporting can be displayed through the use of certain stylistic features that can influence how meaning is constructed, forming people's understanding of other cultures and places (see discussion in Riggs, 2019a, 2019b). While MT is not a human activity, its outputs are derived from human communications and can still carry the political biases or culturally preconceived ideas of humans. MT systems are trained on data and algorithms that are created and developed by humans, and research has revealed that the data used to train these machines contain and reproduce bias (see Vanmassenhove et al., 2021). These biases can subtly or overtly influence the output of the system. For example, instances of gender bias, such as the over-representation of masculine pronouns, or even racism, have been observed (Fan et al., 2019; Ghosh & Caliskan, 2023). This study will demonstrate how linguistic bias can be embedded into automated systems in ways that cause real harm, leading to the marginalisation and dehumanisation of entire groups of people. It highlights a broader concern that, without adequate oversight, these technologies can replicate and even amplify the prejudices of the societies that design them.

In this work, we look at MT-driven misinformation from the perspective of how it changes the *inferred intent* and sentiment of the original text. This results in the reader receiving a distorted message that could trigger a certain action or change their opinion about a place, person, action, or event. The current study tackles the challenge of disseminating distorted information through MT by directly utilising real-life examples of mistranslations, and as such, will address social media and MT providers' policy goals on misinformation, automation, and information sharing. The research seeks to highlight the risks associated with these technologies and discuss the ethical and social implications, along with corporate responsibility regarding their development and use. This work answers the following research questions (RQs):

- RQ1.** To what extent a machine can distort facts, fuel misinformation and reflect inherent biases?
- RQ2** What are the potential consequences on the end-user and service providers?
- RQ2.1** Do these consequences, if any, vary depending on the political and cultural contexts, the affected groups and status of the users?
- RQ3.** What are the ethical implications of deploying and using MT systems in sensitive contexts?
- RQ4.** What obligations do tech companies and service providers have in mitigating harms, ensuring transparency, and safeguarding users from machine-generated misinformation?

To fulfil the objectives of this study and answer these RQs, four real-life case studies of mistranslations, generated by the auto-translate feature on Meta's Facebook and Instagram platforms, are analysed in depth. It may be argued that reported cases of

mistranslation in social media might be considered as rare, meaning that the **likelihood** of a mistranslation problem causing severe consequences is low given the amount of machine-translated social media content that is consumed on a regular basis. However, these severe consequences do occur, and cannot be ignored, making it an issue deserving of our attention. In this study, misinformation can result from low MT performance, where the machine may, or may not, have necessarily been programmed to produce incorrect information. It is an attempt to highlight the issue and a call for more action to ensure a safer online experience for all users

First, we provide an overview of prior research on the relationship between social media, misinformation, and automated translation. Second, we explain Meta's policy on misinformation and automated translation. Third, we present and analyse four case studies of mistranslations caused by the auto-translate feature on Meta's Facebook and Instagram, discussing its impact on users and the ethical considerations associated with their use and development. The analysis highlights the responsibility of social media platforms to ensure the accuracy of its automated translation systems and explores how existing regulatory frameworks can support equal access to information and the protection of language rights in the digital age. The paper concludes with recommendations for future action and proposes directions for further research.

2. Social Media, Misinformation and Machine Translation

Research has discussed how the use of social media could influence users' political and social attitudes, resulting in the adoption of false or misleading ideas that can sometimes lead to tangible consequences in both behaviour and political discourse (Martin et al., 2019, Pedace; 2021; Saaida, 2023). For example, by investigating findings from multiple sources, Allcott and Gentzkow (2017)'s study demonstrates how misinformation played a significant role in shaping political outcomes during the 2016 US election, benefiting Donald Trump. Their research reveals that false stories favouring Trump's campaign were shared more widely than factual ones, and that many users believed the misleading content they came across (see also Martin et al., 2019). Baker (2022) indicates that social media influenced users' attitudes during Covid by allowing wellness influencers who promote alternative health practices to spread misinformation by presenting themselves as relatable and authentic. The platforms enabled them to challenge mainstream institutions such as governmental research and public health institutions while promoting alternative beliefs and conspiracy theories, encouraging distrust in science and resistance to public health measures (Baker, 2022).

Social media is a communication environment that is extremely diverse. With automatic translation in the form of MT now being very common on these platforms, users consume media produced in different places and in a range of languages. Social media platforms provide raw, unchecked MT output, leaving users with possibly distorted translations, raising ethical concerns due to their potential to generate inaccurate information (Al Sharou & Specia, 2022). Mistranslations by MT can even introduce toxicity or offensive language that was not in the original text, as shown in Al Sharou and Specia (2022). In their study, they looked into the use of MT for user-generated content and the type of errors MT can generate when dealing with such content. Their examples show how some human-generated errors in the content - such as grammatical or orthographical errors - and writing techniques like shortening, word lengthening or dis-

guising words using symbols, can lead to inaccurate translations (see also Al Sharou et al., 2021's study that presents a taxonomy of such non-standard features of the text). These translations may cause misunderstandings or misinterpretations for users who rely entirely on the translation to access the meaning or message of the original content. For example, the sentence "your killing the f***ing planet" in English was translated by Google Translate into Arabic as "may the damn planet kill you", reversing the sentiment and introducing violence (Al Sharou & Specia, 2022). Users can face serious consequences if they believe and act on these inaccurate translations (Al Sharou and Specia, 2022). However, some of these errors may even be difficult to detect, presenting significant risks to users (Canfora & Ottmann, 2020; see also Lee & Qian, 2022). This difficulty in detecting errors is a critical liability concern especially considering that users who are not fluent in the source language can be at higher risk of being misled by mistranslations (Lee & Qian, 2022; Martindale, 2024).

Nevertheless, there is little research on evidence that MT spreads misinformation. To date, only two studies focus on MT-driven misinformation. Narayanan's study examines the spread of misinformation on Facebook due to mistranslations of news headlines from English to Tamil (2022). For example, the English headline, "Trump For Rushing To Defend Tomi Lahren While Ignoring Real Victims," was translated into Tamil as "Trump was for coming early to protect Tomi Lahren ignoring the real victims." The translated version changed the meaning, leading to a misleading interpretation that suggested Trump's actions were more neutral or positive than the original intended critical tone. The study found that twenty percent of general and ambiguous translated headlines, and thirty percent of sarcastic and domain-specific headlines, did not accurately reflect the meaning of the original source. Lee and Qian's (2022) study examines MT-driven misinformation by analysing four English-to-Chinese MT texts, finding that most misinformation was caused by polysemy/named-entity errors and non-equivalent idiomatic expressions. One example from Lee and Qian (2022, pp.538-539), categorised as a semantic error, shows how MT can seriously change the meaning of a sentence: the original English sentence, "...but that we cannot burden people with a carbon tax or a gasoline tax to slow global warming," was mistranslated into Chinese as "...but that we cannot impose a carbon tax or a gasoline tax to lessen people's burden to slow global warming." This completely reverses the intended meaning, creating misleading information.

3. Meta's policy on Misinformation and Machine Translation

Meta is the parent company of Facebook and Instagram, two of the most popular social media platforms globally, with billions of active users interacting and sharing content (see, Meta, n.d.a). Meta's auto-translate feature supports this global communication in several languages. Facebook automatically translates posts based on the user's default language settings (Facebook, n.d). For example, if a user has Arabic set as their default language on Facebook, posts in English will be automatically displayed in Arabic. The auto-translate feature is also available for comments. Furthermore, users have the option to view the content in its original language by clicking on the "See Original" option. Users can rate the translation, but no option to report incorrect translations is given. Instagram provides a "See Translation" option to translate a post's caption, comments, and profiles (Instagram, n.d.). In terms of its effort to develop their MT systems and allow people to communicate in their native languages across all its platforms, Meta launched

the "No Language Left Behind (NLLB)" initiative (Costa-jussà et al., 2022; Meta, n.d.b). By utilising cutting-edge modelling techniques, the project aimed to achieve high translation accuracy and make MT accessible for major languages as well as low-resource languages, including dialects. As of its latest report, NLLB covers over 200 languages with 150 low-resource languages included (see, Figure 1)².

Real-World Applications		Experience the Tech		NLLB Innovations		Research Milestones	
Arabic (Iraqi/Mesopotamian)	Welsh	Italian	North Azerbaijani	Finnish	Kyrgyz		
Arabic (Yemen)	Danish	Javanese	Bashkir	Fon	Kimbundu		
Arabic (Tunisia)	German	Japanese	Bambara	Scottish Gaelic	Konga		
Afrikaans	French	Kabyle	Balinese	Irish	Korean		
Arabic (Jordan)	Friulian	Kachin Jinghpao	Belarusian	Galician	Kurdish (Kurmanji)		
Akan	Fulfulde	Kamba	Bemba	Guarani	Lao		
Amharic	Dinka(Rek)	Kannada	Bengali	Gujarati	Latvian (Standard)		
Arabic (Lebanon)	Dyula	Kashmiri (Arabic script)	Bhojpuri	Haitian Creole	Ligurian		
Arabic (MSA)	Dzongkha	Kashmiri (Devanagari script)	Banjar (Latin script)	Hausa	Limburgish		
Arabic (Modern Standard Arabic)	Greek	Georgian	Tibetan	Hebrew	Lingala		
Arabic (Saudi Arabia)	English	Kanuri (Arabic script)	Bosnian	Hindi	Lithuanian		
Arabic (Morocco)	Esperanto	Kanuri (Latin script)	Buginese	Chhattisgarhi	Lombard		
Arabic (Egypt)	Estonian	Kazakh	Bulgarian	Croatian	Latgalian		
Assamese	Basque	Kabiye	Catalan	Hungarian	Luxembourgish		
Asturian	Ewe	Thai	Cebuano	Armenian	Luba-Kasai		
Awadhi	Faroese	Khmer	Czech	Igobo	Ganda		
Aymara	Iranian Persian	Kikuyu	Chokwe	Ilocano	Dholuo		
			Central Kurdish	Indonesian	Mizo		

Figure 1. Full list of supported languages by Meta (Meta, n.d.b)

According to Meta’s Code of Conduct for Virtual Experiences (Meta, n.d.c) and Hateful Conduct (Meta, n.d.d), its mission is to empower individuals and give them the opportunity to express themselves freely by creating an online communication environment that is welcoming, non-intimidating, and that does not encourage offline harm or violence. Meta’s Community Standards define what is and is not permitted on its various platforms, including Facebook and Instagram (Meta, n.d.e). Meta claims that these standards “apply to everyone, all around the world and to all types of content, including AI-generated content” (Meta, n.d.e).

Meta’s misinformation policy aims to provide flexible guidelines to manage false content while balancing free speech with the need to prevent harm (Meta, n.d.f). It recognises that truth is not static but can evolve, and that misinformation is often context-dependent; accordingly, it avoids a blanket ban, acknowledging the nuanced challenges of verification and enforcement (Meta, n.d.f). In 2025, however, Meta stopped its Third-Party Fact-Checking Program in the US, which relied on external organisations to verify content, replacing it with Community Notes, a user-driven system where individuals attach contextual notes to potentially misleading posts (Kaplan, 2025). The move was framed by CEO Mark Zuckerberg as reaffirming Meta’s “commitment to free expression” (Kaplan, 2025). However, analysts cautioned that without professional fact-checking, hate speech and disinformation could go undetected, affecting marginalised communities (e.g., ethnic, religious groups), who already faced disproportionate online targeting (Booth, 2025). The European Commission also voiced concern over Meta’s de-

²Please note that Figure 1 has Modern Standard Arabic (MSA) mentioned twice. It was not possible to confirm with Meta whether it is a mistake or they refer to different variants of Arabic language.

cision, stressing that any content moderation system implemented within the EU must undergo a formal risk assessment, demonstrate its effectiveness, and comply with the Digital Services Act (Tsimitakis, 2025).

Nonetheless, Meta's policies do not include explicit guidelines directed at users specifically on how to report cases of misinformation, generated by its auto-translate feature, nor are there warnings about its tool's limitations. The motivation behind this study is the real-life examples of mistranslations that were caused by the auto-translate feature on Facebook and Instagram. These mistranslations demonstrate how Meta's auto-translate feature can generate errors that are critical, and can significantly distort the intended meaning of content, sometimes with serious and harmful consequences for users.

4. Misinformation Through Mistranslation: Real-life Examples

Four real-life case studies of mistranslations, produced by the auto-translate feature on Meta's Facebook and Instagram platforms, were selected and analysed in depth. We focused specifically on cases that:

- Drew significant attention (reported by major news outlets),
- Directly affected communities (e.g., causing political, social, or cultural harm), and
- Prompted a public reaction from Meta (e.g., corrections, official apologies).

These examples will be discussed in terms of four topics (issues), providing answers to the four aforementioned RQs:

1. **Accuracy of Translation (RQ1):** Exploring how the MT system misinterpreted the original message and the extent to which the translation deviated from the intended meaning.
2. **Impact on Users (RQ2 & RQ2.1):** Covering consequences of the errors, such as embarrassment, violence, and/or harm, on the end-users.
3. **Ethical Considerations (RQ3):** Including ethical implications of deploying and using MT systems in sensitive contexts, showing how the content, and/or the sensitive nature of the context, has made the mistranslation particularly dangerous and damaging.
4. **Corporate Responsibility (RQ4):** looking at the role of companies such as Meta in ensuring the accuracy and reliability of their MT systems and their responsibility towards addressing issues when they arise, implementing safeguards to protect users from harm caused by incorrect translations.

First, the four case studies will be presented briefly before embarking on discussing them in light of these four topics.

4.1. Example One (EX1)

In a recent incident in 2023, some Palestinian Instagram users' profiles were incorrectly translated, with the term "terrorist" being inserted into their bios. This issue affected profiles that contained the word "Palestinian," the Palestinian flag emoji, and the Arabic phrase "alhamdulillah" ("Praise be to God"). When clicking "See Translation", the English translation read as: "Praise be to God, Palestinian terrorists are fighting for

their freedom” (McMahon & Tidy, 2023). As reported by 404media, the user who initially posted about the issue on Tiktok tried to translate the phrase (“Praise be to God”) alone without including the word Palestinian or the flag emoji. The result was a correct translation, “Thank God”, that did not include the word “terrorist” (Cole, 2023). Instagram’s auto translate feature not only provided a distorted translation but also attached a politically-charged label “terrorist” to a group of people. This led to widespread backlash from users about the platform’s biases. The issue was fixed by Meta, who attributed this error to a “technical bug” and apologised (Taylor, 2023). However, critics called for greater clarity around how its translation system operates. As reported by the Guardian, Fahad Ali, the secretary of Electronic Frontiers Australia and a Palestinian based in Sydney, said:

There is a real concern about these digital biases creeping in and we need to know where that is stemming from. Is it stemming from the level of automation? Is it stemming from an issue with a training set? Is it stemming from the human factor in these tools? There is no clarity on that. (Taylor, 2023)

4.2. Example Two (EX2)

The Guardian reported that a Palestinian man was wrongly arrested due to a critical error in a machine-translated post in Facebook, which led to a belief that he was planning an attack (Hern, 2017). The man, a construction worker in the West Bank, posted “yusbihukum” on his Facebook profile, which translates to “good morning.” However, Facebook’s MT system mistakenly translated this phrase into “hurt them” in English and “attack them” in Hebrew. The man was detained for several hours, and only after questioning did the police release him. Notably, the post had not even been reviewed by any Arabic-speaking officer before action being taken based purely on the machine-generated translation (Hern, 2017). This is concerning and shows that MT systems can inadvertently create hostility and aggression. The BBC (2017) mentioned that the post was deleted without adding any further details as to why. Nor did the article reflect on the user’s experience of the incident. In sensitive environments, the ability of MT systems to distort meaning and generate potentially harmful content is a significant issue. Facebook apologised for the mistake, stating that it was working to address the issue, though it also acknowledged that such errors can occur, even with improvements to their MT systems.

4.3. Example Three (EX3)

In 2020, Facebook’s auto-translate feature mistranslated Chinese President Xi Jinping’s name from Burmese to English as “Mr. Shithole”, an offensive and inappropriate term that was described by the media as “embarrassing” (Serrano, 2020; The BBC, 2020). The mistranslation appeared in Facebook posts shared on the official Facebook page of Myanmar’s State Counsellor, Aung San Suu Kyi (McPherson, 2020; The BBC, 2020). In response, Facebook issued an apology, attributing the mistake to a “technical issue” (Serrano, 2020). A spokesman for Facebook said: “this should not have happened and we are taking steps to ensure it doesn’t happen again,” and “We sincerely apologize for the offense this has caused” (McPherson, 2020). Facebook admitted that the Burmese-into-English database on which the MT system was trained did not contain Xi Jinping’s name and that led to the mistranslation (McPherson, 2020; Serrano, 2020). After that,

the English translation function did not appear to be working on the Burmese posts of official Facebook pages belonging to Ms. Suu Kyi and the Myanmar government, suggesting ongoing technical issues (McPherson, 2020; The BBC, 2020). This controversy highlights the political and technical limitations of MT, especially when it comes to incidents involving political figures and cultural sensitivities, leading to potentially harmful consequences.

4.4. Example Four (EX4)

In 2020, Facebook faced another significant criticism when its auto-translate feature mistranslated a headline about the live broadcast of the ceremony celebrating the King of Thailand's birthday posted on the Facebook page of the Thai Public Broadcasting Service (Thai PBS) (Marking, 2020). The original English headline, which was meant to convey respect and celebration, was automatically turned into an offensive phrase. "King's birthday" was changed to "King's Memorial Day," in the Thai translation (Soponvijit, 2022), a term usually associated with mourning someone's death. The mistranslated headline caused a public outcry and serious legal and social repercussions in Thailand, with some viewers demanding the resignation of Thai PBS executives (Soponvijit, 2022). The Royal Thai Police initiated a formal investigation into the mistranslation incident following an official complaint submitted by the Thai PBS to the cybercrime division (Marking, 2020). This example highlights the legal responsibility platforms bear when a mistranslation results in spreading the wrong information. Facebook issued a formal apology, and temporarily deactivated its auto-translate feature on both Facebook and Instagram, promising to enhance the quality of their translation (Marking, 2020). This case also underscores the significant impact that translation errors can have on both legal liability and public trust. According to Marking (2020), Thailand's strict *lèse-majesté* laws make this case particularly significant, as even unintentional mistranslations can lead to imprisonment. Moreover, the fact that the Thai press avoided publishing the exact mistranslation further illustrates the severity and sensitivity of the issue (Marking, 2020). This aligns with the argument that MT errors can have serious real-world impacts, especially in high-stakes situations.

The following discussion examines the central issues brought to light by these examples, namely: the accuracy of MT, its potential impact on users, ethical considerations, and corporate responsibility. It offers a deeper analysis of how machine-generated mistranslations can influence and misinform end-users and the broader implications of such errors.

5. Discussion - Accuracy of Machine Translation (MT)

The presented examples of mistranslations highlight the risks of relying on MT in politically and socially sensitive contexts. They show how MT has the potential to mislead and produce translations that misrepresent the original message and can even criminalise individuals. The mistranslations occurred within the context of user-generated content, which is often more challenging for MT systems to handle. User-generated content is informal, frequently uses slang, and can include context-specific language that automated systems often struggle to interpret correctly (Al Sharou et al., 2021). The broader implication is that automated systems, which may work well in some contexts, still fail to comprehend the complexities of real-world communication and are not

fully ready to deal with such content and context (see discussion in Al Sharou & Specia, 2022).

In some cases (EX1 and EX2), mistranslations include biases and misconceptions that could be the result of training MT systems on biased data (Fan et al., 2019; Vanmassenhove et al., 2021). Some Arabic words related to Islam, such as “alhamdulillah”, “Allahu Akbar” and others, are often misinterpreted, leading to misunderstanding and misrepresentation. “Alhamdulillah,” meaning “Praise be to God,” is a common expression of gratitude and appreciation. “Allahu Akbar” translated as “God is the Greatest,” is a phrase used in worship to express devotion and admiration. The media has played a role in distorting their meaning by associating Muslims with violence and terrorism (Riggs, 2019). According to Corbin (2017), the US media often portray Muslims as terrorists with their prevalent biased narrative that “all terrorists are Muslims” which leads to the belief that “all Muslims are terrorists” (p. 457). The study highlights how post-9/11 films often present Arabs and Muslims negatively, and how news outlets such as Fox News during the Quebec City mosque attack in 2017 quickly linked Muslims to terrorism. The issue is further compounded by the predominant framing of Palestinians in public discourse, i.e. the dehumanisation of Palestinians, in particular being described collectively as “terrorist”, which has intensified significantly during the current war on Gaza, according to many media studies and news reports (see Khurma, 2024; Osama, 2025). This misrepresentation does not occur in isolation; instead, it echoes and reinforces pre-existing prejudices against the Palestinians. It is plausible that this, along with similar content, has been used to train existing MT systems, thereby allowing dominant narratives to influence and shape the output (see discussion in Nee et al., 2021). For example, in 2022, Google Translate faced criticism after its translation of the Arabic word “takhteet”, meaning “to plan,” included an example sentence, “planning to blow up the car” (Clark, 2022; Warner, 2022). Google acknowledged the issue, explaining that the error was a result of biased data used to train the MT system and removed the offensive example (Clark, 2022; Warner, 2022). When training MT systems for Arabic language, it is important to approach religious and cultural terms with an awareness of their true meanings and the contexts in which they are used to avoid biases due to selective or inaccurate translations. The issue of terminology in MT is discussed in Canfora and Ottmann (2020), who emphasise the importance of improving both the accuracy of MT and the consistency of terminology to enhance the quality of MT systems, particularly in contexts where specific meaning is critical.

Another significant challenge for MT is the lack of sufficient training data for low-resource languages. MT models depend on large amounts of parallel texts (source texts and their translations) to learn accurate language mappings. However, low-resource languages often have limited digital content, resulting in less accurate and less robust models (Bender et al., 2021; Nee et al., 2021). Both Thai and Burmese are classified as low-resource languages in the context of MT due to the limited availability of high-quality, parallel corpora for training translation models (Tzoneva, 2023; San et al., 2024). This limits the effectiveness and accuracy of MT models trained for these languages. Arabic dialects are also considered low-resource languages because they have limited parallel corpora and lack standardised written forms, especially when compared to Modern Standard Arabic (MSA). This lack of resources creates translation challenges, par-

ticularly when translating between dialects and languages that carry distinct cultural or contextual meanings (Slim & Melouah, 2024).

5.1. Potential Impact on Users

The first two case studies offer an illustration of how MT errors can result in serious and immediate consequences for individuals and can lead to dangerous and unjust actions. They display distinct yet interconnected harms arising from unverified MT. EX1 shows broad harm caused by MT, where Palestinian users were collectively misrepresented through the insertion of the term “terrorist” into their Instagram bios. This critical error led to a distorted and harmful narrative, falsely associating an entire group with terrorism. The effects of such a translation are substantial, as it could incite mistrust, fear, or even aggression towards specific individuals on Instagram, incorrectly presented as a threat. Furthermore, people who read or interact with such output may start to believe those biased views are accurate (Bender et al., 2021). This is especially concerning, as research shows that even when false or misleading information is corrected, it can still influence what people think and believe unless the correction is highly credible and clearly communicated (see discussion in Sanna & Lagnado, 2025). In this case, Meta should have taken further, more effective, steps to address the issue and reduce the potential for misunderstanding and harm. The company could have pinned clarifications to the affected profiles, publicly explained the mistranslation, and outlined concrete measures to prevent future occurrences. It might be argued that such measures are not technologically feasible, but service providers still have a responsibility to act properly when the impact on users is significant.

Furthermore, the deployment of automated systems without adequate quality control has serious repercussions for the reputation as well as the safety of individuals. EX2 shows that such mistranslations are particularly concerning because the fluency and grammatical accuracy of such outputs can mask the fact that they are mistranslations. Such critical errors can lead to misinformation and real-world harm, as evidenced by the wrongful arrest of the Palestinian worker who posted “good morning” on his Facebook page. It is important to note that this unfortunate incident could likely have been avoided if the authorities/security forces had a policy requiring human validation of machine-translated content before taking critical action based on these translations. As Martindale (2020) suggests, institutions relying on MT for high-stakes decisions should ensure verification by someone fluent in the source language. It can be added that the specific political and social context could have also contributed to such an unfair action against the Palestinian worker, underscoring the need for human supervision when deploying MT systems in politically-sensitive cultural/national contexts.

Additionally, these four examples show discrepancies in Meta’s response to MT mistakes, depending on who is being affected. For instance, in response to the distorted translations involving prominent world leaders, namely the King of Thailand (Maha Vajiralongkorn) and the President of China (Xi Jinping), Meta took significant action, including apologising, deactivating the auto-translate feature, and offering detailed explanations for the errors. This suggests that Meta gave these incidents more attention due to their high-profile nature even though the impact on Palestinian users, who were not public figures and faced similar (if not worse) consequences, was more severe. Meta’s less transparent and uneven responses in these cases point to an ethical double-standard

that is in contrast with Meta's stated Community Standards and content policies that should be applicable equally to all users anywhere (see Meta, n.d.e). These inconsistencies underscore a broader concern that the pursuit of automation and user engagement is outpacing the development of any robust mechanisms for ethical scrutiny and accountability. In the case of MT, ethical responsibility demands more than simply fixing technical issues; it also requires safeguards that ensure fair and equal treatment for all users, regardless of nationality, religion, or political standing. Without such principles in practice, the promise of inclusivity on global platforms remains unfulfilled.

6. Ethical Considerations

EX1 and EX2 raise serious concerns about algorithmic bias and the ethical implications of automated translation systems. Whether the error arose from biased training data, flawed algorithms, or lack of human oversight, the result was the amplification of a harmful stereotype. Given the political sensitivities surrounding the situation in Palestine/Israel, MT tools can inadvertently reinforce existing biases and stereotypes. This reflects a broader concern that automated systems, without proper checks, can reflect the prejudices of the societies that design them. Meta's lack of transparency about how the error occurred fails to address these concerns. Commenting on the Instagram incident of EX1, critics including Palestinian advocates argue that it highlights both digital biases in automated tools and broader issues of content censorship on platforms such as Facebook and Instagram (Paul, 2023). Many Palestinians claim their content has been shadow-banned (shadow-banning is a type of censorship practiced by online platforms where a user's content is hidden or ranked lower, reducing its visibility to others, without the user's knowledge (Suzor et al., 2019), particularly in the context of the ongoing war on Gaza (Paul, 2023). This necessitates that we ask a critical ethical question about how companies such as Meta ensure that their systems are fair, accurate, and non-discriminatory, particularly for individuals who depend on these systems to access information and may be linguistically marginalised, especially in politically sensitive or volatile contexts. These users are more vulnerable to the disproportionate impact of biases in automated systems. As mentioned before, according to Canfora and Ottmann (2020), liability for harm caused by MT mistranslations is still an unresolved issue because machines are not legally liable, leaving users and providers in an ambiguous legal position.

Putting forward ethical guidelines to ensure that MT systems are used responsibly is a necessity, with appropriate safeguards in place to protect vulnerable users from harmful consequences and actions. The way forward could be to explicitly label machine-translated content with a warning, written in plain language, indicating that it may contain errors. This would encourage users to engage with the content more cautiously, reducing the possibility of taking harmful or incorrect actions based on faulty translations. Additionally, users could be given the option to report potentially harmful translations, with a quick review and correction by a human team who can monitor, assess, and resolve translation-related issues. Currently, Facebook users can only rate the quality of a translated post by clicking a star (Facebook, n.d.), offering limited feedback on translation accuracy or appropriateness. Conducting regular evaluation to identify bias and inaccuracies in MT systems is essential, particularly for underrepresented languages with sensitive political or social contexts. However, the feasibility of

implementing these measures remains uncertain given the recent decision by Meta to put an end to its fact-checking program, leaving content moderation to the community without expert oversight or critical evaluation.

7. Corporate Responsibility

By treating the incidents as isolated technical failures, Meta avoids addressing the deeper accountability and ethical challenges that its systems pose, particularly in politically and socially sensitive contexts. These errors can be seen as indicators of underlying structural flaws in the design and deployment of their MT systems. EX1 and EX2 reflect how their MT systems might have been trained on corpora that over-associate Arabic with violence. Instagram's insertion of "terrorist" shows how linguistic bias can be turned into algorithmic violence, dehumanising a whole group of people. The anger and frustration voiced because of the mistranslations demonstrate how such errors can undermine users' confidence in MT systems and their providers. A former Facebook employee with insight into internal discussions regarding Meta's censorship of Palestinian content said the incident (EX1) deeply disturbed many, noting that it can no longer be excused as a technical problem when it spreads misinformation and dehumanises Palestinians (Paul, 2023). While Meta denied intentional censorship, similar claims were made during the major escalation of violence in Gaza in May 2021, when users reported reduced reach, and even removal, of pro-Palestinian posts (Paul, 2021). This incident led to a letter signed by over 200 employees, prompting an independent review commissioned by Meta, which found the company had indeed censored pro-Palestinian content and violated users' rights (Paul, 2023). Furthermore, it has also been reported that Meta applies content moderation unfairly, with Arabic-language posts about Palestine being removed more often than Hebrew content about Israel (Paul, 2024). A recent report by Human Rights Watch (2023), *Meta's Broken Promises: Systemic Censorship of Palestine Content on Instagram and Facebook*, criticised Meta for censoring posts related to Palestinian human rights, despite promises to protect free expression and access to information. Meta has also been criticised for making repeated promises to address these issues, but have often failed to result in meaningful action (Human Rights Watch, 2023). With calls for more clarity on the company's moderation policies, and based on the examples discussed, it can be said that the role of social media giants in shaping public discourse appears to extend beyond moderating original content to also influencing how content is presented through its MT systems.

As a technology subject to ongoing development, MT is currently at the "Peak of Inflated Expectations" stage, which can lead to "a certain blindness regarding the risks" associated with the use of machine-generated content (see discussion in Canfora and Ottmann, 2020, p. 58). Overestimation of MT capabilities can cause users to overlook the need for human supervision and quality assurance measures. Providers of these services need to be transparent regarding their limitations and capacity to ensure that users are aware and can make informed decisions.

As shown in Examples 1-4, MT can be unreliable, especially for languages that are less documented. For example, the English language, due to its history, is overrepresented in datasets used to train AI, whereas languages such as Thai and Burmese lack sufficient representation, making MT less reliable. This is because MT systems are often trained on large datasets that are biased towards dominant languages. Furthermore,

MT is not always effective at understanding slang or cultural context, which can alter the meaning of the source text (Monzó-Nebot & Wallace, 2024). The MT error that led to the unjust arrest of the Palestinian worker shows the extent of the problem of using MT without human supervision. Despite all of these concerns and legitimate debate, tech companies continue to develop and deploy free-to-use MT tools for various purposes. The key question here is what are the ethical implications of making unreliable MT tools available in contexts where they can potentially cause harm? The previous accounts of MT errors with harmful impact highlight that tech companies who provide such services need to:

- Show an ethical obligation to develop and deploy technology that protects vulnerable populations and respects basic human rights.
- Ensure human oversight to reduce the risks of relying on MT systems.
- Carry out frequent assessments of training datasets to identify and mitigate harmful linguistic patterns and underrepresented varieties.
- Use diverse data sources by intentionally including non-dominant language varieties and marginalised community texts.

Regulatory frameworks such as the Digital Services Act (DSA), the UK Online Safety Act (OSA), Ireland's Online Safety and Media Regulation Act, and the EU Code of Practice on Disinformation should be expanded to include content distributed via MT technology. These frameworks prioritise algorithmic transparency, accountability, content moderation and risk mitigation in digital spaces, principles that are essential for promoting responsible AI development, protecting language rights in the digital age and ensuring equal access to information. They can be used to strengthen protection for marginalised communities (socially, politically and linguistically) and vulnerable users, while also increasing corporate accountability. For linguistically marginalised users, MT errors and biased moderation can significantly limit their ability to engage, communicate, and access reliable information. By requiring platforms to carry out impact assessments, improve algorithmic oversight, and establish accessible mechanisms for users to report harms, these regulations can help ensure that all users are treated fairly and protected from digital harm. Furthermore, including specific provisions within the DSA and other similar regulatory frameworks to address linguistic diversity such as requiring evaluation of MT performance by language or mandating human review for high-risk content would close a critical gap. Without such measures, non-dominant language speakers will remain vulnerable to systemic mistranslation and, as a result, misinformation.

8. Concluding Remarks

The four real-life case studies of mistranslations presented have shown that MT can be a double-edged sword, offering convenience but also introducing significant risks, especially if the output is not properly reviewed or contextualised. They have revealed a serious issue with MT when dealing with user-generated content and low-resource languages, where the potential for misinterpretation and misuse is considerable. With its widespread reach and influence, misinformation generated by MT systems on social media has the potential to shape public perceptions, reinforce biases, and lead to harmful action, thus posing substantial risks to the users. Therefore, the responsibility

for ethical AI deployment is a shared one, involving not only developers but also those deploying or disseminating these technologies, e.g., social media providers.

The legal, political, and social consequences associated with these errors highlight the need to address the inherent risks and biases posed by machine-translated content. What we need is more robust, transparent, and ethical approaches to developing, deploying and evaluating MT systems. Social media platforms that also provide MT services should invest in enhancing the quality and context-awareness of their translation algorithms through high-quality training data, neutral data, in multiple languages. It would also be helpful to explicitly label machine-translated content with a warning, written in plain language, indicating that it may contain errors. This might help users better understand the potential downsides of these tools and encourage them to engage with the content carefully, reducing the likelihood of taking harmful or incorrect actions based on faulty translations. In sensitive contexts, they should also provide users with options to report translation errors or offer more human oversight for translations. Human supervision and engagement of end-users must be central to the deployment of AI technologies, particularly when dealing with vulnerable populations to reduce the risks of misrepresentation and harmful stereotypes. This is vital to prevent reinforcing inequalities and power imbalances, such as those between authoritative and vulnerable users, or between more dominant languages versus marginalised ones (see Bender et al., 2021 and Nee et al., 2021 for a discussion on power imbalance in language technologies and linguistic justice). Legal frameworks need to address these issues by promoting responsible AI development and corporate accountability to protect language rights and ensure fair access to information for all users in this digital age.

Future research should critically explore the provision of free-to-use translation services by tech companies and service providers in the context of national and international legal frameworks, aiming to improve access to accurate, multilingual information and promote linguistic justice. This is especially relevant for users in countries with limited access to, or influence over, content created about them, which may still be used to train these systems. Such efforts ensure that users are not just passive recipients of these systems, but active participants in shaping their use and development. Further research may also examine user attitudes towards the reliability of automatic translation tools and platform credibility. This may involve identifying instances where original content was significantly changed in terms of meaning due to mistranslation and interviewing these users to gain insights into their awareness of the errors, their perceptions of the impact, and their trust in the platform's translation tools. This approach can build upon the current study by extending the analysis from the technical and ethical dimensions of mistranslation to user-centred perspectives, exploring how these translation failures affect individual users' sense of agency, identity, and communication.

Acknowledgment

The research conducted in this publication was funded by the Irish Research Council under grant number (GOIPD/2022/341).

References

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election.

- Journal of economic perspectives*, 31(2), 211–236.
- Al Sharou, K., Li, Z., & Specia, L. (2021). Towards a better understanding of noise in natural language processing. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)* (pp. 53–62). Retrieved from <https://aclanthology.org/2021.ranlp-1.7/>
- Al Sharou, K., & Moorkens, J. (2024). Transitude: Machine Translation on Social Media: MT as a potential tool for opinion (mis) formation. In *Proceedings of the 25th Annual Conference of the European Association for Machine Translation* (Vol. 2, pp. 2–3).
- Al Sharou, K., & Specia, L. (2022). A taxonomy and study of critical errors in machine translation. In *Proceedings of the 23rd annual conference of the European Association for Machine Translation* (pp. 171–180).
- Baker, S. A. (2022). Alt. Health Influencers: how wellness culture and web culture have been weaponised to promote conspiracy theories and far-right extremism during the COVID-19 pandemic. *European Journal of Cultural Studies*, 25(1), 3–24.
- 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).
- Booth, R. (2025). Ditching of Facebook factcheckers a ‘major step back’ for public discourse, critics say. *The Guardian*. Retrieved from <https://www.theguardian.com/technology/2025/jan/07/ditching-facebook-factcheckers-major-step-back-public-discourse>
- Calzada-Pérez, M. (2014). *Apropos of ideology: translation studies on ideology-ideologies in translation studies*. Routledge.
- Canfora, C., & Ottmann, A. (2020). Risks in neural machine translation. *Translation Spaces*, 9(1), 58–77.
- Clark, M. (2022). Google Translate Suggested ‘Blow Up The Car’ When Arabic Word ‘Plan’ Entered. *Newsweek*. Retrieved 2024-12-11, from <https://www.newsweek.com/google-translate-suggests-blow-car-when-arabic-word-plan-entered-1726537>
- Cole, S. (2023). Instagram ‘Sincerely Apologizes’ For Inserting ‘Terrorist’ Into Palestinian Bio Translations. *404 media*. Retrieved from <https://www.404media.co/instagram-palestinian-arabic-bio-translation/>
- Comito, C. (2023). The role of social media in the battle against COVID-19. In *Artificial Intelligence in Healthcare and COVID-19* (pp. 105–124). Academic Press.
- Comito, C. (2024). Polarization and Misinformation: Anticipating Early Signs of Potential Fake News on Social Media. In *2024 15th International Conference on Information, Intelligence, Systems & Applications (IISA)* (pp. 1–8).
- Corbin, C. M. (2017). Terrorists are always Muslim but never white: At the intersection of critical race theory and propaganda. *Fordham L. Rev.*, 86(455).
- Costa-jussà, M. R., Cross, J., Çelebi, O., Elbayad, M., Heafield, K., Heffernan, K., ... Team, N. (2022, August). *No language left behind: Scaling human-centered machine translation*. arXiv. Retrieved from <http://arxiv.org/abs/2207.04672> (arXiv:2207.04672 [cs])
- Department for Science, Innovation & Technology. (2025). *Online Safety Act: explainer. Updated 24 April*. Retrieved 2025-05-12, from <https://www.gov.uk/government/>

- publications/online-safety-act-explainer/online-safety-act-explainer
 Department of Culture, Communications & Sport. (2022). *Online Safety and Media Regulation Act*. Retrieved 2025-12-12, from <https://www.gov.ie/en/publication/d8e4c-online-safety-and-media-regulation-bill/>
- European Commission. (n.d.). *The 2022 Code of Practice on Disinformation*. Retrieved 2025-05-11, from <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation>
- European Union. (2022). *Digital Services Act (DSA) - Regulation (EU) 2022/2065*. Retrieved 2025-04-12, from <https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng>
- Facebook. (n.d.). *Translate Feed*. Retrieved 2025-05-11, from <https://www.facebook.com/help/509936952489634>
- Fan, L., White, M., Sharma, E., Su, R., Choubey, P. K., Huang, R., & Wang, L. (2019). *In plain sight: Media bias through the lens of factual reporting*. arXiv. Retrieved from <http://arxiv.org/abs/1909.02670> (arXiv:1909.02670 [cs]) doi: 10.48550/arXiv.1909.02670
- Farhan, A. K. (2017). *Ideological manipulation in the translation of political discourse: a study of presidential speeches after the Arab Spring based on corpora and critical discourse analysis* (Doctoral dissertation). University of Surrey.
- Ghosh, S., & Caliskan, A. (2023). Chatgpt perpetuates gender bias in machine translation and ignores non-gendered pronouns: Findings across bengali and five other low-resource languages. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, (pp. 901–912).
- Hern, A. (2017). Facebook translates 'good morning' into 'attack them', leading to arrest. *The Guardian*. Retrieved 2025-01-03, from <https://www.theguardian.com/technology/2017/oct/24/facebook-palestine-israel-translates-good-morning-attack-them-arrest>
- Human Rights Watch. (2023). *Meta's Broken Promises Systemic Censorship of Palestine Content on Instagram and Facebook*. HRW. Retrieved from <https://www.hrw.org/report/2023/12/21/metass-broken-promises/systemic-censorship-palestine-content-instagram-and>
- Instagram. (n.d.). *How Instagram Feed Works*. Retrieved 2025-01-17, from https://help.instagram.com/512686498916530/?helpref=related_articles
- Ireton, C., & Posetti, J. (2018). Journalism, 'Fake News' & Disinformation: Handbook for Journalism Education and Training. *UNESCO Publishing*. Retrieved 2025-01-14, from <https://web.archive.unesco.org/web/20230926213448/https://en.unesco.org/fightfakenews>
- Kaplan, J. (2025). More Speech and Fewer Mistakes. *Meta*. Retrieved 2025-05-03, from https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/?utm_source=chatgpt.com
- Khurma, M. (2024). A Year of War Since October 7: On Narrative and Dehumanization in Arab and Israeli Societies. *Wilson Center*. Retrieved 2025-06-25, from <https://www.wilsoncenter.org/article/year-war-october-7-narrative-and-dehumanization-arab-and-israeli-societies>
- Lee, K. W., & Qian, M. (2022). Misinformation in Machine Translation: Error Categories and Levels of Recognition Difficulty. In *International Conference on Human-Computer Interaction* (pp. 533–545). Cham.

- Marking, M. (2020). Thai Mistranslation Shows Risk of Auto-Translating Social Media Content. *Slator*. Retrieved 2025-05-06, from <https://slator.com/thai-mistranslation-shows-risk-of-auto-translating-social-media-content/#:~:text=A%20July%2028%2C%202020%20post,in%20Thai%20on%20Thai%20PBS>
- Martin, D. A., Shapiro, J. N., & Nedashkovskaya, M. (2019). Recent trends in online foreign influence efforts. *Journal of Information Warfare*, 18(3), 15–48.
- Martindale, M. J. (2020). Responsible ‘Gist’ MT Use in the Age of Neural MT. In *Workshop on the Impact of Machine Translation (iMpacT 2020)* (pp. 18–45).
- Martindale, M. J. (2024). *When Good MT Goes Bad: Understanding and Mitigating Misleading Machine Translations* (Doctoral dissertation). University of Maryland, College Park.
- McMahon, L., & Tidy, J. (2023). Instagram sorry for adding ‘terrorist’ to some Palestinian user bios. *BBC*. Retrieved 2024-12-01, from <https://www.bbc.com/news/technology-67169228>
- McPherson, p. (2020). Facebook says technical error caused vulgar translation of Chinese leader’s name. *Reuters*. Retrieved 2025-01-01, from <https://www.reuters.com/article/us-myanmar-facebook/facebook-apologizes-after-vulgar-translation-of-chinese-leaders-name-idUSKBN1ZH01B/>
- Meta. (n.d.a). *About*. Retrieved 2025-05-11, from <https://www.meta.com/en-gb/about/>
- Meta. (n.d.b). *No Language Left Behind Driving inclusion through the power of AI translation*. Retrieved 2025-05-11, from <https://ai.meta.com/research/no-language-left-behind/>
- Meta. (n.d.c). *Code of Conduct for Virtual Experiences*. Retrieved 2025-05-11, from <https://www.meta.com/gb/legal/quest/code-of-conduct-for-virtual-experiences/>
- Meta. (n.d.d). *Hateful Conduct*. Retrieved 2025-05-11, from <https://transparency.meta.com/en-gb/policies/community-standards/hateful-conduct/>
- Meta. (n.d.e). *Community Standards*. Retrieved 2025-05-11, from <https://transparency.meta.com/en-gb/policies/community-standards>
- Meta. (n.d.f). *Misinformation*. Retrieved 2025-05-11, from <https://transparency.meta.com/en-gb/policies/community-standards/misinformation/>
- Monzó-Nebot, E., & Wallace, M. (2024). Gender and ethnolinguistic lawfare: Weaponizing the law. *Just. Journal of Language Rights & Minorities, Revista de Drets Lingüístics i Minories*, 3(2), 7–116.
- Narayanan, S. (2022). Automated misinformation: Mistranslation of news feed using multi-lingual translation systems in Facebook [Abstract]. In *Affinity Workshop: Global South in AI*. Retrieved 2025-05-01, from <https://nips.cc/virtual/2022/62672>
- Nee, J., Macfarlane Smith, G., Sheares, A., & Rustagi, I. (2021). October. Advancing social justice through linguistic justice: Strategies for building equity fluent NLP technology. In *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (pp. 1–9).
- Nurminen, M., & Koponen, M. (2020). Machine translation and fair access to information. *Translation Spaces*, 9(1), 150–169.
- Osama, Y. (2025). Deconstructing Narratives: An Analysis of Dehumanization Techniques in the US Media Representation of Palestinians at the Onset of the War on Gaza. *The Arab Journal of Media and Communication Research (AJMCR)*(48), 143–185.

- Paul, K. (2021). Facebook under fire as human rights groups claim ‘censorship’ of pro-Palestine posts. *The Guardian*. Retrieved 2025-01-10, from <https://www.theguardian.com/media/2021/may/26/pro-palestine-censorship-facebook-instagram>
- Paul, K. (2023). Instagram users accuse platform of censoring posts supporting Palestine. *The Guardian*. Retrieved 2025-01-10, from <https://www.theguardian.com/media/2021/may/26/pro-palestine-censorship-facebook-instagram>
- Paul, K. (2024). Meta struggles with moderation in Hebrew, according to ex-employee and internal documents. *The Guardian*. Retrieved 2025-05-11, from <https://www.theguardian.com/technology/article/2024/aug/15/meta-content-moderation-hebrew>
- Pedace, L. (2021). Misinformation and Manipulation on Social Media: User-based and Network-based view. *iSCHANNEL*, 16(1).
- Petrosyan, A. (2025). Number of internet and social media users worldwide as of February 2025 (in billions). *Statista*. Retrieved from <https://www.statista.com/statistics/617136/digital-population-worldwide/>
- Riggs, A. (2019a). On France, terrorism and the English press: Examining the impact of style in the news. In C. Declercq, J. Munday, & M. F. Federici (Eds.), *Intercultural crisis communication: Translation, interpreting and languages in local crises* (pp. 193–214). Bloomsbury Publishing.
- Riggs, A. (2019b). *Stylistic deceptions in online news. Journalistic style and the translation of culture* (Bloomsbury ed.).
- Saaïda, M. (2023). The Role of Social Media in Shaping Political Discourse and Propaganda. *Science for all Publication*, 3(2), 1–8.
- San, M. E., Usanavasin, S., Thu, Y. K., & Okumura, M. (2024). A Study for Enhancing Low-resource Thai-Myanmar-English Neural Machine Translation. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 23(4), 1–24.
- Sanna, G. A., & Lagnado, D. (2025). Belief updating in the face of misinformation: The role of source reliability. *Cognition*, 258, 106090.
- Serrano, J. (2020). *Facebook Apologizes for Translating Chinese President’s Name as ‘Mr Shithole’*. Retrieved 2025-01-01, from <https://gizmodo.com/facebook-apologizes-for-translating-chinese-president-s-1841095962>
- Slim, A., & Melouah, A. (2024). Low resource Arabic dialects transformer neural machine translation improvement through incremental transfer of shared linguistic features. *Arabian Journal for Science and Engineering*, 1–17.
- Soponvijit, K. (2022). *Lost in translation: Facebook’s royal translation error*. Retrieved 2025-01-11, from <https://tankytech.net/2021/04/05/facebooks-royal-translation-error>
- Suzor, N. P., West, S. M., Quodling, A., & York, J. (2019). What do we mean when we talk about transparency? Toward meaningful transparency in commercial content moderation. *International Journal of Communication*, 13(18).
- Taylor, J. (2023). Instagram apologises for adding ‘terrorist’ to some Palestinian user profiles. *The Guardian*. Retrieved 2024-12-16, from <https://www.theguardian.com/technology/2023/oct/20/instagram-palestinian-user-profile-bios-terrorist-added-translation-meta-apology>
- The BBC. (2017). No Arabic-speaking police officer was consulted before the man was

- arrested on suspicion of incitement, online. *BBC*. Retrieved 2025-01-19, from <https://www.bbc.com/news/world-middle-east-41714152>
- The BBC. (2020). Facebook blames 'technical issue' for offensive Xi Jinping translation. *BBC*. Retrieved 2025-01-11, from <https://www.bbc.com/news/world-asia-51166339>
- Tsimitakis, M. (2025). *EU Demands Comprehensive Risk Assessment from Meta Before Fact-Checking Elimination*. Retrieved 2025-05-13, from <https://creativesunite.eu/article/eu-demands-comprehensive-risk-assessment-from-meta-before-fact-checking-elimination>
- Tzoneva, D. (2023). Fixing a Low-Resource Language's Quality Issues — Burmese. *Pulse of Asia*. Retrieved 2025-01-11, from <https://www.1stopasia.com/blog/low-resource-language-issues-burmese/>
- UN. (2022). UN: ARTICLE 19 welcomes report on the right of access to information. *Article 19*. Retrieved 2025-05-01, from <https://www.article19.org/resources/un-article-19-welcomes-report-on-the-right-of-access-to-information/>
- Vanmassenhove, E., Shterionov, D., & Gwilliam, M. (2021). *Machine translationese: Effects of algorithmic bias on linguistic complexity in machine translation*. arXiv. Retrieved 2025-10-25, from <http://arxiv.org/abs/2102.00287> (arXiv:2102.00287 [cs]) doi: 10.48550/arXiv.2102.00287
- Vincent, J. (2022). *Meta open sources early-stage AI translation tool that works across 200 languages*. Retrieved 2025-01-11, from <https://www.theverge.com/2022/7/6/23194241/meta-facebook-ai-universal-translation-project-no-language-left-behind-open-source-model>
- Wang, B., & Feng, D. (2018). A corpus-based study of stance-taking as seen from critical points in interpreted political discourse. *Perspectives*, 26(2), 246–260.
- Warner, A. (2022). *Following outcry, Google Translate removes offensive example phrase for Arabic entry*. Retrieved 2025-04-11, from <https://multilingual.com/google-translate-arabic-bias/>

Deaf in AI: AI language technologies and the erosion of linguistic rights

Maartje De Meulder ¹

¹HU University of Applied Sciences Utrecht, The Netherlands; Faculty of Healthy and Sustainable Living, Research Group Speech and Language Therapy: Participation through Communication, maartje.demeulder@hu.nl

Abstract

This paper explores the interplay of AI language technologies, sign language interpreting, and linguistic access, highlighting the complex interdependencies shaping access frameworks and the trade-offs these technologies bring. While AI tools promise innovation, they also perpetuate biases, reinforce technoableism and deepen inequalities through systemic and design flaws. The historical and contemporary privileging of sign language interpreting as the dominant access model – and the broader inclusion ideologies it reflects – shape AI's development and deployment, often sidelining deaf languaging practices and introducing forms of linguistic subordination to technology. Drawing on Deaf Studies, Sign Language Interpreting Studies, and crip technoscience, this paper critiques the framing of AI as a substitute for interpreters and examines its implications for access hierarchies. It calls for deaf-led approaches to foster AI systems that remain equitable, inclusive, and trustworthy, supporting rather than undermining linguistic autonomy and contributing to deaf-aligned futures.

Keywords: AI language technologies, deaf, sign language interpreting, linguistic rights, access.

Resumo

Este artigo explora a interação entre as tecnologias linguísticas de IA, a interpretação de língua gestual e o acesso linguístico, destacando as interdependências complexas que moldam os quadros de acesso e as soluções de compromisso que estas tecnologias trazem. Embora as ferramentas de IA prometam inovação, também perpetuam preconceitos, reforçam o technoableismo e aprofundam as desigualdades através de falhas sistêmicas e de concepção. O privilégio histórico e contemporâneo da interpretação de língua gestual como modelo de acesso dominante - e as ideologias de inclusão mais amplas que reflecte - moldam o desenvolvimento e a implementação da IA, muitas vezes marginalizando as práticas linguísticas dos surdos e introduzindo novas formas de subordinação linguística à tecnologia.

Com base nos Estudos Surdos, nos Estudos de Interpretação de Língua Gestual e na tecnociência crítica, este artigo critica o enquadramento da IA como um substituto dos intérpretes e examina as suas implicações para as hierarquias de acesso. Apela a abordagens lideradas por surdos para promover sistemas de IA que permaneçam equitativos, inclusivos e fiáveis, apoiando em vez de minar a autonomia linguística e contribuindo para futuros alinhados com a surdez.

Palavras-chave: *Tecnologias linguísticas em IA, surdos, interpretação de língua gestual, direitos linguísticos, acesso.*

1. Introduction: AI, access, and the erosion of linguistic rights

After a recent presentation I gave on ethical considerations in sign language AI, a European deaf attendee approached me. Like me, they had an advanced degree, high socioeconomic status, fluency in several signed and written languages, and digital and AI literacy, along with reliable access to human sign language interpreting services. They were also a public figure engaged in interpreting and linguistics research. With visible frustration, they confided that they were fed up with the current state of human sign language interpreting services in their country and beyond – the logistical hurdles, the pervasive audism, the loss of privacy, and the economics of interpreting services. “AI”, they said, “might finally free us from this system. Maybe in the not-so-distant future, we won’t need human interpreters anymore.” They paused, then added with a wry smile, “Of course, saying that out loud would cause a storm.”

Yet, they were not the first person (of a comparable position) telling me this, and I assume they will not be the last. In a recent paper U.S. deaf authors Hall, Brick, and Millios (2024, p. 10) state: “It must be acknowledged that if Deaf people could go without interpreters, they would.” Those deaf people involved in research on interpreting, working as interpreters, and/or being on the receiving end of interpreting services are aware of the limitations of human interpreting in terms of providing equitable “access” (Haualand, De Meulder, & Napier, 2023). For some, AI tools such as speech-to-text applications may offer a (partial) relief from the access labour associated with working with human sign language interpreters. At the same time, deaf communities are generally acutely aware of how access is collective and interconnected, recognising that access is not just about personal choices but about shared opportunities and systemic structures.

There is also an implicit deaf understanding of access rights as, what I call, “slippery slope rights”, referring to measures initially rejected by deaf communities due to their limitations, only to be normalised over time as acceptable solutions. For instance, while deaf people have objected against the use of Video Remote Interpreting (VRI) – a service that connects users to remote interpreters through video technology – in hospitals because of unreliable technology, issues with establishing trust, and privacy, the use of VRI in some medical situations is now considered acceptable by health care providers (James et al., 2022; Kushalnagar, Paludneviene, & Kushalnagar, 2019).

The attendee who approached me knew there is a difference between saying (like Hall et al., 2024) “if we could do without interpreters, we would” or saying out loud that AI could be a replacement for or even serve as a tentatively acceptable addition to human interpreters. They knew very well this could trigger governments and institutions to abandon interpreter services altogether, in favour of “cheaper” AI-driven solutions. For many deaf people without similar privileges – those with lower AI literacy, less access to reliable AI tools, lower socioeconomic status, and/or restricted access to human sign language interpreting services (the reality for most deaf people around the world) – this shift could have devastating consequences, potentially eroding hard-won linguistic and accessibility rights. The implementation of AI technologies risks exacerbating access hierarchies and shifting labour onto individuals without adequate resources or support. This raises critical questions about how these tools redistribute access and reshape social hierarchies. Deaf-aligned futures depend on ensuring deaf people retain control over their access choices, resisting the imposition of technologies that may fail to align with their preferences. While some deaf users may embrace AI tools as empowering, others who prefer working with human interpreters might face a narrowing of choices.

This paper examines this push and pull of AI tools in the context of deaf people, sign language interpreting, and linguistic access, highlighting the complex deaf interdependencies that underpin access frameworks, the trade-offs inherent in these technologies, and the critical need to navigate these shifts carefully. Because in practice, the erosion of linguistic rights may already be underway. Like many technological shifts, this does not happen overnight but gradually, through subtle shifts in policy, funding priorities, and service provision. These incremental changes often remain unnoticed until they become deeply embedded in systems, leaving those most reliant on access services with fewer options and greater disparities. I explored some initial thoughts on this in an earlier article (De Meulder, 2021). This was just before the rapid rise of generative AI and its uptake by everyday users, which has significantly shifted the landscape. Today, I am writing in a constantly and quickly evolving environment, as AI capabilities advance, and the role of sign language interpreters evolves.

These rapid shifts are reshaping the very fields they affect, challenging established assumptions, theories, and practices. Firstly, AI disrupts long-standing models in Sign Language Interpreting Studies and Deaf Studies that primarily centre human interpreters and human-to-human interaction as the dominant means of access and mediation (Napier, McKee, & Goswell, 2010; Stone et al., 2022; Young, Napier, & Oram, 2019). With some exceptions (e.g. Clark, 2021), these theories have yet to fully address the implications of these developments. Secondly, AI has significant implications for the theoretical and methodological framework of sign language (socio)linguistics and has the potential to disrupt deeply held attitudes and ideologies about (signed) languages (Backus et al., 2023; Kelly-Holmes, 2024). With a few exceptions (e.g. Lee, Hill, & Smith, 2021), research on sign language ideologies and sign language sociolinguistics so far has focused exclusively on human-to-human interactions, relying on established theoretical and methodological frameworks (Kusters et al., 2020). These fields must now grapple with the implications of human-AI interactions, including how these influence perceptions of access, mediation, interpreting, authority, correctness, and acceptability in language use.

2. Setting the stage

The widespread adoption of AI language technologies by everyday users is reshaping how we communicate, work, and engage with the world. Generative language AIs, powered by Large Language Models (LLMs), understand, process and (co-)produce human-like language across various contexts. Current examples of these technologies include tools like Google Translate and automated captions on platforms such as Zoom, intelligent personal assistants like Siri and Alexa, chatbots, and speech-to-text applications such as Otter.ai. More advanced conversational AIs, such as ChatGPT and Gemini, further demonstrate the growing sophistication of LLMs. Beyond text, AI's generative capabilities extend to the visual domain with tools like DALL-E and Midjourney, producing images outputs from textual prompts. At the same time, AI systems can interpret visual input in real-time, describing objects, environments, or scenes through spoken or written descriptions.

In the not-so-distant future, these tools are expected to include specialised sign language technologies – those designed for sign language recognition, generation and translation. Examples of these technologies are virtual sign language avatars and automated translation interfaces between signed and spoken languages. The barriers to developing sign language technologies (primarily the volume and quality of data) have been well-documented (De Coster et al., 2023; Desai et al., 2024; Shterionov, Leeson, & Way, 2024; Vandeghinste et al., 2024; Yin et al., 2024). While these challenges mean that these technologies currently remain underdeveloped for broad use, the state-of-the-art is advancing rapidly. As I write this in January 2025, another new project “SignGPT” (£8,45m) has just been announced, aiming to build a British Sign Language LLM capable of end-to-end translation English/BSL and build “the largest sign language dataset in the world”. Reflecting broader trends in the field of sign language AI (Desai et al., 2024), all the Principal Investigators are hearing academics, and the data (also) include interpreted datasets English/BSL (Albanie et al., 2021). I will come back to these trends and biases later in the article.

These AI language technologies are evolving into semi-autonomous communicative agents that mimic ‘human’ forms of interaction. Important in the context of sign languages and deaf people is that these tools operate across various modalities (text, speech, gestures, movement, sign) and take on different degrees of embodiment, from text-based interfaces to human-like robots and virtual avatars. Yet, the implications of these AI technologies for deaf people (who use sign languages and navigate diverse communication practices) remain largely unexplored or primarily examined from hearing, non-disabled perspectives, although there are emerging exceptions (e.g. Angelini, Tran, De Meulder, & Spiel, in prep; Glazko et al., 2023; Huffman et al., 2024).

In language policy and legislation, deaf people are often positioned as both linguistic minorities and individuals with disabilities (De Meulder & Murray, 2017). This dual status grants deaf people (primarily in Europe, Australia, New Zealand, Canada and the United States) access to sign language interpreting services. While there are parallels between public service interpreting for deaf people and other groups who use interpreters for accessing public services such as immigrants, refugees, and other language groups (Stone, 2010), there are also differences. For immigrants and refugees, the demand for interpreting services often fluctuates with migration patterns, which means the request

for interpreting in certain languages rises and falls over time (Giambruno, 2014; Piller, 2017; Skaaden & Wadensjö, 2014). For immigrants, interpreting services are frequently framed as a temporary measure, intended to bridge the gap until they acquire proficiency in the majority language. This creates an intersecting dynamic for deaf migrants and refugees (Duggan & Holmström, 2024; Sivunen, 2019). The perceived temporality of spoken language interpreting services is shaped by discourses of who is deemed a ‘worthy’ recipient of interpreting, particularly within assimilationist institutional practices that enforce the learning of the majority language (Piller, 2017). By contrast, sign language interpreting services are generally provided over deaf people’s lifetime and are not limited to the public sector, as is often the case for spoken language interpreting, but provided across a wide range of domains. For example, in several Western countries, a significant portion of sign language interpreters work in educational settings, including early childhood and primary education, domains where spoken language interpreters tend not to work. This divergence reflects deaf people’s categorisation as disabled and the associated ideologies of inclusion, which justify placing young deaf children in regular education settings with interpreters. This practice, and the larger practice of sign language interpreting services to fulfil ideologies of inclusion has been criticised as an “illusion of inclusion”, masking deeper systemic inequities (Caselli, Hall, & Henner, 2020; De Meulder & Haualand, 2021; De Meulder & Murray, 2021; Russell, 2021).

This intersection of technological progress and systemic challenges creates a critical moment of both ruptures and opportunities for resilience, with profound implications for access and communication practices, as well as for theory-building in the field. How do these dynamics unfold in practice? To answer this question, I will turn to how “access” is conceptualised in Deaf Studies, Sign Language Interpreting Studies, and crip technoscience, providing a framework to critically evaluate discourses surrounding the development and use of AI (sign) language technologies. Indeed, the issue of ‘access’ (to information, education, political participation, employment, health care, the justice system, etc.) has been paramount on international deaf agendas, having been the central advocacy issue for national, regional and international deaf NGOs for many years, and takes a central place in discourses around the development and use of AI language technologies.

Before proceeding, let us first have a look at the complex interplay between AI-driven language technologies and human sign language interpreting services in deaf lives, and how these systems interact in everyday practice. Consider a typical day of a literate deaf person with a high socioeconomic status (like me) in 2025: a vibrating smartwatch wakes them, real-time captions generated by automatic speech recognition relay the news. Their hearing family members ask Alexa for weather updates through voice commands – it does not support sign language yet. At work, Zoom meetings offer live captions alongside a remote human sign language interpreter, but the captions often misrepresent jargon and the interpreter, unfamiliar with the topic, struggles to keep up. During lunch, a call to the doctor involves a remote human interpreter via a Video Relay Service. While effective in mediating the conversation, the reliance on a third party to access private health information feels intrusive. Later, a visit to the doctor’s office prompts a different choice: rather than using an interpreter, they opt for a speech-to-text app, which feels more discreet. Yet, the app’s accuracy is inconsistent, creating moments of uncertainty, misunderstanding or just plain awkwardness. At a local shop, they use a speech-to-text

app to communicate with the shop assistant. The app struggles with environmental noise and regional accents, turning a simple low-stake conversation into a prolonged source of frustration. For a parent-teacher meeting they request an interpreter since the teacher is not comfortable speaking into a phone and the WIFI at the school is spotty. However, the school hires an unfamiliar interpreter, leading to a strained conversation. Later in the day they attend a court hearing as part of a dispute over workplace discrimination. Although they requested a qualified interpreter, the court arranges for an AI-powered translation instead, citing efficiency and cost-saving measures.

This vignette illustrates how AI-driven language technologies and human sign language interpreting services interact in shaping deaf communication and access for deaf people with high literacy and socioeconomic status privileges. The outcomes would differ significantly for deaf people without those privileges, i.e. without the same technology, legislation, internet access, etc. Moreover, as AI capabilities advance and the role of interpreters evolves, a vignette written in a few years' time could sketch an entirely different landscape.

3. The intersection of AI sign language technology development and ideologies centred on interpreting

The historical and contemporary situation of sign language interpreting services and the associated discourses around 'access' significantly influence how AI language technologies are being developed and (will be) deployed in this domain. In Europe, Australia, New Zealand, Canada and the United States, sign language interpreting services have now existed for more than four decades and have evolved from volunteer charity work to well-established social institutions and professional services. However, the availability of sign language interpreters is often equated with access, with increasing the number of interpreters posited as the primary solution to access issues. This perspective is so entrenched that interpreter services are now a prerequisite for public service provision (De Meulder & Haualand, 2021; De Meulder, Murray, & McKee, 2019). Yet training sign language interpreters, who predominantly represent white, female, hearing, abled, new signer demographics – and thus become a mostly white, female, hearing, abled profession (Hill, Isakson, & Nakahara, 2022; Napier et al., 2022) requires substantial time and resources. Virtually all sign language interpreter training programs struggle with teaching sign language acquisition *and* the acquisition of subject-specific knowledge within a three or four-year (BA or MA) program (Webb, Napier, & Adam, 2025). As a result, interpreter education and provision represent a limited human capacity solution.

Simultaneously, in most of these countries, institutionalised sign language interpreting services are undergoing significant shifts driven by multiple converging forces. They face logistical and capacity constraints, challenging their sustainability as a means of providing access. These challenges manifest differently across contexts: for example, in Sweden, a retirement wave is creating shortages (DIK, 2024), while Finland's saturated labour market raises concerns over interpreters' job security (Rainò, 2022). In Germany, the shortage of sign language interpreters creates tensions between (hearing) sign language interpreters and deaf consumers Maaß (2024). In more general terms, the pressures stem from systemic supply-demand imbalances, ideologies of inclusion that over-rely on human interpreters, and an aging workforce outpacing new entrants.

Simultaneously, in many of the same countries that experience these pressures, some deaf people are achieving higher levels of education and literacy in written languages, especially in professional and academic sectors. This, and the rapid uptake of generative AI, also by deaf users, may be changing reliance on human interpreters to meet access needs, prompting shifts in how and when these services are sought.

These dynamics are compounded by perceptions of governments and other policy actors as interpreting services as logistically complex and costly, contributing to a potential growing push towards AI-driven solutions as a more ‘cost-effective’ alternative. My own work (De Meulder & Haualand, 2021), highlighting the limits of sign language interpreting as a quick fix for inclusion, is regularly cited to motivate the development of AI sign language tools (e.g. Coy, Mohammed, & Skerrit, 2024; Gunarhadi, Wiliyanto, Anggarani, & Anggrellangi, 2024; Nasution & Batubara, 2024). Anecdotal evidence from some European countries indicates that deaf (young) adult consumers of interpreting services are increasingly directed to rely on speech-to-text AI (also in education settings) when their allotted interpreting hours are exhausted or when human interpreters are unavailable. This shift highlights the slippery slope of access rights and raises concerns about choice and agency.

A related issue I preliminarily discussed in earlier work (De Meulder, 2021) is that the primary bottleneck for developing AI sign language technologies is and remains the scarcity and quality of available data. Here again, sign language interpreting plays a significant role. AI systems are inherently data-hungry and due to the lack of data “in the wild”, developers often turn to readily available internet data, including interpreted data sets from, among other things, public broadcasting. Most language AI uses language pairs to learn, so in theory this is not uncommon. What is uncommon is that the datasets are not just language pairs, *but simultaneously interpreted* language pairs. These datasets include both the signed input from (mostly hearing, mostly new signer) interpreters and the spoken source languages, or vice versa, with the signed source language from for example deaf presenters or interpreters paired with spoken output. The use of interpreted datasets as source material to develop sign language AI has been acknowledged as problematic (e.g. by Desai et al., 2024; Fox, Woll, & Cormier, 2023; Vandeghinste et al., 2024). Indeed, interpreted language use differs significantly from usage patterns in deaf communities and the nature of scripted and interpreted language use may result in a distorted representation of sign languages in AI systems. However, this reliance on interpreted datasets not only skews representations of sign languages in AI systems but also reflects and reinforces broader discourses that prioritize interpreting as the dominant framework for accessibility in both policy and practice.

National and international deaf and sign language interpreter associations still largely view sign language interpreting as the primary or preferred method of access, particularly in what they call “high-stakes contexts” such as medical settings, educational settings, live broadcasting, and legal proceedings (World Federation of the Deaf & World Association of Sign Language Interpreters, 2018). This is confirmed by recent data from interviews with deaf users in Europe, who list the same settings as high-stakes. Although, they further fine-tune them depending on whether the information flow is one-way (AI) or two-way (interpreter), short (AI) or long (interpreter), with high (interpreter) or low (AI) impact, planned in advance (interpreter) or of a fleeting, spontaneous nature with low-impact (AI) (“Towards) Sign Language Machine Translation in Practice”, n.d.). Current

scholarly and practice discourses (e.g. Picron et al., 2024) frequently compare AI to human interpreters, emphasising that the quality of human interpreters is, and will likely remain, superior to AI-driven alternatives. Research often frames AI and human interpreters within a hierarchy, with human interpreters consistently positioned as superior (SignOn Consortium, 2023). This framing, taken together with the use of interpreted datasets, perpetuates a problematic benchmark for AI sign language technologies.

Indeed, these technologies are not meant to substitute or compete with interpreters; they are meant to support and reflect deaf languaging practices, i.e. the ways deaf people themselves use and navigate signed and written languages in different contexts. Furthermore, bias risks are amplified when AI systems are trained on an amalgam of data sets containing interpreted input. This concern is heightened by the already significant influence of interpreters on deaf people's signing. For example, sign language interpreters often serve as de facto language models for deaf learners in regular educational settings (Caselli et al., 2020), and deaf people frequently adjust their signing to facilitate comprehension by interpreters. This is a form of linguistic subordination to interpreters that now risks being compounded by linguistic subordination to technologies. In my previous work (De Meulder, 2021), I cautioned against the potential use of AI sign language technologies, especially those trained on interpreted datasets, in sign language interpreter training programs. This concern remains pressing, particularly in a context of ongoing budget cuts to higher education programs and a neoliberal climate that prioritises quantity over quality, with limited resources for staff and students alike (Webb et al., 2025). This would exacerbate existing problems in sign language interpreter training programs, which often already fail to reflect the racial, multicultural and multilingual diversity of deaf communities (Robinson, Sheneman, & Henner, 2020).

However, in the current climate marked by an emerging co-existence of AI language technologies with human interpreting, deaf NGOs may find themselves in a double bind. Their position is shaped not only by their commitment to ensuring reliable and nuanced access, but also by an awareness that access is inherently collective, extending beyond individual accommodations to encompass the diverse needs, preferences, and circumstances of different deaf communities. This means balancing different priorities, levels of technological accessibility, and situational constraints, while advocating for solutions that are both scalable and adaptable across various contexts and individuals. Indeed, access through human interpreters already generates hierarchies among deaf people (De Meulder & Hualand, 2021). Deaf people who can deploy multilingual and multimodal resources, often work with interpreters and have access to the "best" interpreters (because of professional experience, cooperation with interpreters, and network) seem to get more out of interpreter-mediated interactions than deaf people who do not have these privileges or resources. AI introduces additional (although possibly different) layers of inequality, arising from differential access to different AIs and varying levels of digital and linguistic literacy among deaf users. These dynamics illustrate the intersection of sign language technology development and inclusion ideologies centred on interpreting. However, there are even additional layers of complexity.

4. Ambivalence and resilience in interpreting relationships

Building on the dynamics described above, it is crucial to situate them within the broader context of issues that AI sign language technology developers (who in most cases have

no lived experience of being deaf) often overlook. Deaf people have a historical and present complex and ambivalent relationship with sign language interpreters – marked by both individual critiques on interpreters and the broader interpreting system, while at the same time a reliance on this system for access. Hall et al. (2024), whom I referred to in the introduction, state that “at a deeply fundamental level, Deaf people do not want interpreters in their lives; they would prefer to navigate the world on their own” (p. 11). This is an issue that is often not spoken about and not acknowledged by most hearing people, and many interpreters themselves. Interpreters are frequently perceived by deaf people as (un)invited guests in their personal lives (Burke, 2017; Hall et al., 2024) and as oppressive actors and gatekeepers unaware of their own power and privilege (Robinson et al., 2020). For a large part, this also has to do with how they are being trained (De Meulder & Stone, 2024; Sheneman & Robinson, 2021). Yet, in Deaf Studies and Sign Language Interpreting Studies, the relationship between deaf people and sign language interpreters is often still described as being “bound together in mutual systems of precarious interdependence” (Marie, 2019 in Marie & Friedner, 2021, p. 5), with deaf lives seen as “predicated on interpreters” and “deaf selves” largely existing in translation (Young, Napier, & Oram, 2019; Young, Oram, & Napier, 2019). Sign Language Interpreting Studies has also examined the issue of ‘trust’ as a crucial element in the working relationships between sign language interpreters and deaf people, with interpreters and deaf consumers often claiming that trust is a prerequisite to a successful interaction (De Meulder, Napier, & Stone, 2018; Napier, 2011). Although, this concept of trust in interpreted interactions has also been critically examined by deaf scholars (O’Brien et al., 2023). Emerging research, again not coincidentally, mostly done by deaf scholars (Hualand et al., 2023), has exposed the “access labour” for deaf people working with human sign language interpreters. This refers to the (often invisible) additional emotional, cognitive and logistical efforts deaf consumers are dealing with in these situations. This includes “gaze work” (De Meulder & Stone, 2024), preparing and “grooming” interpreters (De Meulder et al., 2018), dealing with interpreters’ insecurities, and negotiate relationships with them (Chua et al., 2022; Crawley & O’Brien, 2020). Chua et al. (2022) found that working with sign language interpreters is one of the main reasons for impostor syndrome in deaf academics.

It is therefore not surprising that some deaf people, particularly those with (high) literacy privileges, start to see AI as offering resilience: AI tools reduce the need for emotional labour, have no personal attitudes (but, biases nonetheless), reduce the need for gaze work, and are generally free of charge and accessible around the clock. Another notable form of resilience is the mainstream adoption of these technologies: hearing people are also integrating them into their practices. Many hearing people are used to speaking into a speech recognition app on a phone and a range of technologies now incorporate mainstream accessibility features such as live captions, vibrational alerts, and visual notifications such as flash alerts. Wearable technologies, such as smartwatches and emerging devices like smart glasses are gaining broader social acceptance. These devices blend function with fashion in a way that makes them “cool” (see also Profita et al., 2016). In contrast, the presence of a human interpreter is a very visible form of mediation and often carries a social stigma for deaf people, with interpreters being perceived as family members or caregivers, or their performance leading to negative assumptions about deaf peoples’ capabilities (Feyne, 2018; Heldens & Gent, 2020). Yet, these perceived

benefits must be critically examined against their trade-offs. Crip Technoscience offers a lens to do this, interrogating not just how these tools are being used but also how they redistribute access and reshape social hierarchies.

5. Access and crip technoscience

Crip Technoscience (Hamraie & Fritsch, 2019) offers a critical lens to understand the complex, often ambiguous relationship between disabled people and technology. This may mean disabled people adopt imperfect, uncomfortable, or less functional tools, while simultaneously critiquing their limitations (notice the parallels here with deaf people's ambiguous relationships with sign language interpreters). They position themselves as both users and design experts, hacking, adapting, and reimagining existing technologies to fit their desires and needs (Hamraie, 2023). Using a similar lens, recent interdisciplinary deaf-led work in Human-Computer Interaction (HCI) is exploring which technologies – and what futures – are truly *worth wanting* for deaf communities (Angelini, 2024; Angelini, Spiel, & De Meulder, 2024a, 2024b).

Crip Technoscience provides a critical lens for understanding deaf experiences with AI, particularly in the complex and often tense relationship between disability and technology. For example, automated captions may benefit some deaf people with strong written language skills (in dominant languages), but disadvantage others without such skills. Technologies may not work effectively in certain contexts, such as remote locations. They can lead to social inequalities between deaf people, because not all can pay for an (often more performant) AI model. Remote access can benefit some, while at the same time 'distantism' (the societal bias that prioritizes the distance senses of sight and hearing over touch) creates accessibility barriers for deafblind people (Clark, 2021). In a future projection users of larger, better institutionalized sign languages like American Sign Language (ASL) or Libras (the majority sign language in Brazil) might benefit more from sign language technologies than those sign languages with much smaller datasets available. They might also privilege some varieties of these sign languages over others, and disadvantage those deaf users with idiosyncratic signing that might influence their interaction with sign language AI.

This, and especially the combination with access (or lack of access) to human sign language interpreters, creates new hierarchies of accessibility, with "the same conditions enabling access for some can create hierarchies of belonging for others" (Hamraie, 2023, p. 310). This positions communication access as a collective entity grounded in interdependency (Bennett, Brady, & Branham, 2018; McDonnell & Findlater, 2024). Access is co-constructed: changes in access practices, such as new technologies, often redistribute access in ways that improve opportunities for some, while potentially creating barriers for others. AI language technologies highlight these dynamics. While they may reduce stigma and access labour associated with working with human interpreters, this resilience also must be seen against the unavoidable trade-offs that accompany their adoption. These trade-offs include the fact that, like human interpreters, AI systems are prone to bias, including ableism, racism, and gender bias, which are deeply embedded in the data and algorithms that underpin them (Broussard, 2023; Börstell, 2023; Glazko et al., 2023). AI tools are also prone to hallucinating – Hicks, Humphries, and Slater (2024) argue that "bullshitting" is a more fitting description – generating misleading outputs that may compromise trust. Additionally, while some of these tools are freely

available, they often come with considerable ethical concerns. These include granting large corporations access to sensitive personal data and uncertainty about how this data is stored, processed and used to train models. This can lead to a lack of transparency and accountability. Furthermore, the ecological footprint and the “ghost work” involved in AI development and deployment remains a growing concern (De Vries, 2023; Gray & Suri, 2019).

Additionally, the line between resilience and breakdown is razor-thin: when the limits of technology are reached, these tools risk becoming marginalizing. For example, speech recognition apps may fail to recognise ‘deaf accents’, referring to the distinctive speech patterns of some deaf people, or regional dialects spoken by hearing people into a phone. In a future projection, sign language AI may encounter similar challenges in understanding regional sign languages, sign language dialects (e.g. Black ASL, which is significantly underrepresented in datasets), sociolinguistic variation (e.g. queer signing, influence of International Sign) or idiosyncratic signing styles, such as those resulting from additional disabilities, or just lack of access to signing spaces (Haualand et al., 2024). Those who can calibrate their signing to be recognised by AI sign language technologies often mirror the same group accustomed to adapting to human interpreters. This dual adaptability privileges already privileged “normative” users of well-resourced sign languages, potentially widening existing inequalities.

6. Deaf leadership in sign language AI research

These dynamics – access labour, hierarchies of accessibility, and the collective nature of access – highlight the need to study the socio-political contexts in which both these tools, and human sign language interpreting services, operate (Erdocia, Migge, & Schneider, 2024; Sayers, Sousa-Silva, & Höhn, 2021; Seargeant, 2023). Currently, the field of AI sign language is dominated by non-disabled, often non-signing, actors. While some of their work reflects systemic biases such as technochauvinism (the belief that technology is the best solution for social problems), technoableism (the drive to “fix” disabled bodies), and modality chauvinism (the privileging of speech over sign) (Broussard, 2023; Henner & Robinson, 2023; Shew, 2023), other challenges stem from plain ignorance and working in disciplinary silos. Many researchers and developers lack meaningful engagement with deaf communities. These embedded biases shape designs like smart gloves that translate signs to text or speech (Lu, 2016). This foregrounds the belief that deaf people’s ideas make more sense if converted to spoken form (Hill, 2013), or that deaf people are not worth talking to since there is no assumption of bilateral engagement, as the gloves do not translate the hearing person’s speech back to sign. Similarly, AI tools are often designed with normative bodies and minds in mind: voice assistants like Alexa or Siri often fail to account for deaf accents (Devries et al., 2024; Tran et al., 2024).

A systematic review conducted by Desai et al. (2024) based on 101 papers in AI sign language research, found that while calls for the ethical development of sign language AI (e.g. Bragg et al., 2021; De Meulder, 2021; Fox et al., 2023) have helped some individual researchers improve their practices, the field continues to face systemic biases. These include an overfocus on addressing perceived communication barriers, a lack of use of representative datasets, reliance on annotations lacking linguistic foundations, and the development of methods built on flawed models. To echo Broussard (2023), these are

‘more than just glitches’: these biases are now baked-in as standards, and risk causing harm by marginalising the very target users of sign language AI.

For a large part, these biases are introduced, magnified, and sustained by a lack of deaf leadership in the field of sign language AI (a problem extending to nearly every single field and discipline dealing with deaf people and sign languages). Desai et al. (2024) found that the field is mainly driven by what hearing, abled, researchers perceive as the most urgent problems and by decisions that they perceive as the most convenient. This also significantly impacts on how co-creation and co-design are filled in. While these concepts have become central in technology development and HCI, both in research and in policy (cf. Official Journal of the European Union, n.d., the EU 2024 AI Act and 2019 Ethics Guidelines), they often appear to function more as strategies to secure funding than as genuinely well-integrated principles in research design and development. “Co-creation” often serves as user studies or data collection methods, without fundamentally challenging the status quo for disabled people. In some cases, they verge on exploitation (De Meulder et al., 2024; Raman & French, 2022) or are perceived as “participation washing” (Sloane et al., 2020), referring to the superficial or performative involvement of participants in these processes. In large research consortia focused on AI and people with disabilities, I often find myself to be the only disabled researcher in the room. There is a growing acknowledgment of this under-representation of deaf researchers in the field. Shterionov et al. (2024, p. 18) acknowledge the “collective responsibility to ensure that we create pathways to the field that welcomes, supports, and encourages deaf scholars. Indeed, ideally, deaf researchers would be leading work on SLMT”. However, this underrepresentation is often acknowledged in a detached manner, with researchers presenting it as an unfortunate reality rather than a system issues that requires their active intervention. Many projects, including those emphasising co-creation, continue to be funded without deaf Principal Investigators, while the predominant response remains one of lamenting the lack of deaf researchers rather than addressing the structural barriers that contribute to this lack. Moreover, the growth in the number of AI projects is far outpacing the increase in the number of deaf researchers in the same field. This situation differs from what sign linguistics once faced (and perhaps still does) regarding a similar lack of representation, as the development (and funding) in that field progressed at a slower pace. In contrast, AI is advancing at an unprecedented rate.

7. Discussion and conclusion

This paper examined the complex interplay of AI language technologies in the context of deaf people, sign language interpreting, and linguistic access. It highlights the interdependencies underpinning these access frameworks, the trade-offs inherent in these technologies, and the critical need to navigate these shifts carefully. The integration of AI language technologies into deaf accessibility practices brings opportunities and challenges, revealing the tension between innovation and the preservation of hard-won linguistic and accessibility rights. The tensions expressed by the European attendee in the introduction reflect one of the central challenges of this moment: navigating the promises of AI as a potential liberatory tool without undermining deaf communities’ hard-won linguistic rights and access frameworks.

A key issue identified in the article is the problematic framing of AI sign language technologies as a substitute for human interpreters. This framing perpetuates a bench-

mark that undermines the very purpose of these technologies, which are not intended to compete with human interpreters but to support and reflect the diverse languaging practices of deaf communities. At its core, the development of these tools intersects with broader inclusion ideologies centred on interpreting, often reinforcing access hierarchies and introducing new forms of linguistic subordination to technology. These dynamic raises critical questions about not only how these tools are used but also how they redistribute access and reshape social hierarchies.

The concerns from deaf users, sign language interpreters and representative deaf organisations extend beyond who creates AI sign language technologies, to how these tools will be used once they are ‘good enough’ for real-world applications, and how this may lead to the potential erosion of hard-won linguistic and accessibility rights, particularly those rooted in human interpreting services. One major concern is that technological solutions risk being imposed on deaf people in situations where AI tools are not warranted, and the overestimation of AI capabilities might lead to compromised access. Although current AI sign language technologies are mainly targeted at so-called “low-stakes” contexts like public transportation (David & Bouillon, 2018; van Gemert et al., 2022) and hospitality (Leeson et al., 2024), developers are also pursuing “higher-stakes” areas like healthcare (Esselink et al., 2023) and emergency services (Guo et al., 2023; Martin et al., 2013). This shift, if not carefully managed, may exacerbate existing inequalities and undermine trust in these AI systems. These concerns do not come from thin air, considering the cost-effective approach often pursued by governments to address linguistic diversity and the “slippery slope” of access rights in deaf communities. The growing reliance on AI tools in low-stakes contexts, such as public transportation and customer services, might initially be seen as acceptable. However, these small adaptations could expand into higher-stakes contexts like healthcare or legal proceedings, where the stakes for deaf people are much higher. At the same time, policy measures can help but, in some ways, also hinder the implementation of a human-in-command approach. For instance, the EU Accessibility Act, set to take effect in June 2025, requires public and private entities, including broadcasters, to ensure their services and content are accessible to people with disabilities, covering measures like subtitles, audio descriptions, and sign language translation. The “shortage of human sign language interpreters”, perceived or real, and the legal obligation to meet accessibility requirements is frequently used here as a justification for developing sign language avatars for live broadcasting, despite deaf community preferences. This dynamic again highlights the interplay of technoableism and the interconnected discourses surrounding technology and sign language interpreting, where technological solutions are imposed as a quick fix for systemic challenges, without addressing the root causes of inequality. Deaf-led research in sign language AI is essential to challenging these power dynamics, ensuring ethical practices, and preventing harm caused by misaligned technological solutions. This research must emphasise the collective nature of access, emphasising choice, agency, and human oversight to ensure alignment with the diverse desires and preferences of deaf communities.

AI language technologies are not inherently disruptive but also generative, offering practical and theoretical opportunities. In a practical sense, they may alleviate the strain on interpreting services, expand access options, and empower deaf users by promoting AI literacy. Resilience lies not only in the capabilities of these tools but also in the

capacity of deaf communities to adapt, innovate, and co-create alternative forms of access that reflect their values, needs, and aspirations. Theoretically, resilience lies in the development of new theoretical and methodological frameworks in Deaf Studies, Sign Language Interpreting Studies and sign language (socio)linguistics so they remain relevant and equitable in the AI era.

As highlighted in the introduction, technological shifts rarely occur overnight but instead embed themselves quietly into systems and practices. By the time their effects are fully visible, they may have already reshaped the accessibility landscape in profound and irreversible ways—both positively and negatively. Engaging critically with these shifts as they unfold is essential to fostering AI systems that are not only innovative and cutting-edge but also equitable, inclusive, and trustworthy, and serve rather than subordinate linguistic practices to technological imperatives.

References

- Albanie, S., Varol, G., Momeni, L., Bull, H., Afouras, T., Chowdhury, H., ... Zisserman, A. (2021). *BBC-Oxford British Sign Language Dataset*. arXiv. Retrieved 2025-05-26, from <https://arxiv.org/abs/2111.03635> doi: 10.48550/ARXIV.2111.03635
- Angelini, R. (2024). Deaf Tech Worth Wanting: A Participatory Speculative Investigation. Retrieved from <https://repositum.tuwien.at/handle/20.500.12708/198125> doi: 10.34726/HSS.2024.117205
- Angelini, R., Spiel, K., & De Meulder, M. (2024a). Bridging the Gap: Understanding the intersection of deaf and technical perspectives on signing avatars. In A. Way, D. Shterionov, C. Rathmann, & L. Leeson (Eds.), *Sign Language Machine Translation*. Springer.
- Angelini, R., Spiel, K., & De Meulder, M. (2024b). Experiencing deaf tech: A deep dive into the concept of deafWatch. In *The 26th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 1–4). ACM. Retrieved from <https://dl.acm.org/doi/10.1145/3663548.3688483>
- Angelini, R., Tran, N., De Meulder, M., & Spiel, K. (in prep). Exploring Interactions with ChatGPT by Deaf Users in Europe.
- Backus, A., Cohen, M., Cohn, N., Faber, M., Kraemer, E., Laparle, S., ... De Vos, C. (2023). Minds: Big questions for linguistics in the age of AI. *Linguistics in the Netherlands*, 40, 301–308. Retrieved from <http://www.jbe-platform.com/content/journals/10.1075/avt.00094.bac> doi: 10.1075/avt.00094.bac
- Bennett, C. L., Brady, E., & Branham, S. M. (2018). Interdependence as a frame for assistive technology research and design. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 161–173). Galway Ireland: ACM. Retrieved from <https://dl.acm.org/doi/10.1145/3234695.3236348> doi: 10.1145/3234695.3236348
- Bragg, D., Caselli, N., Hochgesang, J. A., Huenerfauth, M., Katz-Hernandez, L., Koller, O., ... Ladner, R. E. (2021). The FATE landscape of Sign Language AI datasets: An interdisciplinary perspective. *ACM Transactions on Accessible Computing*, 14(2), 1–45. Retrieved from <https://dl.acm.org/doi/10.1145/3436996> doi: 10.1145/3436996
- Broussard, M. (2023). *More than a Glitch: Confronting Race, Gender, and Ability Bias in Tech*. The MIT Press. Retrieved from <https://direct.mit.edu/books/book/5555/>

- More-than-a-GlitchConfronting-Race-Gender-and doi: 10.7551/mitpress/14234.001.0001
- Burke, T. B. (2017). Choosing accommodations: Signed language interpreting and the absence of choice. *Kennedy Institute of Ethics Journal*, 27(2), 267–299. Retrieved from <https://muse.jhu.edu/article/665108> doi: 10.1353/ken.2017.0018
- Börstell, C. (2023). Ableist Language Teaching over Sign Language Research. In *Second Workshop on Resources and Representations for Under-Resourced Languages and Domains* (pp. 1–10).
- Caselli, N. K., Hall, W. C., & Henner, J. (2020). American Sign Language interpreters in public schools: An illusion of inclusion that perpetuates language deprivation. *Maternal and Child Health Journal*, 24(11), 1323–1329. Retrieved from <https://link.springer.com/10.1007/s10995-020-02975-7> doi: 10.1007/s10995-020-02975-7
- Chua, M., De Meulder, M., Geer, L., Henner, J., Hou, L., Kubus, O., ... Robinson, O. (2022). 1001 Small victories: Deaf academics and imposter syndrome. In M. Addison, M. Breeze, & Y. Taylor (Eds.), *The Palgrave Handbook of Imposter Syndrome in Higher Education* (pp. 481–496). Cham: Springer International Publishing. Retrieved from https://link.springer.com/10.1007/978-3-030-86570-2_29 doi: 10.1007/978-3-030-86570-2_29
- Clark, J. L. (2021). *Against Access*. Retrieved from https://audio.mcsweeneys.net/transcripts/against_access.html
- Coy, A., Mohammed, P. S., & Skerit, P. (2024). Inclusive deaf education enabled by artificial intelligence: The path to a solution. *International Journal of Artificial Intelligence in Education*, 35(1), 96–134. Retrieved from <https://link.springer.com/10.1007/s40593-024-00419-9> doi: 10.1007/s40593-024-00419-9
- Crawley, V., & O'Brien, D. (2020). Multilingual and multicultural supervision meetings: The case of a deaf supervisor and hearing postgraduate researcher. *The Learner Development Journal*, 4, 6–19.
- David, B., & Bouillon, P. (2018). Prototype of automatic translation to the sign language of French-speaking Belgium evaluation by the deaf community. *Modelling, Measurement and Control C*, 79(4), 162–167. Retrieved from http://www.iieta.org/journals/mmc_c/paper/10.18280/mmc_c.790402 doi: 10.18280/mmc_c.790402
- De Coster, M., Shterionov, D., Van Herreweghe, M., & Dambre, J. (2023). Machine translation from signed to spoken languages: state of the art and challenges. *Universal Access in the Information Society*, 23(3), 1305–1331. Retrieved from <https://link.springer.com/10.1007/s10209-023-00992-1> doi: 10.1007/s10209-023-00992-1
- De Meulder, M. (2021). Is “good enough” good enough? Ethical and responsible development of sign language technologies. In *Proceedings of the 18th Biennial Machine Translation Summit Proceedings of the 18th Biennial Machine Translation Summit 1st International Workshop on Automatic Translation for Signed and Spoken Languages* (pp. 12–22). Retrieved from <https://aclanthology.org/2021.mtsummit-at4ssl.2/>
- De Meulder, M., & Haualand, H. (2021, May). Sign language interpreting services: A quick fix for inclusion? *Translation and Interpreting Studies*, 16(1), 19–40. Retrieved from <http://www.jbe-platform.com/content/journals/10.1075/tis.18008.dem> doi: 10.1075/tis.18008.dem
- De Meulder, M., & Murray, J. J. (2017). Buttering their bread on both sides?: The

- recognition of sign languages and the aspirations of deaf communities. *Language Problems and Language Planning*, 41(2), 136–158. Retrieved from <http://www.jbe-platform.com/content/journals/10.1075/lplp.41.2.04dem>
- De Meulder, M., & Murray, J. J. (2021). The illusion of choice in inclusive education. *International Journal of Inclusive Education*.
- De Meulder, M., Murray, J. J., & McKee, R. L. (2019). *The Legal Recognition of Sign Languages: Advocacy and Outcomes Around the World*. Multilingual Matters. Retrieved from <https://www.degruyter.com/document/doi/10.21832/9781788924016-002/html> doi: 10.21832/9781788924016-002
- De Meulder, M., Napier, J., & Stone, C. (2018). Designated or preferred? A deaf academic and two signed language interpreters working together for a PhD defence: A case study of best practice. *International Journal of Interpreter Education*, 10(2), 5–26. Retrieved from <https://www.cit-asl.org/new/ijie-10-2-designated-or-preferred/>
- De Meulder, M., & Stone, C. (2024). Finally Free from the Interpreter's Gaze? Uncovering the Hidden Labor of Gaze Work for Deaf Consumers of Interpreter Services. *International Journal of Interpreter Education*, 15(1), 87–98. Retrieved from <https://open.clemson.edu/ijie/vol15/iss1/8>
- De Meulder, M., Van Landuyt, D., & Omardeen, R. (2024). *Lessons in co-creation: The inconvenient truths of inclusive sign language technology development*. arXiv. Retrieved 2025-05-26, from <https://arxiv.org/abs/2408.13171> doi: 10.48550/ARXIV.2408.13171
- Desai, A., De Meulder, M., Hochgesang, J. A., Kocab, A., & Lu, A. X. (2024). Systemic biases in Sign Language AI research: A deaf-led call to reevaluate research agendas. In *LREC-COLING 2024, 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources* (pp. 54–65). Torino, Italia: ELRA and ICCL. Retrieved from <https://aclanthology.org/2024.signlang-1.6/>
- De Vries, A. (2023). The growing energy footprint of artificial intelligence. *Joule*, 7(10), 2191–2194. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S2542435123003653> doi: 10.1016/j.joule.2023.09.004
- Devries, P. S., Tran, N., Delk, K., Miga, M., Taulbee, R., Pidathala, P., ... Vogler, C. (2024). *Sign Language-Based versus Touch-Based Input for Deaf Users with Interactive Personal Assistants in Simulated Kitchen Environments*.
- DIK. (2024). *Språkbransch i förändring - en rapport om arbetsmiljö, arbetsvillkor och AI:s påverkan på språkbranschen (A changing language industry – a report on the work environment, working conditions and AI's impact on the language industry)*. (Tech. Rep.). Retrieved from <https://dik.se/om-oss/nyheter/sprakbransch-i-forandring-en-rapport-om-ai-och-sprakyrkenas-framtid>
- Duggan, N., & Holmström, I. (2024). Deaf migrants in Sweden: exploring linguistic and bureaucratic challenges through the lens of Crip Theory and Crip Linguistics. *Multilingua*, 43(5), 667–693. Retrieved from <https://www.degruyter.com/document/doi/10.1515/multi-2023-0203/html> doi: 10.1515/multi-2023-0203
- Erdocia, I., Migge, B., & Schneider, B. (2024). Language is not a data set—Why overcoming ideologies of dataism is more important than ever in the age of AI. *Journal of Sociolinguistics*, 28(5), 20–25. Retrieved from <https://onlinelibrary.wiley.com/doi/10.1111/josl.12680> doi: 10.1111/josl.12680
- Esselink, L., Roelofsen, F., Dotlačil, J., Mende-Gillings, S., De Meulder, M., Sijm, N., &

- Smeijers, A. (2023). Exploring automatic text-to-sign translation in a healthcare setting. *Universal Access in the Information Society*, 23(1), 35–57. Retrieved from <https://link.springer.com/10.1007/s10209-023-01042-6>
- Feyne, S. (2018). Variation in perception of the identity of interpreted Deaf lecturers. In C. Stone & L. Leeson (Eds.), *Interpreting and the politics of recognition* (Vol. 7, pp. 119–137). Routledge.
- Fox, N., Woll, B., & Cormier, K. (2023). Best practices for sign language technology research. *Universal Access in the Information Society*, 24(1), 69–77. Retrieved from <https://link.springer.com/10.1007/s10209-023-01039-1> doi: 10.1007/s10209-023-01039-1
- Giambruno, G. (2014). Dealing with languages of lesser diffusion. In G. Giambruno (Ed.), *Assessing Legal Interpreter Quality through Testing and Certification: the Qualitas Project* (pp. 93–108). Sant Vincent del Raspeig: Alicante Publications.
- Glazko, K. S., Yamagami, M., Desai, A., Mack, K. A., Potluri, V., Xu, X., & Mankoff, J. (2023, October). An Autoethnographic Case Study of Generative Artificial Intelligence’s Utility for Accessibility. In *The 25th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 1–8). New York NY USA: ACM. Retrieved from <https://dl.acm.org/doi/10.1145/3597638.3614548>
- Gray, M., & Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books.
- Gunarhadi, Wiliyanto, D. A., Anggarani, F. K., & Anggrellangi, A. (2024). Virtual reality as a solution: Meeting the needs of deaf students with digital sign language interpreters. *Edelweiss Applied Science and Technology*, 8(5), 1189–1199. Retrieved from <https://learning-gate.com/index.php/2576-8484/article/view/1823>
- Guo, Y., Zhao, J., Ding, B., Tan, C., Ling, W., Tan, Z., ... Lu, S. (2023). Sign-to-911: Emergency Call Service for Sign Language Users with Assistive AR Glasses. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking* (pp. 1–15). Madrid Spain: ACM. Retrieved from <https://dl.acm.org/doi/10.1145/3570361.3613260> doi: 10.1145/3570361.3613260
- Hall, W. C., Brick, K., & Millios, W. (2024). From principles to practice: Creating language equity for deaf professionals. *Journal of Interpretation*, 32(1).
- Hamraie, A. (2023). Crip Authorship. Disability as Method. In M. Mills & R. Sanchez (Eds.), *Crip Making* (pp. 303–317). New York University Press.
- Hamraie, A., & Fritsch, K. (2019). Crip Technoscience Manifesto. *Catalyst: Feminism, Theory, Technoscience*, 5(1), 1–33. Retrieved from <https://catalystjournal.org/index.php/catalyst/article/view/29607> doi: 10.28968/cftt.v5i1.29607
- Haualand, H., De Meulder, M., & Napier, J. (2023). Unpacking sign language interpreting as a social institution: The missing macro perspective? *Translation and Interpreting Studies*, 17(3), 351–358. Retrieved from <http://www.jbe-platform.com/content/journals/10.1075/tis.00055.hau> doi: 10.1075/tis.00055.hau
- Haualand, H., Hjulstad, J., Ingjerd Legreid Ødemark, & Kermit, P. (2024). Where can I sign my language?: A systematic literature review (2000–2019) of young people’s access to sign Language socialisation spaces in the Nordic countries. *Disability Studies Quarterly*, 44(1). Retrieved from <https://dsq-sds.org/index.php/dsq/article/view/8409> doi: 10.18061/dsq.v44i1.8409
- Heldens, T., & Gent, P. V. (2020). Echoing attitudes: How the regional accent of the

- interpreter affects the perception of a deaf signer. *Sign Language Studies*, 21(1), 5–31. Retrieved from <https://muse.jhu.edu/article/773578> doi: 10.1353/sls.2020.0027
- Henner, J., & Robinson, O. (2023). Crip linguistics goes to school. *Languages*, 8(1), 48. Retrieved from <https://www.mdpi.com/2226-471X/8/1/48>
- Hicks, M. T., Humphries, J., & Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology*, 26(2), 38. Retrieved from <https://link.springer.com/10.1007/s10676-024-09775-5> doi: 10.1007/s10676-024-09775-5
- Hill, J. (2013). Language ideologies, policies, and attitudes toward signed languages. In R. Bayley, R. Cameron, & C. Lucas (Eds.), *The Oxford Handbook of Sociolinguistics* (1st ed., pp. 680–698). Oxford University Press. Retrieved from <https://academic.oup.com/edited-volume/28007/chapter/211782313> doi: 10.1093/oxfordhb/9780199744084.013.0033
- Hill, J., Isakson, S. K., & Nakahara, C. (2022). Infusing social Justice in interpreting education. In L. Swabey & R. E. Herring (Eds.), *Signed Language Interpreting Pedagogy: Insights and Innovations from the Conference of Interpreter Trainers* (pp. 394–407). Gallaudet University Press.
- Huffman, S., Chen, S., Mack, K. A., Su, H., Wang, Q., & Kushalnagar, R. (2024). "We do use it, but not how hearing people think": How the Deaf and Hard of Hearing Community Uses Large Language Model Tools. arXiv. Retrieved from <https://arxiv.org/abs/2410.21358> doi: 10.48550/ARXIV.2410.21358
- James, T. G., Coady, K. A., Stacciarini, J.-M. R., McKee, M. M., Phillips, D. G., Maruca, D., & Cheong, J. (2022). "They're not willing to accommodate deaf patients": Communication experiences of deaf American Sign Language users in the emergency department. *Qualitative Health Research*, 32(1), 48–63. Retrieved from <https://journals.sagepub.com/doi/10.1177/10497323211046238>
- Kelly-Holmes, H. (2024). Artificial intelligence and the future of our sociolinguistic work. *Journal of Sociolinguistics*, 28(5), 3–10. Retrieved from <https://onlinelibrary.wiley.com/doi/10.1111/josl.12678> doi: 10.1111/josl.12678
- Kushalnagar, P., Paludnevičienė, R., & Kushalnagar, R. (2019). Video Remote Interpreting Technology in Health Care: Cross-Sectional Study of Deaf Patients' Experiences. *JMIR Rehabilitation and Assistive Technologies*, 6(1), e13233. Retrieved from <http://rehab.jmir.org/2019/1/e13233/> doi: 10.2196/13233
- Kusters, A., Green, M., Moriarty, E., & Snoddon, K. (2020). Sign language ideologies: Practices and politics. In A. Kusters, M. Green, E. Moriarty, & K. Snoddon (Eds.), *Sign Language Ideologies in Practice* (pp. 3–22). De Gruyter. Retrieved from <https://www.degruyter.com/document/doi/10.1515/9781501510090-001/html> doi: 10.1515/9781501510090-001
- Lee, Y., Hill, J., & Smith, A. (2021). The Challenge of Preserving Captured Sign Language Data in Human Avatar Models. *Frameless*, 3(1).
- Leeson, L., Morrissey, S., Shterionov, D., Stein, D., Van Den Heuvel, H., & Way, A. (2024). How It Started and How It's Going: Sign Language Machine Translation and Engagement with Deaf Communities Over the Past 25 Years. In A. Way, L. Leeson, & D. Shterionov (Eds.), *Sign Language Machine Translation* (Vol. 5, pp. 27–72). Cham: Springer Nature Switzerland. Retrieved from https://link.springer.com/10.1007/978-3-031-47362-3_2 doi: 10.1007/978-3-031-47362-3_2
- Lu, A. (2016). Deaf people don't need new communication tools — everyone else

- does. Retrieved from <https://medium.com/@alexijie/deaf-people-dont-need-new-communication-tools-everyone-else-does-df83b5eb28e7>
- Maaß, L. M. (2024). Managing the shortage: A quantitative-qualitative study on the scale and effects of the severe lack in sign language interpreting resource in the German market. *VAKKI Publications*, 16. Retrieved from <https://journal.fi/vakki/article/view/145616> doi: 10.70484/vakki.145616
- Marie, A. (2019, February). Enacting Dependence. *Somatosphere*. Retrieved from <https://somatosphere.com/2019/enacting-dependence.html/>
- Marie, A., & Friedner, M. (2021). Entangled Interdependence: Sign Language Interpreting without Recognition in India and Vietnam. *PoLAR: Political and Legal Anthropology Review*, 44(2), 192–206. Retrieved from <https://anthrosource.onlinelibrary.wiley.com/doi/10.1111/plar.12450> doi: 10.1111/plar.12450
- Martin, M., Belhe, S., Mudliar, S., Kulkarni, M., & Sahasrabudhe, S. (2013). An Indian Sign Language (ISL) Corpus of the Domain Disaster Message using Avatar. In *Proceedings of the Third International Symposium in Sign Language Translations and Technology*.
- McDonnell, E. J., & Findlater, L. (2024). Envisioning Collective Communication Access: A Theoretically-Grounded Review of Captioning Literature from 2013-2023. St. John's, NL, Canada.
- Napier, J. (2011). “It’s not what they say but the way they say it”. A content analysis of interpreter and consumer perceptions towards signed language interpreting in Australia. *International Journal of the Sociology of Language*, 2011(207). Retrieved from <https://www.degruyter.com/document/doi/10.1515/ijsl.2011.003/html> doi: 10.1515/ijsl.2011.003
- Napier, J., McKee, R. L., & Goswell, D. (2010). *Sign language interpreting: theory and practice in Australia and New Zealand* (2nd edition ed.). Annandale, NSW: The Federation Press.
- Napier, J., Skinner, R., Adam, R., Stone, C., Pratt, S., Hinton, D. P., & Obasi, C. (2022). Representation and diversity in the sign language translation and interpreting profession in the United Kingdom. *Interpreting and Society: An Interdisciplinary Journal*, 2(2), 1–22. Retrieved from <https://journals.sagepub.com/doi/10.1177/27523810221127596> doi: 10.1177/27523810221127596
- Nasution, A. K., & Batubara, M. H. (2024). Research trends in technology-enhanced language learning: A bibliometric analysis. *Journal of Linguistics, Literature and Language Teaching (JLLLT)*, 3(2), 111–130.
- O’Brien, D., Hodge, G., Gulamani, S., Rowley, K., Adam, R., Emery, S., & Walker, J. (2023). Deaf professionals’ perceptions of ‘trust’ in relationships with signed language interpreters. *The International Journal of Translation and Interpreting Research*, 15(2), 25–42. Retrieved from <http://www.trans-int.org/index.php/transint/article/view/1717/454> doi: 10.12807/ti.115202.2023.a02
- Official Journal of the European Union. (n.d.). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/1139 and (EU) 2019/2144 and Directives 214/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*. Retrieved from <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng#:~:text=Regulation%20%>

- 2D%20EU%20%2D%202024%2F1689%20%2D%20EN%20%2D%20EUR%2DLex
 Picron, F., Van Landuyt, D., Omardeen, R., Efthimiou, E., Wolfe, R., Fotinea, S. E., ...
 Hanke, T. (2024). The EASIER Mobile Application and Avatar End-User Evaluation
 Methodology. In *11th Workshop on the Representation and Processing of Sign
 Languages* (pp. 276–281). ELRA Language Resources Association.
- Piller, I. (2017). *Linguistic Diversity and Social Justice: An Introduction to Applied Sociolin-
 guistics*. New York: Oxford University Press USA - OSO.
- Profita, H., Albaghli, R., Findlater, L., Jaeger, P., & Kane, S. K. (2016). The AT effect: How
 disability affects the perceived social Acceptability of Head-Mounted Display Use.
 In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*
 (pp. 4884–4895). San Jose California USA: ACM. Retrieved from [https://dl.acm.org/
 doi/10.1145/2858036.2858130](https://dl.acm.org/doi/10.1145/2858036.2858130)
- Rainò, P. (2022). Saturation or shortage? Does sign language interpreter education
 in Finland match the demand in the labour market? In L. Segler-Heikkilä &
 L. Halkosaari (Eds.), *The many faces of communication* (pp. 56–70). HUMAK.
- Raman, S., & French, T. (2022). Towards a shared understanding of genuine co-design
 with people with lived experience: reflections from co-designing for relational and
 transformational experiences in health and social care in the UK. In S. Morton
 (Ed.), *Designing Interventions to Address Complex Societal Issues*. Routledge.
- Robinson, O. E., Sheneman, N., & Henner, J. (2020). Toxic Ableism Among Interpreters:
 Impeding deaf people’s linguistic rights through pathological posturing. In C. Mc-
 Dermid, S. Ehrlich, & A. Gentry (Eds.), *Conference proceedings of the 2019 WASLI
 conference* (pp. 14–41). WASLI.
- Russell, D. (2021). Critical Perspectives on Education Mediated by Sign Language
 Interpreters: Inclusion or the Illusion of Inclusion? In K. Snoddon & J. Weber (Eds.),
 (pp. 217–239). Multilingual Matters.
- Sayers, D., Sousa-Silva, R., & Höhn, S. (2021). *The Dawn of the Human-Machine Era: A
 forecast of new and emerging language technologies* (Report for EU COST Action
 CA19102 ‘Language in the Human-Machine Era’). Open Science Centre, University
 of Jyväskylä. Retrieved from <https://jyx.jyu.fi/handle/123456789/75737> doi: 10
 .17011/jyx/reports/20210518/1
- Seargeant, P. (2023). *The future of language: how technology, politics and utopianism
 are transforming the way we communicate*. London New York Oxford New Delhi
 Sydney: Bloomsbury Academic.
- Sheneman, N., & Robinson, O. (2021). Helpers, professional authority, and pathologized
 bodies. Ableism in interpretation and translation. In O. Carbonell i Cortés &
 E. Monzó-Nebot (Eds.), *Translating Asymmetry – Rewriting Power* (pp. 55–76).
 Benjamins Translation Library.
- Shew, A. (2023). *Against Technoableism: Rethinking Who Needs Improvement (a Norton
 Short)*. Erscheinungsort nicht ermittelbar: W. W. Norton & Company, Incorporated.
- Shterionov, D., Leeson, L., & Way, A. (2024). The pipeline of sign language machine
 translation. In A. Way, L. Leeson, & D. Shterionov (Eds.), *Sign Language Machine
 Translation* (Vol. 5, pp. 1–25). Cham: Springer Nature Switzerland. Retrieved
 from https://link.springer.com/10.1007/978-3-031-47362-3_1 doi: 10.1007/978-3-
 031-47362-3_1
- SignOn Consortium. (2023). *Sign Language Translation Mobile Application and Open Com-*

- munications Framework* (Tech. Rep. No. Deliverable 7.4: Final Report). Retrieved from <https://signon-project.eu/publications/public-deliverables/>
- Sivunen, N. (2019). An ethnographic study of deaf refugees seeking asylum in Finland. *Societies*, 9(1), 2. Retrieved from <https://www.mdpi.com/2075-4698/9/1/2> doi: 10.3390/soc9010002
- Skaaden, H., & Wadensjö, C. (2014). Some considerations on the testing of interpreting skills. In C. Giamb Bruno (Ed.), *Assessing Legal Interpreter Quality through Testing and Certification: the Qualitas project* (pp. 17–26). Sant Vincent del Raspeig: Alicante Publications.
- Sloane, M., Moss, E., Awomolo, O., & Forlano, L. (2020). Participation is not a design fix for machine learning. In *Proceedings of the 37th International Conference on Machine Learning*. Vienna, Austria: PMLR 119.
- Stone, C. (2010). Access all areas – Sign language interpreting, is it that special? *The Journal of Specialized Translation*, 14, 41–54.
- Stone, C., Adam, R., Quadros, R. M. d., & Rathmann, C. (Eds.). (2022). *The Routledge handbook of sign language translation and interpreting*. London New York: Routledge, Taylor & Francis Group. doi: 10.4324/9781003019664
- (Towards) Sign Language Machine Translation in Practice. (n.d.). In *Machine translation: Technologies and applications: Sign language machine translation* (Vol. 5). Cham. Retrieved from https://link.springer.com/10.1007/978-3-031-47362-3_14
- Tran, N., Kushalnagar, R., Devries, P. S., Glasser, A., Seita, M., & Vogler, C. (2024). Assessment of sign language-based versus touch-based input for deaf users interacting with intelligent personal assistants. In *CHI Conference on Human Factors in Computing Systems*.
- Vandeghinste, V., De Sisto, M., Gómez, S. E., & De Coster, M. (2024). Challenges with Sign Language Datasets. In A. Way, L. Leeson, & D. Shterionov (Eds.), *Sign Language Machine Translation* (Vol. 5, pp. 117–139). Cham: Springer Nature Switzerland. Retrieved from https://link.springer.com/10.1007/978-3-031-47362-3_5 doi: 10.1007/978-3-031-47362-3_5
- van Gemert, B., Cokart, R., Esselink, L., De Meulder, M., Sijm, N., & Roelofsen, F. (2022). First steps towards a signing avatar for railway travel announcements in the Netherlands. In *7th International Workshop on Sign Language Translation and Avatar Technology (SLTAT): The Junction of the Visual and the Textual: Challenges and Perspectives* (pp. 109–116). Marseille, France.
- Webb, S., Napier, J., & Adam, R. (Eds.). (2025). *International Perspectives on Sign Language Translator and Interpreter Education*. Washington, D. C: Gallaudet University Press.
- World Federation of the Deaf & World Association of Sign Language Interpreters. (2018). *WFD and WASLI statement on the use of signing avatars*. Retrieved from <https://wfdeaf.org/wp-content/uploads/2017/02/WFD-and-WASLI-Statement-on-Avatar-FINAL-14032018-Updated-14042018-1.pdf>
- Yin, K., Atwell, K., Hochgesang, J. A., & Alikhani, M. (2024). The importance of including signed languages in natural language processing. In A. Way, L. Leeson, & D. Shterionov (Eds.), *Sign Language Machine Translation* (Vol. 5, pp. 73–87). Cham: Springer Nature Switzerland. Retrieved from https://link.springer.com/10.1007/978-3-031-47362-3_3 doi: 10.1007/978-3-031-47362-3_3

- Young, A., Napier, J., & Oram, R. (2019). The translated deaf self, ontological (in)security and deaf culture. *The Translator*, 25(4), 349–368. Retrieved from <https://www.tandfonline.com/doi/full/10.1080/13556509.2020.1734165> doi: 10.1080/13556509.2020.1734165
- Young, A., Oram, R., & Napier, J. (2019). Hearing people perceiving deaf people through sign language interpreters at work: On the loss of self through interpreted communication. *Journal of Applied Communication Research*, 47(1), 90–110. Retrieved from <https://www.tandfonline.com/doi/full/10.1080/00909882.2019.1574018> doi: 10.1080/00909882.2019.1574018

The complexity of gender and language: Digitization of a physical board game deconstructing gender stereotypes

Dimitra Anastasiou ¹, Christian Moll ², Marie Gallais ³, Laurence Johannsen ⁴,
and Carole Blond-Hanten ⁵

¹Luxembourg Institute of Science and Technology (LIST), dimitra.anastasiou@list.lu

²Luxembourg Institute of Science and Technology (LIST), christian.moll@list.lu

³Luxembourg Institute of Science and Technology (LIST), marie.gallais@list.lu

⁴Luxembourg Institute of Science and Technology (LIST), laurence.johannsen@list.lu

⁵Luxembourg Institute of Socio-Economic Research (LISER),
Carole.Blond-Hanten@liser.lu

Abstract

In this paper, we present a physical board game, known as the Gender Game¹ (GG), and detail how we enhanced it with digital elements. The goal of the GG is to deconstruct gender stereotypes by highlighting a scientific approach: the GG is rooted in social sciences and aims to inform the public about gender equality by presenting scientific results, mainly from research projects. The game is played with several players seated around a giant board; players take turns to roll the dice, pick a card from the deck held by the moderator and move their pawn to a domain where stereotypes exist. After a question is shown on a screen, the moderator standing in the middle asks the team players a question and all players discuss the answer together. A core component of the game addresses gendered language directly, serving as a practical tool for promoting linguistic rights and raising awareness of how language shapes perceptions of gender and power. We conducted an exploratory study with generative artificial intelligence (AI) by submitting some of the GG questions to ChatGPT 3.5 and 4.0. In this paper, we discuss the results of this study. Generative AI and, in particular, Large Language Models (LLMs) are impactful tools for decision-making and process automation. However, bias is often found in these tools and is perpetuated in the generated content. Thus, an ethical and responsive usage of LLMs is of utmost importance. From this perspective, the GG is positioned as a proactive, human-centred intervention designed to address the root cause of the biases that AI amplifies.

Keywords: Gender stereotypes, digital divide, generative AI, board game.

Resumo

Neste artigo, apresentamos um jogo de tabuleiro físico, conhecido como "Gender Game" (GG), e detalhamos como o aprimoramos com elementos digitais. O objetivo do GG é desconstruir estereótipos de género, destacando uma abordagem científica: o GG tem origem nas Ciências Sociais e visa informar o público sobre a igualdade de género, apresentando resultados científicos, resultantes principalmente de projetos de investigação. É jogado com vários jogadores sentados em torno de um tabuleiro gigante; os jogadores revezam-se para lançar os dados, escolher uma carta do baralho, que deve estar na posse de um moderador, e mover o seu peão para um domínio onde existem estereótipos. Depois de uma pergunta ser exibida no ecrã, o moderador, que está no meio, faz uma pergunta aos jogadores da equipa e todos discutem a resposta juntos. Uma componente central do jogo aborda diretamente a linguagem de género, servindo como uma ferramenta prática para promover os direitos linguísticos e aumentar a consciencialização sobre como a linguagem molda as perceções de género e poder. Realizámos um estudo exploratório com inteligência artificial (IA) generativa, submetendo algumas das perguntas do GG ao ChatGPT 3.5 e 4.0. Neste artigo, discutimos os resultados desta experiência. A IA generativa e, em particular, os Grandes Modelos de Linguagem (LLMs) são ferramentas com grande impacto na tomada de decisões e automação de processos. No entanto, muitas vezes é encontrado um viés nessas ferramentas, que se perpetua no conteúdo gerado. Assim, o uso ético e crítico dos LLMs é de extrema importância. A partir dessa perspectiva, o GG posiciona-se como uma intervenção proativa e centrada no ser humano, projetada para abordar a causa raiz dos vieses que a IA amplifica. **Palavras-chave:** Estereótipos de género, exclusão digital, IA generativa, jogo de tabuleiro.

1. Introduction



Figure 1. Gender Game as a physical giant square board game (3x3m) working in Q&A mode

Gender equality, besides being a fundamental human right, is essential for achieving peaceful societies, with full human potential and sustainable development (United

Nations, 2020). UNESCO highlights that all forms of discrimination based on gender are violations of human rights. Together with its partners, it works to provide solutions to reduce inequalities in and through education, to empower women in science and technology for environmental action, to promote inclusion and combat gender-based violence, to bridge the digital gender divide and to support women's empowerment in crisis, emergency and early recovery contexts². Deconstructing gender stereotypes is essential because when we fail to question or challenge them, we risk reinforcing outdated norms that limit individuals based on their gender. These stereotypes not only shape attitudes and behaviors but also contribute to systemic inequalities—particularly in access to goods, services, and opportunities. Breaking down these stereotypes is a crucial step toward building a more inclusive, fair, and equitable society. Besides UNESCO, other examples of organizations which reinforce and promote gender equality are detailed below:

- **European Women's Lobby** (EWL)³: a non-governmental organization working on social issues in Europe. Its mission is to ensure that social rights and equality are respected, upheld and mainstreamed in EU policy-making. Uniting over 40 European-wide networks, EWL represents the voices of citizens and members of civil society (e. g. NGOs, community groups etc.) who fight for social justice, bringing their needs, concerns and challenges to the forefront of EU democracy.
- **Equality Now**⁴: a network combining legal expertise, advocacy, collaboration and resourcing to realize its vision of a more just and equal world for women and girls. In cooperation with local, national, regional and international partners, it uses the power of the law to establish legal equality, and take steps to end sexual violence, sexual exploitation and harmful practices.
- **WIDE+**⁵ (Women in Development Europe+): a European network of associations and activists that fights for women's rights, as part of a larger struggle for social justice, sustainable livelihoods and human rights. They advocate changing European policies that affect people both within and outside of Europe. WIDE+ promotes inclusive and intersectional feminist movement-building in Europe.
- **Stop the pushback on women's rights**⁶: a growing collective of women from around the world, united in response to the alarming resurgence of attacks on women's rights. Started as a conversation among a small group of women within the Geneva Global Health Hub (G2H2), the collective has grown and now includes the voices of women from around the world united in response to the alarming resurgence of attacks on all women's rights, transcending borders and affiliations. G2H2 continues to serve as a vital facilitator within Geneva's multilateral landscape.

Moreover, the Council of Europe's Gender Equality⁷ focuses on six strategic areas, one of which is "preventing and combating gender stereotypes and sexism". Under the authority of the Committee of Ministers, and building on the achievements in gender

²<https://www.unesco.org/en/gender-equality>

³<https://www.socialplatform.org>

⁴<https://equalitynow.org>

⁵<https://wideplus.org>

⁶<https://stopthepushback.org>

⁷<https://www.coe.int/en/web/genderequality/gec1>

equality, the Gender Equality Commission steers the Council of Europe's intergovernmental work in this area. They advise the Committee of Ministers on appropriate action to be taken in its field of competence, taking the relevant transversal perspectives into account.

Furthermore, the Directorate-General for Communication of the European Commission provided an EU Roadmap for Women's Rights and published the 2025 report on gender equality in the EU⁸. According to this report, in May 2024, the first-ever Directive on combating violence against women and domestic violence, as well as two Directives on strengthening the role of equality bodies, were adopted.

From another European policy perspective, the European Institute for Gender Equality (EIGE) is composed of members from competent bodies specialized in gender equality issues from every Member State of the European Union. There are two members designated by the European Parliament, as well as three members designated by the European Commission. EIGE publishes the Gender Equality Index, which indicates levels of equality between women and men in the domains of violence, health, work, money, knowledge, power and time⁹. EIGE gives the EU and its member states a score from one to 100. A score of 100 would mean that a country has achieved full equality between women and men. In 2023, the EU surpassed 70 points for the first time, showing a growth of 1.6 points since 2022. The increase in the overall EU score is the highest year-on-year rise since the first edition of the Index in 2013.

While policy instruments like the EIGE Index are vital for measuring the outcomes of gender inequality, addressing the disparities they highlight requires their foundational causes to be tackled. Pervasive gender stereotypes are a primary driver of the imbalances seen across the EIGE's domains of work, money and power. These same stereotypes pose a profound risk in the digital era. When AI systems are trained on societal data, they inevitably learn and often amplify these embedded biases, which can lead to discriminatory outcomes in hiring, finance and beyond. Therefore, interventions that challenge and deconstruct stereotypes on a human level are more critical than ever, as they address the root of both societal and algorithmic bias. Based on EIGE domains as the main motivation, the GG was conceptualized to both 1) deconstruct gender stereotypes, including the promotion of a gender-inclusive language, and 2) promote science and research to a wider audience. The design of the GG, as shown in Figures 1 and 2, is based on three of the abovementioned EIGE domains: *work*, *knowledge* and *power*, as well as a fourth domain added specifically for Luxembourg: *society*. The current collaboration between two research institutions, the Luxembourg Institute of Science and Technology (LIST) and the Luxembourg Institute of Socio-Economic Research (LISER), provides advancements in the digitization of the GG. This paper outlines the digital elements that were added to the initial, physical version of the game. We present the main output of the project: a blended physical and digital gamified device, elaborated as an extension and development of the purely physical GG version. The core of the GG remains the physical game, but it is enhanced with digital elements, developed based on a systematic review of the game elements. The GG is an example of a physical game with integrated digital human computer interaction elements. At the same time, it is also a social situation-experiment that deconstructs gender stereotypes through the discus-

⁸https://commission.europa.eu/document/055fdbab-5786-425e-a072-652bf53d8fe4_en

⁹<https://eige.europa.eu/gender-equality-index/2023>

sions the between moderator and players. We also compare bias by a human versus a machine by asking ChatGPT some of the questions from the GG.

As mentioned above, the GG is a giant physical board game (nine square metres) that works in question and answer (Q&A) mode in four areas in which gender stereotypes persist according to the European Index on Gender Equality: work, knowledge, power and society (see Figure 2).



Figure 2. The four domains of the GG: work, knowledge, power and society

The four domains are described below:

- **Work:** The *work* domain represents access to employment for women and men and access to appropriate working conditions. The objective is to reduce poverty by providing equal access to economic resources.
- **Knowledge:** The field of *knowledge* represents the equal participation of women and men in education and training. Equal access to education and knowledge for girls and boys, and women and men enables balanced participation in the labour market, equal access to economic resources, and the realization of everyone's potential.
- **Power:** The *power* domain stands for the representation of women and men in decision-making positions in the political, social and economic fields. The objective of balanced representation refers to the equal access of women and men to decision-making positions.
- **Society:** The domain of *society* concerns the participation of women and men in society in Luxembourg. This field is composed of different sub-domains such as mobility, health, wealth, etc. The objective is to ensure the equal access and equal participation of women and men in order to enable a democratic society that allows everyone, both women and men, to attain economic and social well-being.

The game consists of 144 cards corresponding to 144 questions and answers prepared in advance; 48 were designed for children, 48 for youths and 48 for adults, with 12 questions per area for each age group (see two example questions in Figure 3). In the GG, the focus is not only on answering the question correctly, but also highlighting the gender-inclusive language used by the moderators both in the written form (of the questions and answers) and in their verbal discussion to promote linguistic justice by deconstructing stereotypes.

Because of the plurality of languages spoken in Luxembourg, the questions are prepared and presented in four languages (French, German, Luxembourgish, English), as can be seen in Fig.3, but the moderators interact with the children in the language in which they are the most familiar with. In order to avoid players selecting an answer to a question arbitrarily, the moderators always ask the players to justify why they provided this answer.

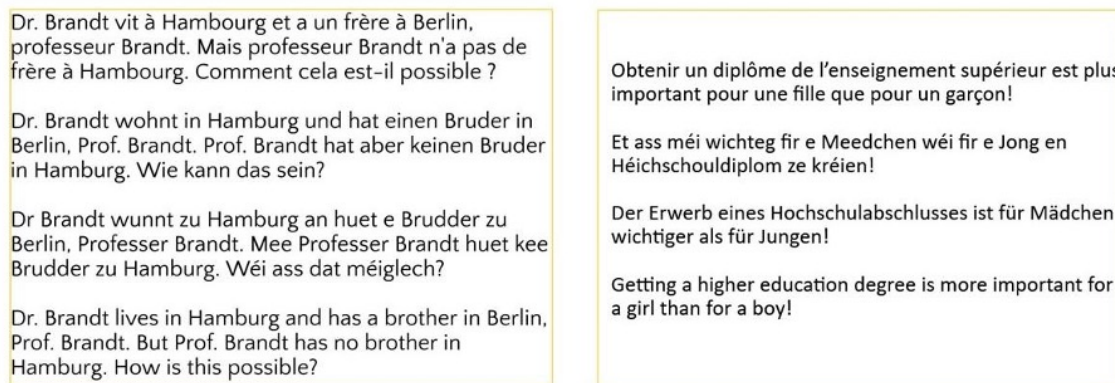


Figure 3. Two quiz questions in the GG

The paper is laid out as follows. In Section 2, we present summaries of related work with regards to gender digital divide, gamification and digital game-based learning. In Section 3, we present the procedure followed during the GG and its focus on gender-inclusive language. Section 4 describes the digital elements of the GG; more precisely, pawns, cards, software system, voting system and questionnaires. In Section 5, we briefly describe our exploratory study with some GG questions tested in generative AI (GAI). We conclude the paper and share some future prospects in Section 5.1.

2. Related Work

2.1. Gender Digital Divide

In this section, we refer to some related work on the gender digital divide and the way in which genders are represented in physical and digital games. We begin with research projects at the European level, which address gender imbalances. One of these projects is the GEARING-Roles (Gender Equality Actions in Research Institutions to traNsform Gender ROLES¹⁰), which targets the questioning and transformation of traditional gender roles at the micro, meso and macro levels, with four main objectives: female career progression, education and research, leadership and decision-making, as well as the promotion of gender equality in research organizations. Another project is ATHENA,¹¹ which implements gender equality plans to unlock the research potential of research-performing organizations and research-funding organizations in Europe. Last but not least, the CALIPER¹² project enhances gender balance in STEM fields and promotes a greater engagement of female researchers with research and innovation. This stimulates collaboration between academia, public authorities, professionals and industry players in order to tackle gender inequalities across the research-transfer-to-market chain.

A systematic literature review by Acilar and Sæbø (2023) reveals a crucial aspect of the gender digital divide. They found that gender inequalities do not solely concern who can access ICT. The more significant issue lies in how that access is transformed into valuable, meaningful use. It indicates the continuing challenge of women accessing the digital-enabled society. Here, sociocultural factors play an important role in explaining

¹⁰<https://gearingroles.eu/project>

¹¹www.athenaequality.eu/

¹²<https://caliper-project.eu/>

the gender digital divide. Issues related to the second-level digital divide that comprises both ICT use and digital skills must be addressed. Especially in developing countries, training in ICT and digital media use should be offered to women and girls to learn and improve their digital skills. Moreover, Acilar and Sæbø (2023) suggested the importance of well-formulated implementable policies to bridge the gender digital divide. However, the success of such policies often depends on practical tools that can raise public awareness and foster critical understanding. The digitized GG is designed to be such an ICT-based tool. It illustrates (or explains) the gender digital divide by examining the (under-)representation of females in several domains, particularly in the STEM (Science, Technology, Engineering and Mathematics) fields.

Gamification enhances learning and increases motivation by incorporating game elements like points, badges, levels and challenges, and transforms abstract or complex material into elements that are more accessible and attractive for the user. Gamification is an informal umbrella term for the use of video game elements in non-gaming systems to improve user experience (UX) and user engagement (Deterding, Sicart, Nacke, O'Hara, & Dixon, 2011). Reward and reputation systems with points, badges, levels and leader boards are parts of gamification. In the case of the GG, the gamification elements are reflected in the players rolling a dice and picking a card from a moderator. Moreover, when a question is answered correctly, a small circular light appears on the pawn representing the colour of the domain (work, knowledge, power, society) in question (see Figure 11). Through the GG, we promote gamification as an effective methodology both for raising awareness and educating and disseminating important societal values like gender equality.

Concerning games and fun, Havukainen, Laine, Martikainen, and Sutinen (2020) proposed a co-design process model that considers the inter-generational perspective. In their findings, older adults identified seven game elements as essential to make games fun: appearance and aesthetics, competition, manageability of game-play, social impact, familiarity, unpredictability and inter-generational gameplay. The GG fulfils these elements, since it is indeed an attractive game, promotes competition between players through gaining points and can also be played by several generations. In the following paragraphs, we describe games from a learning perspective, as well as demonstrating how genders are represented in board and video games.

According to Prensky (2001), what attracts and engages children in today's video and computer games is the learning opportunities the games provide, i.e. that the game players i) learn to do things, ii) take in information from many sources and make decisions quickly, and iii) collaborate with others.

Furthermore, the GG is based on student-centred learning (SCL), which shifts the focus from the teacher as the primary source of knowledge to the student as an active participant in their own learning process. Coleman and Money (2020) presented a conceptual framework of student-centred digital game-based learning and used it to classify a range of literature published between 2007 and 2018. They found that "social" elements were integrated to a lesser extent, particularly in areas such as mutual respect, teacher and learner interdependence, and the development of reflexive attitudes to learning and teaching.

As far as the relation of gender with physical or video games is concerned, the video game industry has been considered a field dominated by men, while female players have often been excluded (directly or symbolically) from the gaming community (Dovey & Kennedy, 2006). Moreover, the video game industry is often criticized for certain genders behaving or thinking in specific ways, e.g. men are tough and unemotional, and women are passive, or when active, are often hyper-sexualized and dressed in highly inappropriate clothing. However, there are certain exceptions in which video games challenge traditional gender roles, such as *Omori* (emotionally charged characters), *Monster Prom* (multiple LGBTQ+ characters), and *Resident Evil* (female soldier), to name just a few.

Several years ago, Pobuda (2018) ran a study that analysed the gender and racial representation of game designers and illustrators in the Top 200 board games, as ranked by BoardGameGeek (BGG). This was a table-top gaming community discussion board and online listing of 100,426 table-top games. (Pobuda, 2018) found that 869 white characters were depicted on the cover art compared to 169 representations of non-white or people of colour in the Top 100-ranked games.

2.2. Gender Bias in Generative AI

In the following paragraphs, we provide summaries of related work, first on bias and stereotypes in general, and then on bias in LLMs. After this, we introduce the design and the results of our exploratory study, which entailed posing 35 questions from the GG to ChatGPT 3.5 and 4.0 and evaluating its answers.

Bias is an unjustifiable judgment about a person or situation, whereas stereotype is a widely held, simplified and essentialist belief about a specific group. Stereotypes are often more rigid and resistant to change than biases. According to Beeghly (2015), when we call something a stereotype, we tend to mean it as a criticism, as if calling something a stereotype explains why it is wrong. Beeghly (2015, p. 675) focused his attention on this thought in particular: stereotypes are false or inaccurate, defining this as “the falsity hypothesis”; the falsity hypothesis is widespread and often one of the first reasons people cite to explain why we should not use stereotypical views in cognition, reasoning or speech.

Gender bias in technology originates from gender bias in the data humans provide through text, speech and images. LLMs and Multimodal LLMs are trained using publicly available datasets or via web crawling, which allows for unconstrained data that often includes biases, misinformation, hate speech and stereotypes (Luccioni & Viviano, 2021). In her review of ChatGPT from a social perspective, Gross (2023) mentions that biases live in the training data: whatever the LLM finds, ingests and uses is often already laden with biases and these biases are absorbed back into the model.

Ferrara (2023) categorized bias in LLMs into various types: demographic (e.g. gender, race or age), cultural (e.g. stereotypes), linguistic (e.g. English), temporal (e.g. period applicable to the training data), confirmation (e.g. seeking out information that confirms certain beliefs), or ideological and political (e.g. favouring certain political perspectives or ideologies). Zhou et al. (2024) recently tested three of the most popular text-to-image AI generators (Midjourney, Stable Diffusion, and DALL·E 2) and documented systematic gender and racial biases. They also noted subtle biases in facial expressions and appearances. In their descriptive study, Sun et al. (2024) revealed that

DALL·E 2 systematically underrepresents women in male-dominated occupations and overrepresents them in female-dominated jobs. Furthermore, they found that DALL·E 2 images tend to portray more women than men with smiling faces and faces looking down, particularly in female-dominated occupations, which risks reinforcing traditional gender stereotypes. This is because a downward gaze is often interpreted in visual and psychological studies as a sign of modesty, shyness, passivity, or submission. Their findings emphasize the importance of studying presentational gender biases and the need for them to be continuously monitored and evaluated in generative AI (GAI) technologies. Moreover, Zhou, Abhishek, Derdenger, Kim, and Srinivasan (2024) found that, on the one hand, women express more happiness than men across all three AI generators. Men, on the other hand, typically express more neutral emotion and anger than women across all three AI generators. More recently, Duan, McNeese, and Li (2025) highlighted that AI's use of natural language can evoke social perceptions, including gender attribution, making it susceptible to gender associations. They conducted two online experiments using interactive vignettes, in which they manipulated the gender of the GAI, its gender expertise, and linguistic cues exuded in its responses, and found that while removing gender did reduce gender stereotypes toward GAI, it only did so in terms of the ratings of its gender stereotypical traits.

3. Gender Game

In this section, we describe how the Gender Game is constructed and the procedure that is followed during a GG session. Since it is a large physical board game, the moderators make an appointment in advance with the school or exhibition in order to find a suitable location for the GG. Two moderators are usually onsite to prepare the GG starting play, i.e. constructing the puzzle pieces, projecting the questions, laying the cards, pawns etc., as well as ensuring that the software application is running on the tablet and PC. The GG has been played in its physical version at exhibitions and conferences over 130 times. In this paper, we present the new digital version of the GG, the goal of which is to enhance gameplay not only for players, but also for moderators and researchers. The GG game is a part of a nationally funded project running from 2022 to 2025.

When the players arrive to play the GG (up to 20 players per session), we ask them to take seats and form four mixed-gender teams. Each team is represented by one of the animal pawns (cat, dog, hedgehog, fox). The youngest player starts the game. This player rolls the dice, moves the pawn in one of the four designated areas, and draws a question card from the moderator in relation to this area. Then the player moves close to the pawn with the card and places the card in the pawn's designed pocket. These cards feature radio frequency identification (RFID) and inside the pawns, there is an RFID reader. This displays the question on a projector automatically, so that it can be seen by the whole team. The group must answer the question. However, they may only provide an answer after consultations with other groups and discussions with the moderator. After the discussions, the moderator presents the answer on the next slide.

The moderator stands in the middle of the GG (see Figure 4) to be visible by all players, and to engage in close interaction with the players. If the answer given by the team is correct, the team is awarded a point. The winning team is the first one to obtain a correct answer in all four areas.



Figure 4. A moderator standing in the middle initiating the discussion

3.1. Gender-inclusive language

It is well known that gender inequalities are perpetuated in literature. Striving for gender-inclusive language can have many benefits in linguistics, as well as language technology. Kotek et al. (2021) examined example sentences in three journals *Language*, *Linguistic Inquiry*, and *Natural Language & Linguistic Theory* and found a stark imbalance of male (N = 12,117) to female (N = 5,571) arguments, where male-gendered arguments were more likely to be subjects (as Part of Speech in a sentence), and female-gendered arguments non-subjects. Kotek et al. (2021) showed that female-gendered arguments are more likely to be referred to using a kinship term, to exhibit positive emotions and to be the object of affection, whereas male-gendered arguments are more likely to have jobs, to exhibit negative emotions and to perpetrate violence.

Moreover, Dev et al. (2021) explained the complexity of gender and language referring to gender. They surveyed non-binary persons to understand the harm associated with the treatment of gender as binary in English language technologies. Among three common natural language processing (NLP) tasks (Named Entity Recognition, Coreference Resolution, and Machine Translation), the primary forms of harm they discussed are *misgendering* and *erasure*.

Misgendering is the act of accidentally or intentionally addressing someone (one-self or others) using a gendered term that does not match their gender identity. Erasure is the accidental or intentional invalidation or obscuring of non-binary gender identities. In the GG, all questions and answers are written in gender-inclusive and inclusive language. In addition, the moderator of the GG uses gender-inclusive language during the discussion of the questions that addressed within the GG session.

Some questions refer to gendered nouns, as can be seen in an example question presented in Figure 5. The question is: “What is the name of the person who bakes bread/teaches at a school/makes decisions in a commune?”. In linguistics, *grammatical gender* refers to the practice of grouping nouns into the gendered classes of masculine and feminine. Some languages also use a third class known as the neuter gender and a fourth class known as the common gender, while others have no gender system at all¹³.

¹³<https://www.ebsco.com/research-starters/language-and-linguistics/grammatical-gender>

Although gendered nouns are found in many languages, the system has very limited applications in modern English.

In our specific example in Figure 5, the masculine classes in German are Bäcker, Lehrer and Bürgermeister and in French, *boulangier*, *enseignant* and *bourgmestre*, while the feminine classes are *Bäckerin*, *Lehrerin* and *Bürgermeisterin* in German and *boulangère*, *enseignante* and *bourgmestre* in French. In most cases, players answer only with the male form in German and French (which are grammatically gendered languages), i.e. Bäcker/boulangier, Lehrer/enseignant, Bürgermeister/bourgmestre (this noun has the same form in French for male/female), even if we explicitly ask female participants to answer the question. The correct answer in all four languages is provided in Figure 6.

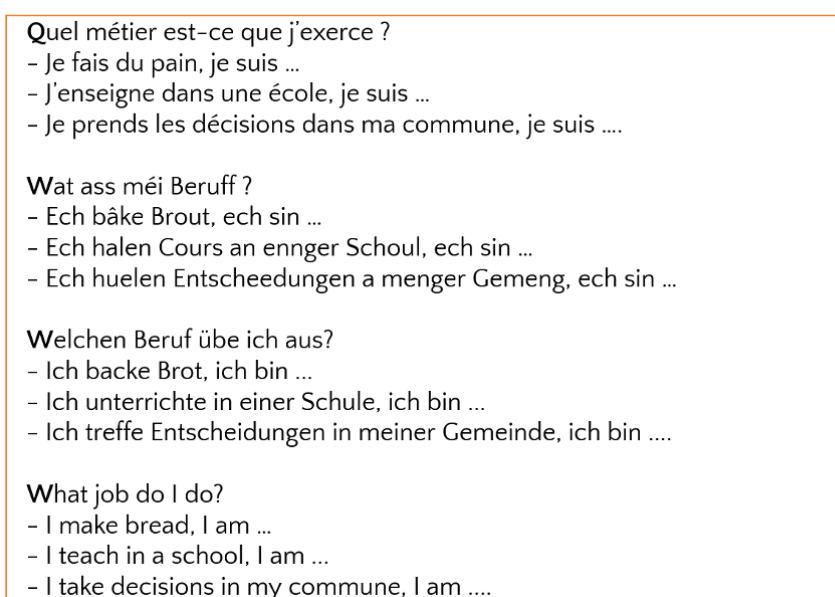


Figure 5. Question to raise awareness about nouns and grammatical gender

When deciding which nouns should be prioritized in communication, the Council of Europe published Guidelines for the use of language as a driver of inclusivity¹⁴. These strategies recommend the following:

- take into account the type of text/oral communication, the context, the audience and the purpose of the communication;
- ensure that the text is readable and the text/oral communication is clear, flowing and concise;
- seek to combine different strategies throughout the text/oral communication.

As far as the gender of nouns and pronouns is concerned, Figure 7 presents the strategies that the European Commission suggests using regarding the language type. The main strategy is neutralization in natural gender languages, such as Danish, English and Swedish, and feminization in grammatical gender language, which is the plurality of European languages. There are also genderless languages, such as Estonian, Finnish and Hungarian and in these cases, usually no strategy is needed.

¹⁴<https://rm.coe.int/guidelines-for-the-use-of-language-as-a-driver-of-inclusivity/1680aec235>



Figure 6. Awareness of masculine and feminine forms of nouns

Natural gender languages DA, EN and SV	Grammatical gender languages BG, CS, DE, EL, ES, FR, GA, HR, IT, LT, LV, MT, NL, PL, PT, RO, SK and SL	Genderless languages ET, FI and HU
<ul style="list-style-type: none"> personal nouns are mostly gender-neutral there are personal pronouns specific to each gender 	<ul style="list-style-type: none"> every noun has a grammatical gender the gender of personal pronouns usually matches the reference noun 	<ul style="list-style-type: none"> there is no grammatical gender and no pronominal gender
Main strategy: NEUTRALISATION	Main strategy: FEMINISATION	Usually no strategy needed

Figure 7. Gender and types of languages

4. Digitization of the Gender Game

The GG tackles the societal challenge of gender equality from a research perspective. At the same time, it provides a constructive and efficient learning environment with respect to a fair educational assessment for diverse learners (children, youth and adults), and in particular with respect to the goal of ensuring more equal opportunities in education.

Enhancing the physical GG with digital elements aims to both improve the gaming experience of the players while they learn, as well as making the organization and assessment of the learning impact easier for the researchers. Thus, the goal of deconstructing stereotypes is reached more efficiently. We consider the advancements with digital elements for three stakeholders of the GG (see Figure 8):

- **Players:** the GG participants (children, youths and adults) who answer the questions posed by the moderator by discussing them with other players and the moderator;
- **Moderators:** the people who stand in the middle of the circle and moderate the GG;
- **Researchers:** the designers of the game (including psychologists, engineers and computational linguists), who collect data to assess its impact on the deconstruction of gender stereotypes.

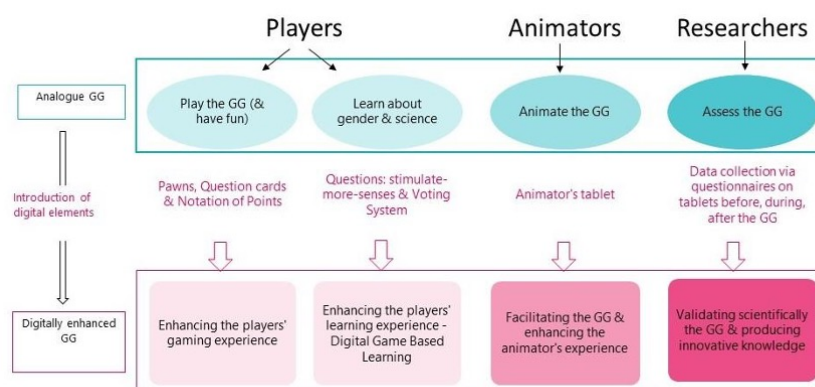


Figure 8. Benefits of digitally enhanced GG

As can be seen in Figure 8, we designed and developed several digital elements in order to transition from an analogue GG to a digitally enhanced version. The pawns were integrated with RFID scanners to be able to automatically project a question by placing a card on it. More questions were added including multimodal (audio-visual) data, thus stimulating more senses. In the digitized version of the GG, all players receive a tablet during the session. This enables another kind of question-answer type, that is, an anonymous voting system. For example, one form of voting question is featured as: Getting a higher education degree is more important for a girl than for a boy! (see picture on the right in Figure 2). The results of the voting system can be shown on the projector in real-time, therefore reinforcing the competition and making the gameplay even more attractive.

The moderator is now provided with a tablet featuring a dedicated GG application (see picture on the right in Figure 11). Within the application, there is a dashboard, where they can check the status of the teams (which pawn is playing next, how many answers each group answered correctly, etc). Last but not least, the evaluation of the GG has become more efficient, since the questionnaires are now in digital form and several research components can be validated through the answers (knowledge acquisition from GG, usage of gender-inclusive language, UX, etc.).

4.1. Design workshop

In this section, we discuss the design process behind the digitization of GG, which first started with a multidisciplinary design workshop. In May 2022, a research team of 10 persons with different backgrounds and competences (sociologists, computer scientists, psychologists) brainstormed how to digitize the five elements involved in the GG: pawns, cards, software system, voting system and questionnaires. Two moderators animated this design workshop, proposing the problematization of the gaming elements and suggesting diverging and converging steps.

A Design Thinking (DT) approach was used to develop the advanced Gender Game. The methodology is solution-based, an approach usually considered in engineering (Plattner, Meinel, & Weinberg, 2009). To create the new blended digital and physical version of the game, we followed five nonlinear steps of DT (*Empathize, Define, Ideate, Prototype and Test*).

The first stage of the process (*Empathize*), is to gain an empathic understanding of both the limitations of the current GG, and the potential of a new version of the GG. It involves finding out more about the area of concern through observing, engaging and empathizing with the players and moderators, via playing and training sessions, thanks to the assessment framework. The *Define* stage analyses observations from stage 1 and synthesizes them in UX maps to define the needs and requirements for developing the game. The design team started by mapping the current gaming experience of the players and of the animator in the form of a journey representing the different phases of a game and for each phase: actions, motivation, feelings and "pains". Iterating from field to theory, we also deconstructed the game into gaming mechanisms, as also considering how they contribute to the gaming experience, and learning.

The subsequent *Ideate* stage allows for 'thinking outside the box' to identify future developments. The *Prototype* step aims to identify the best answers to fulfil the needs and requirements of phases 1 and 2. As a result, we were able to select which gaming mechanisms should be digitized for the players: dice rolling, selecting and displaying questions, answering the questions and displaying the answers, and finally, scoring. During the *Ideation* phase, when the dice was rolled the following ideas arose: suppressing the dice totally; making it fully digital; lighting the dice of a team to indicate when it is their turn; having a camera on the ceiling to read the number on the dice and light up the corresponding square on the board; adding a "joker" side of the dice for the team to choose any category they wanted. When converging on the gaming element entitled "question selection and display", we opted to keep the physical cards and have an RFID tag on each question card. In this way, the players can place the card on the reader and display it directly on the screen (instead of handing it to the moderator who read it out loud, as in the previous version).

The *Test* is the final stage of the five stage-model, which must be considered as an iterative process (playing the GG). The results generated during the testing phase are often used to redefine one or more problems. They are also used to better understand the users, conditions of use, thinking processes, behaviours and feelings, and to empathize. These aspects are supported by scientific assessment and validation. The design team prototyped different options of the selected ideas (e.g. a unique RFID reader versus an RFID reader on each team pawn) and tested them during gaming sessions. They collected game play observations of players during playing the game, submitting a Player Inventory Experience questionnaire¹⁵ and a semi-structured interview to collect more qualitative feedback. In the physical GG game, pawns were made out of machine-cut hard foam with printed stickers. There are 4 pawns with 4 animal prints: cat, dog, hedgehog and fox (see Figure 9).

In the digitized version, the pawn outlines were initially used to make a 3D model from hard foam with cut-outs for hiding the electronics and a RFID reader in the pawn (see Figure 11). A printed circuit board was fitted in the base of the pawns, hosting the microcontroller, battery and charging circuit. The RFID reader and addressable RGB-LED for feedback and scoring were positioned in the chest region behind the breast pocket of the pawn on the other side. In the latest iteration of the pawn's design, we unified the body for 3D printables and all the electronics are integrated into the pawns'

¹⁵<https://playerexperienceinventory.org/>



Figure 9. The four GG game pawns

body. The unified body models made it possible to produce the individual heads as a "puzzle" piece with a press fit for assembly. The electronics are now positioned behind a lid in the breast area for maintenance reasons. The base can be interchanged with an AA battery case or power banks to offer different power options (see Figure 10). This makes the digitized version of the GG easily transportable, e.g. for conferences abroad.

4.2. Pawns

In the physical version of the GG, the player shows the card to a second moderator, who then selects the question number and projects it on a screen. In the new version, cards are tagged with RFID stickers. The players, after drawing a card from the deck of cards, moves the card close to the built-in RFID reader (designed as a pocket area). The unique ID of the card is used to select the question category, and a question from this category is then randomly generated. This makes it possible to use one card-set to serve all age groups, as well as avoiding maintenance of 1-to-1 relations between card IDs and questions.



Figure 10. Hard-foam and 3D printed version of the pawn (left & centre), showing the different electronics iterations (right)

4.3. Cards

In the physical version of the GG, the player shows the card to a second moderator, who then selects the question number and projects it on a screen. After drawing a card from the deck, the player moves the card close to the built-in RFID reader (designed as a pocket area of the pawns). The unique ID of the card is used to select the question category, and a question from this category is then randomly generated. This makes it

possible to use one card-set to serve all age groups, as well as avoiding maintenance of one-to-one relations between card IDs and questions.

4.4. Software System

The software system for the digitized version makes use of several open-source software components to speed up development. It is a client server design that uses Node-Red¹⁶ as the backend for the business logic and to serve HTML pages for both the moderator and the audience. Game management/setup and information about the game state are visible on a tablet that the moderator holds in their hand during the GG. Figure 11 (right) presents a general view of the whole app. Some of the most important features are shown, such as which team's turn it is. It also shows the question and a preview of the answer, and each team's points, as well as the battery and connection state of the digitized pawns. A second web view displays the questions, answers and information on turns to the players. Node-Red is also used to control the content provided to the players' tablets before, during and after the GG session.

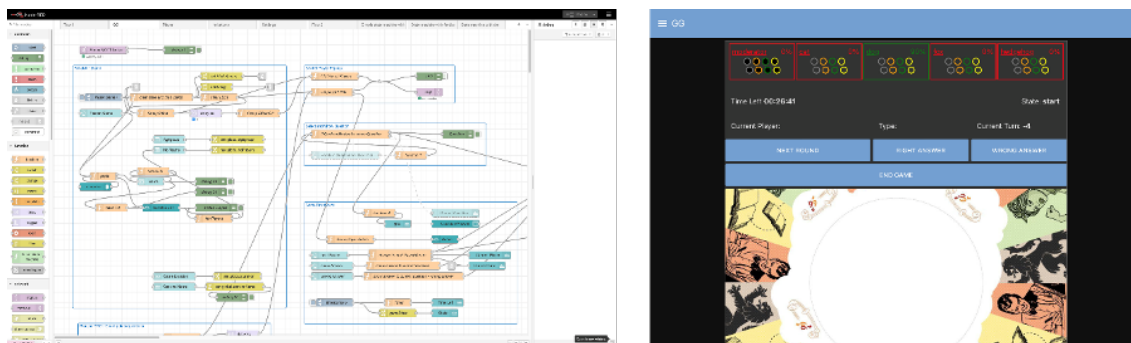


Figure 11. Node-Red and GG application

The digitized pawns use wireless (WiFi) enabled microcontrollers running the open-source firmware Tasmota¹⁷, which is software that can be configured for building Internet of Things (IoT) devices. Tasmota uses Message Queueing Telemetry Transport (MQTT) as a lightweight protocol to establish bi-directional communication with the server component via an MQTT broker that is also served by Node-Red.

4.5. Voting system

In the physical GG, the players usually answer closed questions, and these answers are then discussed with the moderator. A new gamification element added to the digitized GG is the voting system. The process is as follows: if a player picks a card with the voting system, all participants answer a questionnaire on the tablets provided, and they vote during the game. Anonymized results of the voting will be presented during the game, so that participants can discuss the results with the moderators. This includes discussing the effects of playing the GG and how this may or may not have changed their perception of gender roles and stereotypes. An example of a question with the voting system can be found on the picture on the right in Figure 2: "Getting a higher education degree is more important for a girl than for a boy!"

¹⁶<https://nodered.org/>

¹⁷<https://tasmota.github.io/docs/>

4.6. Questionnaires

In order to measure the immediate impact of the game, we follow a “before-and-after-design” methodology using questionnaires provided to players before and after the game. In the digitized version of GG, the questionnaires are available on the tablets.

We also measure the long-term impact of the game by sending participants the same post-activity questionnaire three weeks later. To remember to do this, participants scan a personalized QR code with their personal mobile devices after they finish the session. A calendar entry will be added to remind them to access and fill in the questionnaire three weeks later. The description of the questions as well as the results of the questionnaires are not the focus of this article and will be published at a later time.

Besides the “before-and-after-design” questionnaire measuring players’ knowledge on gender equality, we also provided a UX questionnaire after playing the GG. We took the mini Player Experience Inventory (PXI) questionnaire as the basis and adapted it slightly for the GG, as shown in the table below. The PXI questionnaire draws on the Means-End theory (Gutman, 1982) and the related ‘Mechanics - Dynamics - Aesthetics’ (MDA) framework (Hunicke, LeBlanc, Martikainen, & Sutine, 2004). Our questionnaire consists of seven closed questions with Likert scale response options (from 1 low to 5 high) and three open-end questions to collect qualitative data.

1	Playing the game was meaningful to me.
2	I felt I was good at playing this game.
3	I liked the look and feel of the game.
4	The game was not too easy and not too hard to play.
5	It was easy to know how to perform actions in the game.
6	The goals of the game were clear to me.
7	I had a good time playing this game.
8	What did you like most?
9	What did you like least?
10	What would you change?

Table 1. Mini PXI questionnaire adapted for the GG

The GG was presented at Researcher Days, a large-scale fair, on 28-30 November 2024; there we collected 105 questionnaire responses. Here, we present the descriptive statistics of the Likert-scale responses (see Figure 12).

In Figure 12, we see that all questions were answered with a score above four in questions 1-7. The lowest score (4,033) was given to "I felt I was good at playing this game" and "The game was not too easy and not too hard to play". The highest score was given in the question "The goals of the game were clear to me".

As far as the answers to questions 8-10 are concerned, we can observe that certain digitization benefits were very positively accepted by the players. As for what they liked most, participants provided the following answers: cards, figures, scanning cards and pawns. Many participants also stated that they liked both the discussions and the moderators, with some of the following answers given: “the women explaining the answers (ID36, ID37)”, “exchanging together information (ID80)”, explanations (ID42, ID78, ID79, ID86)”. This result is a clear indication that the physical human-human

interaction is of critical importance for the GG. Last but not least, the following answers indicate a positive overall look and feel of the game: “the figures are animals” (ID64, ID94, ID104), “decoration” (ID25, ID31).

In answer to the question “What did you like least?”, the main theme in the answers was the “waiting time” (ID13, ID14, ID33), and that the pawns did not “recognize the cards” (ID19, ID23, ID43, ID61, ID74). There was no mention of the fact that once a card has been already scanned, it cannot be scanned again; therefore, a careful organization of cards by the moderator is required.

In terms of question 10 “What would you change?”, participants wanted the game duration to be shorter, but also stated “making questions easier” (ID77, ID80, ID83, ID85, ID104)”, since some questions were difficult for them to answer.

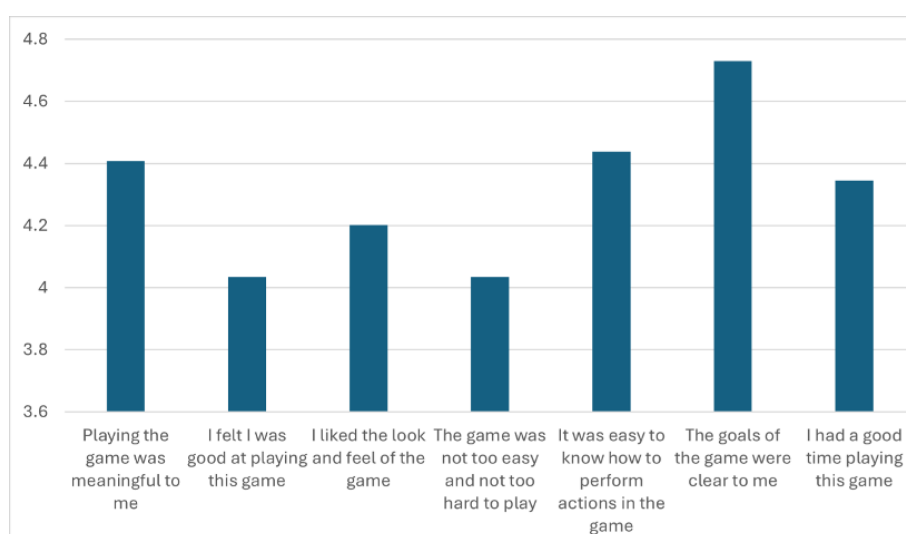


Figure 12. Evaluation of the GG questionnaires (1 strongly disagree - 5 strongly agree)

5. GAI in the Gender Game

We observe that GAI (both text and vision-based) is biased because the models are trained using publicly available datasets or via web crawling. This allows for unconstrained data that often includes biases, misinformation and hate speech. Although solving gender bias is a topic that is being researched by NLP and AI, it is still a difficult challenge to tackle. In the first instance, we wanted to test how GAI addresses some questions that could provoke a biased answer. Thus, we decided to test Open AI’s ChatGPT 3.5 for a sample of 35 questions from the GG; the criterion for selecting these 35 questions specifically was for a gender stereotype to be implicit in the statement or question. This is because explicit statements in the questions would prompt the LLMs to answer in an unbiased way.

The objective of this study was to test whether the answer would include bias, and if yes, of what kind. We found that for the GG questions in quiz form in particular, ChatGPT mainly exhibited ageism and gender bias, as can be shown in the two tables below, respectively.

In this example, we did not reveal whether the mother was 32 and the father 34 or the other way around. The question was intended to raise exactly this point: which

GG question	Sam was born in 2016. At that time, his parents were 32 and 34 years old. In your opinion, how old was his dad?
ChatGPT answer	“If Sam was born in 2016, and his parents are 32 and 34 years old at the time of his birth, we can calculate the father’s age by subtracting 32 (the mother’s age) from 34 (the father’s age). This would mean the father is 2 years older than the mother. Therefore, the father is 34 years old at the time of Sam’s birth in 2016.”

gender is considered the younger and which the older parent? The ChatGPT answer takes for granted that the older parent is male, which is a clear demonstration of ageism.

GG question	Dr Brandt lives in Hamburg and has a brother in Berlin, Prof. Brandt. But Prof. Brandt does not have a brother in Hamburg. How is this possible?
ChatGPT 3.5 answer	“This riddle can be resolved by considering the possibility that Dr. Brandt and Prof. Brandt are siblings, but Dr. Brandt lives in Hamburg, while Prof. Brandt resides in Berlin. In this scenario: <ul style="list-style-type: none"> • Dr. Brandt lives in Hamburg. • Prof. Brandt, who is Dr. Brandt’s brother, resides in Berlin. • Prof. Brandt, who lives in Berlin, does not have a brother living in Hamburg (implying that Dr. Brandt does not have a brother residing in Hamburg, as he is the one living there). So, the explanation lies in the fact that Dr. Brandt and Prof. Brandt are siblings, with Dr. Brandt residing in Hamburg and Prof. Brandt living in Berlin.”

In this example, the answer is that Dr Brandt is a female. This was not revealed by ChatGPT 3.5, which simply described and paraphrased the quiz question.

However, we also tested the two examples above with the newer version of ChatGPT (ChatGPT 4.0) and the answers are depicted below:

ChatGPT 4.0 answer	“The dad is 34 years old because the question states that Sam’s parents are 32 and 34 years old. Since one parent is 32 and the other is 34, the dad is most likely the 34-year-old parent.”
--------------------	--

The answer given by ChatGPT 4.0 is less verbose; it also paraphrases the question to a limited extent. Most importantly though, it adds “most likely” to the answer. However, it would have been appropriate and fair if reasons had been provided as to why it is most likely for the father to be the older parent. As for the second example, the following answer was given:

ChatGPT 4.0 answer	<p>“This is a classic riddle! The trick lies in the way the relationships are described. Dr. Brandt, who lives in Hamburg, is the brother of Prof. Brandt, who lives in Berlin. However, Prof. Brandt does not have a brother—because Dr. Brandt is actually his sister! So, Dr. Brandt is a woman, which is why Prof. Brandt has no brother in Hamburg.”</p>
--------------------	---

Through this correct answer, we show the evolution and advancement of LLMs with regards to gender stereotypes.

Last but not least, we provide below another example:

GG question	Who won the last football World Cup?
ChatGPT 4.0 answer	<p>“Argentina secured their third FIFA World Cup title in 2022, defeating France in a dramatic final that ended 3-3 after extra time and was decided by a 4-2 penalty shootout. Lionel Messi led Argentina with two goals, while France’s Kylian Mbappé scored a hat-trick, becoming the first player to do so in a World Cup final since 1966. This victory marked Argentina’s first World Cup win since 1986 and the first time a non-European team claimed the title since 2002. The 2022 final is widely regarded as one of the greatest football matches in history.”</p>

The answer provided by ChatGPT 4.0 is partially correct. Indeed, Argentina won the last men’s World Cup, but ChatGPT 4.0 completely ignored the FIFA women’s World Cup, which took place in 2023 and thus more recently than the men’s Cup, and is therefore the "last" World Cup. Here, we see that ChatGPT 4.0 prioritizes the men’s Cup over the women’s World Cup. When playing the GG, the moderators describe the significance of biases we have to the players and how these biases finally enter LLMs and the GAI cycle as training data.

Within the GG project, we made video recordings of 10 sessions of the GG and analysed the speech of players, particularly the bias in their speech. Comparing the bias expressed by children versus that of ChatGPT, we documented that bias by children is often unjustifiable or based on their own experience. This is understandable, since they do not yet have the cognitive skills to analyse all factors that can lead to a certain situation. On the other hand, the bias by LLMs (ChatGPT in particular) is more implicit and the answers often unclear. Mitigating gender bias means bringing together experts from various disciplines, including computer science, social sciences, humanities and ethics, in order to foster a comprehensive understanding of the potential biases associated with AI applications.

The GG is a highly interactive game, which tackles the challenge of gender equality by breaking down gender stereotypes through discussions between moderators and players. It is aimed at a diverse audience, including children, youth and adults. The research institutions LIST and LISER have carefully curated 144 questions and answers customized for the age ranges and domains in which most gender stereotypes exist, based on the European Index on Gender Equality: work, knowledge, power and society. Using a gamified approach, which was scientifically validated after a Design Thinking workshop, four mechanisms were digitized: pawns, card scanning, voting system and questionnaires, with these digital elements advancing the original physical game. The

audience now experiences a very interactive session with a high learning impact in parallel.

The stakeholders of the GG, the players, moderators and researchers, all benefit equally from the digital advancement of the GG. The players' gaming experience is particularly enhanced through digital game-based learning. The moderator uses a single application on the tablet, where both questions and answers, as well as the order of the groups and the winning points of each group, are displayed. This provides the moderator with an overall view of the current status of the game. The assessment of the GG is also digitized through questionnaires, which are more efficient and cost-effective in terms of data collection. Based on 105 responses collected, the average rating of all questions exceeded 4 (5 is the highest score), with the highest score being given to the question "The goals of the game were clear to me".

5.1. Limitations and future prospects

The digitized version of the GG is an ICT-based tool which reduces the digital gender divide and enhances learning through gamification and digital game-based learning. Through the usage of gender-inclusive language both in the Q&As and in discussions with the moderator, the GG additionally fosters linguistic justice, as it educates players to avoid gendered language in their everyday verbal and written communication, and thus, it contributes to reducing the amount of bias entering the generative AI. As well as the GG, we described a short study that was undertaken with ChatGPT 3.5 and 4.0 to test the AI bias in order to make participants aware of the gender bias that exists in AI. The limitations of this study relate to the fact that we tested only a small excerpt of the GG questions (35 out of 144). Another limitation is that we did not compare different LLMs, but tested only ChatGPT. However, we tested both versions to observe the evolution of the tool.

In the future, we plan to test all GG questions in GAI systematically and compare at least the GAI answers from two LLMs. Concerning the future prospects of the GG, we plan to include a humanoid social QT robot¹⁸ in the game as an assistant moderator. This robot can stand next to the moderator, welcome the players and briefly explain the rules of the GG. Additionally, it can be used as a joker or expert teacher, providing an answer to a question, if a group requires it.

Acknowledgements

The project is funded by the Luxembourg National Research Fund and was awarded an "Outstanding Award for Promoting Science to the Public".

References

- Acilar, A., & Sæbø, (2023). Towards understanding the gender digital divide: A systematic literature review. *Global Knowledge, Memory and Communication*, 72(3), 233–249. doi: 10.1108/GKMC-09-2021-0147
- Beeghly, E. (2015). What is a Stereotype? What is Stereotyping? *Hypatia*, 30(4), 675–691. doi: 10.1111/hypa.12170

¹⁸<https://luxai.com/>

- Coleman, T. E., & Money, A. G. (2020). Student-centred digital game-based learning: A conceptual framework and survey of the state of the art. *Higher Education*, 79(3), 415–457. doi: 10.1007/s10734-019-00417-0
- Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011). Gamification using game-design elements in non-gaming contexts. In *CHI'11 extended abstracts on human factors in computing systems* (pp. 2425–2428).
- Dovey, J., & Kennedy, H. W. (2006). *Game cultures: Computer games as new media: computer games as new media*. UK: McGraw-Hill Education.
- Duan, W., McNeese, N., & Li, L. (2025). Gender stereotypes toward non-gendered generative AI: The role of gendered expertise and gendered linguistic cues. *Proceedings of the ACM on Human-Computer Interaction*, 9(1), 1–35.
- Gross, N. (2023). What ChatGPT tells us about gender: A cautionary tale about performativity and gender biases in AI. *Social Sciences*, 12(8), 435. doi: 10.3390/socsci12080435
- Gutman, J. (1982). A means-end chain model based on consumer categorization processes. *Journal of marketing*, 46(2), 60–72.
- Havukainen, M., Laine, T. H., Martikainen, T., & Sutinen, E. (2020). A case study on co-designing digital games with older adults and children: Game elements, assets, and challenges. *The Computer Games Journal*, 9(2), 163–188. doi: 10.1007/s40869-020-00100-w
- Hunicke, R., LeBlanc, M., Martikainen, T., & Sutine, E. (2004). MDA: A formal approach to game design and game research. In *Proceedings of the AAAI Workshop on Challenges in Game AI* (Vol. 4, p. 1722).
- Luccioni, A. S., & Viviano, J. D. (2021). *What's in the box? A preliminary analysis of undesirable content in the common crawl corpus*. arXiv. doi: 10.48550/arXiv.2105.02732
- Plattner, H., Meinel, C., & Weinberg, U. (2009). *Design thinking*. Landsberg am Lech: Mi-Fachverlag.
- Pobuda, T. (2018). Inclusivity or invisibility?: Gender and racial representation in top-rated BoardGameGeek Games. *Analog Game Studies*.
- Zhou, M., Abhishek, V., Derdenger, T., Kim, J., & Srinivasan, K. (2024). *Bias in generative AI*. arXiv. doi: 10.48550/arXiv.2403.02726

Professor: Who holds the copyright for AI-assisted and AI-generated contents?

Katerina Zdravkova ¹

¹Ss. Cyril and Methodius University in Skopje, Faculty of Computer Science and Engineering, katerina.zdravkova@finki.ukim.mk

Abstract

Generative artificial intelligence (AI) has had a significant impact on intellectual property protection, mainly because model training is usually based on copyrighted content. This problem will deepen because, on the one hand, corporations with large training resources begin to sell content to AI companies, but also because authors whose copyrighted works are being used without permission react against this illegal theft. Parallel to this dilemma, a bigger one opens up: who is the copyright holder when AI is an assistant in the creation of a work, or when AI has competently created a new work? The discussion of these issues is illustrated with the help of nine case studies, which are elaborated in more detail. Based on this, it becomes evident that works created with the support of AI can be considered the copyrighted works of the person whose creativity is aided by AI. In contrast, fully autonomous works, created without any AI assistance, cannot be granted intellectual property protection.

Keywords: *Authorship, copyright, academic dishonesty, scientific misconduct, lawsuits, infringement, artificial intelligence.*

Resumo

A inteligência artificial generativa (IA) teve um impacto significativo na proteção da propriedade intelectual, principalmente porque o treino dos modelos utiliza geralmente conteúdo protegido por direitos de autor. Esse problema agravar-se-á porque, por um lado, as empresas com grandes recursos para o treino destes modelos começam a vender conteúdo para empresas de IA; por outro lado, os autores, cujas obras protegidas por direitos de autor são utilizadas sem permissão, reagem contra esse roubo ilegal. Paralelamente a este dilema, surge um outro ainda maior: quem é o detentor dos direitos de autor quando a IA é um assistente na criação de uma obra ou quando a IA criou competentemente uma nova obra?

A discussão destas questões é ilustrada recorrendo a nove estudos de caso, elaborados com mais detalhes. Com base nisto, torna-se evidente que as obras criadas com o apoio da IA podem ser consideradas obras protegidas por direitos de autor da pessoa cuja criatividade é auxiliada pela IA. Em contrapartida, obras totalmente autónomas, criadas sem qualquer recurso da IA, não podem receber proteção de propriedade intelectual.

Palavras-chave: *Autoria, direitos de autor, desonestidade académica, má conduta científica, processos judiciais, violação, inteligência artificial.*

1. Introduction

One of the key reasons for human progress is the inherent desire of individuals to become better, faster and more efficient – in essence, superior to others. Many tirelessly strive to achieve these goals, investing enormous efforts in their aspirations. Unfortunately, some people take shortcuts and engage in dishonest practices. Cheating takes many forms: athletes use doping (Nicholls et al., 2020), companies inflate revenues or conceal debts (Schilit & Perler, 2010), counterfeiters forge works of art (Spink, Moyer, Park, & Heinonen, 2013), traders engage in insider trading (Woody, 2020), and politicians manipulate votes (Hernández-Huerta, 2020). Similarly, some students resort to academic dishonesty, such as plagiarism, cheating on exams, or falsifying assignments (Zdravkova, 2023), while researchers may engage in misconduct, including data fabrication, falsification, or unethical authorship practices (Gopalakrishna et al., 2022). Because of this dishonesty, it becomes unclear as to who truly won the race, what the company is actually worth, whether the artwork in the gallery is authentic, and what the true value of the stock is. Counterfeiting, piracy, academic dishonesty and scientific misconduct predominantly affect intellectual property, particularly the copyright of the work.

When an author or a group of authors create an original literary or artistic work, according to the Berne Convention, which was adopted in 1886, the authorship is automatically granted, without the need for registration or other formalities (WIPO, 1979). However, the authorship does not mean that the authors hold the copyright (Jaszi, 2017). Copyright gives legal rights regarding the control of use and distribution of the work. For literary works (poems, novels, research papers, computer codes etc.), copyright is usually transferred by the author to a publisher or a software company who then has the legal rights to control how the work is distributed, reproduced and used (Varian, 2005). In most cases, unless explicitly transferred, the authors of artistic works (paintings, sculptures, engravings, photographs, etc.) retain copyright over their artwork even after selling it to a gallery or an individual buyer (Pila, 2010). In music and film, original creators like songwriters and directors may retain some rights, but they often transfer significant rights to record labels or production companies (Hesmondhalgh, 2018).

The copyright is not eternal. According to the World Intellectual Property Organization (WIPO), the United Nations agency responsible for protecting and promoting the rights of innovators and creators worldwide, it lasts for at least 50 years after the creator's death (WIPO, 1996). In the U.S., the protection lasts 70 years after the creator's death, and for some works, such as the anonymous ones, even '120 years from its

creation' (U.S. Copyright Office, 1976, 2024). After this period, the works are no longer protected by intellectual property laws and become part of the so-called public domain, meaning that everyone can own them (Ochoa, 2002). For the works that belong to the public domain, no one can obtain a copyright or use them in other works that are protected by intellectual property rights (Ochoa, 2002). These works become a public good and belong to all of humanity.

In the past, there have been many cases of recorded copyright infringements that were afterwards revoked, affecting the reputation and standing of those who claimed that these were their original works (Temple, 2003). In addition to this, copyright infringement may have severe legal consequences. For example, the Department of Justice foresees a financial penalty of a quarter of a million U.S. dollars and a maximum sentence of 10 years in prison if the defendant has already been convicted of the same crime (U.S. Department of Justice, 2022).

With the rise of digital technologies, the possibility of copyright infringement has significantly increased (Lunney Jr, 2001). This was primarily made possible by the public display of vast amounts of copyrighted works on the web that are freely shared and pirated in abundance. AI-based technologies have significantly transformed the way literary and artistic works are created, directly impacting areas protected by copyright (Yusuf, Pervin, & Román-González, 2024). These tools enable individuals to create their own literary and artistic contents of satisfactory, or sometimes even high quality (Bagenal, 2024).

The capabilities of AI-generated systems have reached such a high level that it is almost impossible to prove that a literary or artistic work was not created by humans (Frank et al., 2024). Some of these works seem completely original. Despite recent advances, generative AI cannot be recognized as a copyright owner because both U.S. and EU legislation explicitly reserve copyright protection for human creators: U.S. law requires a 'human being' (Library of Congress, 2023), while EU law refers to a 'natural person' (EU, 2019).

Whether a human-AI-assisted work is eligible for copyright protection depends on several factors, including the way AI was used, whether the human author disclosed AI assistance, and whether the training data used comply with copyright laws. These factors are discussed in detail in the extended version of the paper.

2. Copyright and the nine case studies

The following section introduces the nine case studies that were selected to cover a broad spectrum of applications where AI plays a role, ranging from academic assignments to creative fields like art, music, and literature.

Each case study explores a distinct aspect of AI's involvement: for instance, AI-assisted student assignments highlight the educational dimension, while AI-generated art and music focus on the creative process. The inclusion of both poetry and novels acknowledges the different forms of literary expression that AI can impact, with poetry often involving more condensed, emotional content, and novels typically requiring more complex narrative structure. This diversity in case studies ensures that a wide range of AI's capabilities and challenges are fully covered.

2.1. Copyright of student assignments

For most educational institutions, the student holds copyright of their work, automatically, without any need to register the work to obtain a copyright (Crews, 2020). Certainly, in order for students to obtain this copyright, their assignments, whether they are text, image, video or code, should be an original contribution. Unfortunately, students often plagiarize, i.e., steal other people's authorship and try to obtain copyright for the stolen content.

According to Laura Young (2024), academic programming plagiarism is on the rise. The impression of the professors at my faculty is that after the appearance of the LLMs there has been an increase in copyright infringement.

Based on my long-time experience as a computer ethics teacher, student cheating has significantly changed after December 2022 (Zdravkova, 2023). Before the appearance of ChatGPT, cheaters either literally copied already published texts or engaged in cross-lingual plagiarism. Regarding verbatim copies in student assignments, which refer to the direct reproduction of text or content without alteration, there are two possible scenarios related to copyright:

- If the stolen content was protected by copyright, then the act of verbatim copying is a theft and is subject to sanctions.
- If the stolen content was copyleft, meaning it is licensed in a way that allows others to freely use, modify, and distribute it under the same conditions (Free Software Foundation, 2022), then their derivative work also becomes copyleft, and therefore they lose their copyright.

Cross-lingual plagiarism, which means content automatically translated with machine translation services (MTS) into another language (usually from English), is also connected with two scenarios:

- Although translation is a derivative work, the copyright still belongs to the original author.
- Translating a work using MTS does not transfer the copyright to that service.

As of December 2022, students began to make heavy use of ChatGPT and other large language models (LLMs). Discovering text generated by LLMs is a very complex challenge. Plagiarism checkers, such as: iThenticate (www.ithenticate.com/) and GPTZero (gptzero.me/) claim to be able to verify whether a document was written by artificial intelligence. However, to avoid plagiarism, students resort to using paraphrasing tools such as QuillBot (quillbot.com/), Wordtune (www.wordtune.com/), Paraphrase Tool (paraphrasingtool.ai/), Rephrase (www.rephrase.info/), and Grammarly (www.grammarly.com), both premium and free. Although such plagiarism cannot be formally proven, as an experienced teacher with extensive experience, which includes teaching the computer ethics course since 2003, I have developed a keen ability to recognize when the writing style deviates from what is typical for computer science students (Zdravkova & Ilijoski, 2025). I usually do not announce the grade to suspected students and invite them to discuss their assignment. If they answer my questions competently, which happens extremely rarely, then they receive a grade based on the quality of the assignment. It is interesting that most of the suspected students do not try to dispute the suspicion at all.

Students are typically granted authorship for their assignments by default, regardless of whether generative AI was used in the creation process. If the AI helped create an assignment in accordance with the university or course rules, then they deserved that authorship. But, if AI wrote the whole assignment based on their prompts, then they will again get the authorship because AI cannot become a copyright holder. In the second case, that authorship is a crime.

Namely, AI-generated media are trained with various data, including copyrighted ones, as evidenced by the New York Times journalist's lawsuit against OpenAI, which OpenAI refutes (OpenAI, 2024). This is probably the most well-known of the large number of lawsuits that appear ever more often and refer to the greatest generative-AI software companies in the world (Panettieri, 2024).

The symbiosis of academic dishonesty and copyright infringement by generative AI opens up space for mass theft of intellectual property. If this issue is not addressed, it may lead to significant and far-reaching consequences.

2.2. AI-assisted and AI-generated art

Vincent van Gogh (1853 - 1890) is my favourite painter. He died more than 100 years ago, so all his paintings are public domain. This means that everyone can replicate his paintings and even sell them legally. Nicolae, the owner of Niko Prints (<https://nickoprints.com/>) creates digital art and sells it. He digitally combined two famous van Gogh's paintings: *The starry night* and *Wheat field with cypresses*. The name of the digital picture is expectedly, *Wheat Field with Cypresses under a Starry Night*.



Figure 1. Two van Gogh's paintings and their digital combination

Nicolae obviously knows that he cannot protect his works based on works from the public domain by copyright. What is even more important is that Nicolae does not appropriate van Gogh's works and gives credit to the original works. Therefore, his commercial activity is not a copyright violation. His picture is copyleft just like the original van Gogh's works, and thanks to that I was able to copy them into this article. However, in contrast to his approach, many AI companies involved in generating images and videos do not demonstrate the same level of consideration. A group of young artists from several countries launched the so-called class-action lawsuit against Stability AI, DeviantArt, Midjourney, and Runway AI for using their works without consent, credit, or compensation (Justia, 2023). Further information about the movement and the affected artists can be found on the website (<https://imagegeneratorlitigation.com/>).

Generative AI companies that do not use public domain works or have purchased copyright for their training data are seriously infringing copyright (Chu, Song, & Yang, 2024). Moreover, anyone who chooses to use their software to create their own work for commercial purposes will face copyright infringement sanctions if the AI-generated

works are found to be similar to some original copyrighted art. This will open a complex and contentious issue, leading to an influx of lawsuits regarding authorship theft.

2.3. AI-generated music

In July 2024, primary school participants at the European Junior Olympiad in Informatics (<https://ejoi2024.gov.md/>), held in Chisinau, Moldova generated their own anthem (EJOI MK, 2024). Bisera, Vladimir, Marko and Nikola defined the theme of the anthem, and music creator Suno AI (<https://suno.com/>) created and sang it. When I inquired from an experienced colleague about who holds the copyright of the anthem, the response was immediate: 'Children.' This position is entirely accurate, as children, being human beings, fulfil the main prerequisite for obtaining a copyright. Nevertheless, a lawsuit was filed against Suno alleging widespread infringement of copyrighted sound recordings because Suno was trained on copyrighted music (Resnikoff, 2024).

Training the system using copyrighted works is a flagrant example of intellectual property theft. The children's anthem, which by the way, won the desired awards, is their authored work, especially the text that they more or less wrote themselves. They are the copyright holders of that anthem. It is interesting that the melody does not resemble any existing song, so the anthem is not only pleasant to listen to, but also original. However, AI-based systems that compose and create text based on well-known songs violate the copyright of the creators of the works that are part of the system's training base. If musicians use them to enrich their repertoire, they risk being sued whenever another musician recognizes their song. The music industry is powerful and has well-designed methods to protect the rights of creators and performers. Time will tell if composing systems like Suno will survive or if the music machinery will destroy them.

2.4. Deepfakes

Deepfakes are synthetic media created by AI capable of editing or generating images, audio and video. There are many interesting examples of deepfakes, such as, for example, the video of Tom Cruise in the role of Iron Man. If it is not known that Iron Man is portrayed by Robert John Downey Jr, the viewer may not recognize the photo manipulation, as the video appears quite convincing. Deepfakes can be very harmful because they: create confusion and scepticism; initiate the spread of misinformation, pose a threat to privacy and security; convincingly impersonate anyone; orchestrate phishing scams; and identity theft operations with alarming precision (Stanford University IT, 2024).

The connection between deepfakes and copyright lies in the implications they have for intellectual property rights. US Digital Millennium Copyright Act (DMCA) states that: "Deep fakes that take advantage of copyrighted materials could face civil actions for copyright infringement." (U.S. Copyright Office, 2020). According to the recent Artificial Intelligence Act of European Parliament and Council (2024) "artificial or manipulated images, audio or video content ('deepfakes') need to be clearly labelled as such". The act also adds: "These include drawing up technical documentation, complying with EU copyright law and disseminating detailed summaries about the content used for training".

Under these laws, deepfakes are legitimate media as long as they are clearly labelled as such, and do not violate existing copyright laws. However, it is clear that vocal or audio imitations cannot be made without importing the original works. If they are copyrighted, and they almost always are, then the stakes would be high, especially in the US, where copyright infringement is protected by a federal law and the statutory damages can reach \$150,000 per infringed work.

2.5. AI-generated welcome speech

I recently wrote a welcome speech in a language that I spoke bilingually during my youth, but had not used in years. I prepared the welcome speech in my native language and asked the retired GPT-4 to translate it and to improve it. ChatGPT not only corrected it perfectly, but also embellished the sentences.

Although it was no longer my original text, I was pleased with the final result. The backbone of the text was still mine, but the writing style and part of the content was AI-generated. To better illustrate what I am discussing, here is a hypothetical example. First, I wrote a description of the speech and gave a couple of explanations of exactly what I wanted. Before ChatGPT offered me its response, I prepared my own welcome speech and asked ChatGPT to improve it. ChatGPT offered me two responses in both cases, due to space constraints, only the shorter one is shown. Here is how mutual communication took place:

My prompt to ChatGPT:

“Please write me a short welcome speech on the occasion of the 25th edition of the conference organized by my faculty. The conference began at the dawn of the 21st century as a team building event of the Institute of computing and survived the COVID-19. Thank all the participants, representatives of the business community, renowned lecturers, students, members of the program and organizational committee, and at the end declare the conference open.”

The AI-generated response based on my prompts:

“Ladies and gentlemen, esteemed guests, and participants, welcome to the 25th edition of our conference, a remarkable milestone for our faculty and the Institute of Computing. What began as a team-building event at the dawn of the 21st century has evolved into a significant platform for collaboration, innovation, and learning, even persevering through the challenges posed by COVID-19. I want to extend my heartfelt gratitude to all of you: our dedicated participants, representatives from the business community, our esteemed lecturers, enthusiastic students, and the hardworking members of our program and organizational committee. Your commitment and passion are what make this event a success. With that, I am delighted to declare the 25th conference officially open. Let’s make this a memorable and inspiring event together! Thank you.”

This is the draft of my welcome speech that I wrote before asking ChatGPT for help:

“Dear colleagues, dear guests, it is a great honour and pleasure for me to open the 25th anniversary edition of the conference, which began at the dawn of the 21st century as a team building meeting of the Institute of Computing. The conference went through various crises, including COVID-19, and over time turned into a renowned scientific gathering of researchers from all over the world. I am happy that this year the conference was graced by excellent invited lecturers,

a large number of experienced researchers and young students who are taking their first research steps in science. It is a special pleasure that the business community honoured us with their presence. I thank them all from the bottom of my heart. I would like to thank the members of the program committee who wrote valuable reviews, and the members of the organizing committee who were diligent helpers and worked hard. Finally, I sincerely thank all participants who sent valuable works. I declare the conference open and wish you to enjoy all the lectures and exhibitions in the upcoming 3 days and to see you again next year.”

Based on my writing, ChatGPT suggested the final welcome speech:

“Ladies and gentlemen, esteemed guests, and dear participants, it is both an honour and a pleasure to welcome you to the 25th edition of our faculty’s conference. As we gather here today, we celebrate not only a significant milestone but also the resilience and commitment that have brought us to this moment. This conference began at the dawn of the 21st century as a team-building event for the Institute of Computing, and over the years, it has evolved into a vibrant platform for collaboration, innovation, and knowledge sharing. The journey has not been without challenges, particularly as we navigated the unprecedented times brought on by the COVID-19 pandemic. Yet, here we are, stronger than ever, ready to engage, learn, and connect. I would like to extend my heartfelt gratitude to all the participants, representatives of the business community, our distinguished lecturers, dedicated students, and the tireless members of the program and organizational committee. Your contributions and support are invaluable, and it is your passion that fuels the spirit of this conference. As we embark on this exciting journey over the coming days, let us embrace the opportunities for learning and collaboration that lie ahead. Together, we can continue to push the boundaries of knowledge and innovation in our field.”

I find both AI-generated speeches extremely valuable, the first based on my detailed explanation, and the second based on my speech proposal. Both were derived because of my author’s works, so whichever answer I choose, my author’s share lies in the basis of the response. If I decide to use one of them, I must determine whether I have the right to hold the copyright. I cannot definitively answer that. I was happy to say the welcome speech that was the inspiration for this example, especially because ChatGPT used almost all my phrases, so the contribution of AI was primarily translation and polishing of the language, which any professional proof-reader would do without taking over my authorship. Moreover, it was not recorded, and therefore, no copyright is claimed. However, if that option is not disabled (OpenAI, 2024), ChatGPT will insert my query and my welcome speech into its training database, so that they will become the basis of some other similar responses that ChatGPT will generate in the future.

2.6. AI-generated research papers

In the first case study presented in this paper, it was mentioned that students are willing to use generative AI as an assistant or ghost writer for assignment preparation. A brief analysis of the title structure reveals a noticeable trend. Comparing titles from papers published ten years ago, or earlier, with those of more recent publications, it becomes apparent that the use of colons in titles has significantly increased. Specifically, when requesting title suggestions from ChatGPT, it is highly probable that at least five out of ten proposed titles will feature two parts separated by a colon.

Another noteworthy observation can be made. By limiting a Google Scholar search to the past two years, it becomes evident that certain terms, such as *Unveiling' Revolutionizing*, and *Bridging*, as well as the phrase *From Barriers to Bridges* have been used frequently from 2023 to the present, as well as between 2010 and 2022. These are not merely buzzwords, but terms favoured by ChatGPT. This can be easily verified through experimentation.

Recently, it has been observed that some research papers exhibit a writing style that significantly differs from previous works by the same authors. It is likely that these papers contain text generated by a LLM, rather than being solely authored by the individuals listed as the creators. Two studies explicitly confirm this impression (Gray, 2024; Liang et al., 2024). They were using approaches similar to those I presented about the welcome speech and the titles. In the first study, by examining LLMs with prompts containing selected keywords or titles, the proportion of generative AI usage was found to be less than 1% (Gray, 2024). In the second, by searching papers with a set of 12 adjectives, adverbs and controls, the Stanford team estimates that generative AI is used in 6.3 to 17.5 percent of papers, depending on the field of research (Liang et al., 2024). While the frequencies of most words moderately increased, the frequency of adjectives commendable, meticulous and intricate, and particularly the frequency of the adverb meticulously increased more than 100%.

If the apparent inclusion of generative AI is not just a harmless aesthetic intervention, and by all accounts it is not, then we are faced with a worrying misconduct in science. With the obvious presence of falsification and fabrications in these papers resulting from algorithmic bias (Ferrara, 2023), lack of quality control, and generative AI misinformation (Wach et al., 2023), another question arises: Do these researchers have the right to hold copyright? Definitely not, primarily because, as was already illustrated in the introduction of this paper, LLMs violate copyright, so everything they generate is plagiarism. Further, what is more than clear is that stolen intellectual property cannot be protected as one's own (Chesterman, 2025). In the paper, Chesterman addresses the theft of copyright by generative AI during the training phase, as well as the ownership, and consequently the copyright, of the results generated by the AI.

2.7. AI-generated paraphrasing

As mentioned in the case study on student assignments, paraphrasing is a technique used to mask plagiarism. On the other hand, it is directly related to copyright infringement. I will prove this claim by using the applications ChatGPT and Quillbot over the same paragraph twice sequentially. The original paragraph is an explanation of the objectives of the COST action LITHME (<https://lithme.eu/>). It consists of these five sentences:

“How will pervasive human-machine language interfaces affect language in areas such as international law, translation, and other forms of language work? What will this mean for how people identify with specific languages? Could increasing reliance on real-time language technologies actually change the structure of language? Longer term, could developments in brain-machine interfaces serve to complement or even supersede language altogether? Linguistics will be far stronger for robust technological foresight, while developers will bene-

fit from better understanding potential linguistic and societal consequences of their creations.”.

First, let us conduct a thorough examination of the paragraph for potential plagiarism. When the first sentence is searched using quotation marks, the result directs us to LITHME. A search without quotation marks leads to the COST page of LITHME (<https://www.cost.eu/actions/CA19102/>). In both cases, it converged to LITHME. Googling all sentences from the source paragraph behaved similarly and revealed no plagiarism. This is great because it allows the comparison of an original text that LITHME has two copyrights, for the COST action and the action website. Two experiments were done step-by-step. First, the source paragraph was paraphrased twice with ChatGPT, and then with Quilbot. These are the results of the paraphrasing:

ChatGPT paraphrase 1:

“In what ways will widespread use of human-machine language interfaces impact language related fields like international law, translation, and various other linguistic tasks? What implications might this have for people’s connection to particular languages? Could a growing dependence on real-time language technologies lead to changes in the structure of language itself? In the long run, could advancements in brain-machine interfaces either enhance or replace language entirely? A solid grasp of technological trends will strengthen linguistics, while developers will gain from understanding the possible linguistic and societal impacts of their innovations.”

ChatGPT paraphrase 2:

“What impact will widespread human-machine language interfaces have on language in fields like international law, translation, and other linguistic work? What effects might this have on people’s relationships with specific languages? Could an increased reliance on real-time language technologies alter the very structure of language? Looking ahead, might innovations in brain-machine interfaces either improve or even replace language as we know it? A clear understanding of technological developments will enhance the field of linguistics, while creators will benefit from recognizing the potential linguistic and societal effects of their work.”

I undertook my research using all the sentences, but the illustration of the plagiarism check will again be done with the first sentence only. All the words in the sentence, “In what ways will widespread use of human-machine language interfaces impact language-related fields like international law, translation, and various other linguistic tasks?” exist in the blogpost (Schoening, 2023), which is published on the Phrase localization platform (<https://phrase.com/>), which is obviously copyrighted. It can be considered a potluck plagiarism (Myers, 2018) with one major difference: the sentence was a compilation of several sentences from the same and not from different articles. It is possible that the article was used to train the paraphrasing tool, but the plagiarism checker did not detect this.

The first sentence from the second paraphrase: “What impact will widespread human-machine language interfaces have on language in fields like international law, translation, and other linguistic work?” resembles the paper by Vieira, O’Hagan, and O’Sullivan (2021), which is copyrighted by Taylor & Francis. This discloses the copyright controversy of this publishing house (Informa PLC, 2024). The exact focus of the matter

is unclear. Taylor & Francis are owned by Informa PLC. In May 2024, Informa entered into a data access agreement with Microsoft with a term from 2024 to 2027, whereby Microsoft can use their work to train its AI models. The partnership was entered into without the consent of the authors who have already published their research in Taylor & Francis and paid a fee in order for their works to receive open access status. As a result, the publisher faced criticism from authors who had waived their copyright by paying a fee, only to have it sold without consultation or compensation. This situation is, understandably, debatable.

The same paraphrasing experiment was replicated using the professional paraphrasing tool Quillbot. The next two paragraphs present the results.

Quillbot paraphrase 1:

“What effects would ubiquitous human-machine language interfaces have on language use in fields like translation, international law, and other language related tasks? What impact will this have on how individuals relate to particular languages? Could language structure alter as a result of a greater reliance on real-time language technologies? In the long run, may advancements in brain-machine interfaces help to supplement or even replace language? While developers will gain from a better awareness of the possible linguistic and societal repercussions of their creations, linguistics will be significantly stronger for robust technological foresight.”

Quillbot paraphrase 2:

“How might the widespread use of human-machine language interfaces affect language use in domains such as international law, translation, and other language-related tasks? What effect will this have on people’s perceptions of specific languages? Could an increased dependence on real-time language technologies lead to changes in language structure? Could developments in brain-machine interfaces eventually serve to augment or even replace language? Linguistics will be much stronger for solid technical foresight, while developers will benefit from a greater understanding of the potential linguistic and societal effects of their creations.”

I searched each sentence extensively from both paraphrases as a whole and as a combination of keywords. I have not found a single reference in which more than four to five words match. After that, I searched the web looking for any reactions against Quillbot regarding copyright infringement. This search did not reveal any kind of copyright infringement. Commendations are due to Quillbot, which rewrites and paraphrases texts based on AI algorithms without endangering other people’s intellectual property.

2.8. AI-assisted and AI-generated poetry

Using BERT, our young doctoral student managed to convincingly imitate the poetry of the bard of Macedonian literature and the founder of the standard Macedonian language, Blaže Koneski (1921 - 1993). Koneski died recently, so his works are still not part of the public domain. Apparently, no one is concerned about his copyright, neither the publishers who printed his works, nor his daughter, who is the only legal heir to his authorship. Moreover, it turns out that I am the only one who warned the PhD student about the possible copyright implications of his GenAI-generated works. He defends himself that he does not harm anyone, because he does not copyright or publish his

AI-generated poetry with copyright and does not use it for commercial purposes. Is he committing copyright infringement by using Koneski's copyrighted poetry for training his AI model? In my opinion, yes, although some believe that this only promotes Koneski's work.

This example prompted me to find out if there are any examples of imitating someone's poetry using an AI model. I have come across several scientific papers that present similar experiments. For example, the paper by Köbis and Mossink (2021) has already received over 300 citations, which proves the quality of their research. Their system was trained on the works of the famous poet Maya Angelou (1928 - 2014), whose site <https://www.mayaangelou.com/> is copyright protected, so it is very likely that the poetry is also protected by the same type of intellectual property protection. The paper does not give credit to its authorship, but only to the authors of the paper.

An interesting work is related to haiku poetry, which generative artificial intelligence, according to the research participants, was able to imitate (Hitsuwari, Ueda, Yun, & Nomura, 2023). This paper does not mention the authorship of the songs that were used during the training of the system. If it is ancient art, then these works are in the public domain, so there is no copyright infringement. Nevertheless, if the haiku poetry belongs to contemporary authors, then the problem is nothing but obvious.

I will conclude this case study with Hutson & Schnellmann's (2023) suggestions that a compromise must be found between human authorship and technology, primarily for the advancement of artificial intelligence or robotics. I personally do not agree with them at all, and especially with their position that one must always give priority to practical aspects over authorship (Hutson & Schnellmann, 2023).

2.9. AI-assisted and AI-generated novels

By searching Amazon books, I noticed that Amazon offers over 500 books related to AI-assisted content. The amount of novels generated by AI seems to be ten times higher. They mostly include works assisted by AI, where the role of the AI was to support the creation, not to write the novels. Amazon Kindle allows authors and publishers to upload eBooks to their platform, regardless of whether the content was generated by a human author or by AI (Amazon, 2024). Amazon requires authors to notify of AI-generated content when they publish a new book or edit and republish an existing book through Kindle Direct Publishing. Amazon transfers the responsibility of legal violation and violation of property rights to authors, publishers and sales partners who must guarantee that the work does not contradict the laws for the protection of intellectual property. Any additional use of protected works must be guaranteed with the written permission of the copyright holder.

The copyright for books created by generative AI is not mentioned by Amazon. On the other hand, there are many different AI applications that allow for the writing of novels. One example is the reporter at the Verge Adi Robertson (2023) (<https://www.theverge.com/authors/adi-robertson>), who went to great lengths to explain to Sudowrite (copyshark.ai) how to write the novel *The Electric Sea* for her (Tumblr, 2023). The Sudowrite website no longer exists and the Copyright Tools section was transferred towards WordPress. Therefore, I am unable to verify what the position is regarding the copyright of the AI-generated novel.

I watched *The Electric Sea* on Tumblr (<https://www.tumblr.com/>). I have to admit that this “cyberpunk novella” did not grab my attention. But sooner or later, similar AI tools will learn to write more excitingly, so their novels will become fun to read, especially for younger readers. As these books can only be created through the active cooperation of individuals, the authorship belongs to them. When they become completely autonomous, then copyright will become a serious problem.

3. Conclusions

Artificial intelligence is becoming an indispensable part of our everyday life. Based on human prompts, generative AI is capable of autonomously producing creative outputs and intellectual contributions. If generative AI were a human being or a natural person, those creations would be the author’s works. The question of who holds copyright for AI-assisted and AI-generated content is still unanswered.

The current legal framework, especially in the EU, is cautiously trying to balance innovation in AI, including general purpose AI (GPAI), with respect to copyright (Quintais, 2025). The EU AI Act does not change copyright law, but reinforces the need for transparency and lawful data use. Creators have tools to opt out of data mining and to assert their rights, but these mechanisms are still evolving and face many challenges (Lucchi, 2024). This primarily refers to the protection of the integrity of the original works used for training AI models and ensuring adequate compensation if they are not in the public domain.

What can be said with certainty is that translating a work using machine translation does not transfer the copyright to machine translation systems. The same applies to all LLMs with translation abilities. Similarly, the texts, images, and videos generated by LLMs are not copyrighted by that LLM. They belong to the user.

What can be learnt from this paper is that art generated by AI is prohibited for copyright protection because it does not meet the requirement of human authorship according to current laws. Artificial or manipulated images, audio or video content created by AI can become copyrighted works by people who have asked the AI to create them. The only prerequisite is that the share of artificial intelligence must be clearly marked. Book authors are also responsible for reporting content generated by AI.

The greater the contribution of people, the more they have the right to claim authorship. On the other hand, generative AI is becoming more autonomous every day. This leads every human person, especially the inexperienced, to the temptation to appropriate work in which they have invested almost no share. The tension between integrity and ambition highlights a profound dilemma: the desire for greatness can sometimes lead people down a morally ambiguous path. As people navigate their own aspirations, it becomes crucial to reflect on the means they choose to pursue their goals, balancing ambition with ethical standards.

Many new regulatory frameworks have been created to prevent the misuse of AI-assisted and AI-generated content. Training on copyrighted material is a legally grey area. Some argue it’s fair use or allowed under text and data mining (TDM) exceptions, others see it as infringement. Regulatory frameworks can address the concerns related to copyright infringement. Their effect is to protect the true authorship, while preventing the automated creation of both persuasive and irrelevant content.

For now, creators should actively assert their rights wherever possible and stay informed as laws continue to develop. Meanwhile, AI developers must prioritize transparency, fair use, and ethical data practices to ensure compliance and public trust. In the long term, harmonization between AI regulations and copyright laws will be essential. They should explicitly distinguish AI-assisted from AI-generated works.

Laws will soon become stricter to bring order to the sudden influx of new technologies. They will establish licensing frameworks for training data. Moreover, they might suggest easier global mechanisms for creators to control the use of their works in AI systems. But, we must be aware that Walter Mizner was right when he said, "If you steal from one author, it's plagiarism; if you steal from many it's research" (Johnston, 1953, p. 66). Generative AI has thus become a high-calibre researcher. This means that the real challenge is to prevent fraudsters who will be tempted to take the results of generative AI as their own. The obligation of researchers is to prevent them from doing so, at least by avoiding becoming consumers of infringing works (Lucchi, 2024).

Otherwise, an avalanche of stolen works will prevail over the real results of human creativity, mental creations, literary and artistic works. Then the creation of "organic" intellectual property will become extremely rare, which will demotivate most human authors to create new works. I am an optimist that over time "organic" intellectual property will become extremely respected and that its price will become high due to the fact that without new original works even artificial intelligence will not be able to advance a single step forward. This will result in a great regression of humanity. Human civilization has successfully overcome greater trials, so it will surely overcome this too.

References

- Amazon. (2024). *Content Guidelines*. Retrieved 2025-06-26, from https://kdp.amazon.com/en_US/help/topic/G200672390
- Bagenal, J. (2024). Generative artificial intelligence and scientific publishing: Urgent questions, difficult answers. *The Lancet*, 403(10432), 1118–1120. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S0140673624004161> doi: 10.1016/S0140-6736(24)00416-1
- Chesterman, S. (2025). Good models borrow, great models steal: Intellectual property rights and generative AI. *Policy and Society*, 44(1), 23–37. Retrieved 2025-12-05, from <https://academic.oup.com/policyandsociety/article/44/1/23/7606572> doi: 10.1093/polsoc/puae006
- Chu, T., Song, Z., & Yang, C. (2024, March). How to protect copyright data in optimization of large language models? In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, pp. 17871–17879). Retrieved 2025-12-05, from <https://arxiv.org/abs/2308.12247> (Version Number: 1) doi: 10.48550/ARXIV.2308.12247
- Crews, K. D. (2020). *Copyright Law for Librarians and Educators: Creative Strategies and Practical Solutions*. Chicago: American Library Association.
- EJOI MK. (2024). *Odime po medalji*. Retrieved 2025-06-26, from <https://motika.mk/wp-content/uploads/2024/07/05/ejoi.mp4>
- EU. (2019). *Directive (EU) 2019/790 of the European parliament and of the Council*. Retrieved 2025-06-26, from <https://eur-lex.europa.eu/legalcontent/EN/TXT/HTML/?uri=CELEX:32019L0790>

- European Parliament and Council. (2024). Artificial Intelligence Act. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June. In *Official Journal of the European Union, L 1689, July 12*. Retrieved 2025-06-26, from <https://artificialintelligenceact.eu/>
- Ferrara, E. (2023). Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *Sci*, 6(1), 3. Retrieved from <https://www.mdpi.com/2413-4155/6/1/3> doi: 10.3390/sci6010003
- Frank, J., Herbert, F., Ricker, J., Schönherr, L., Eisenhofer, T., Fischer, A., ... Holz, T. (2024). A representative study on human detection of artificially generated media across countries. In *2024 IEEE Symposium on Security and Privacy (SP)* (pp. 55–73). IEEE. Retrieved from <https://arxiv.org/abs/2312.05976> (Version Number: 1) doi: 10.48550/ARXIV.2312.05976
- Free Software Foundation. (2022). *What is copyleft*. Retrieved 2025-06-26, from <https://www.gnu.org/licenses/copyleft.en.html>
- Gopalakrishna, G., Ter Riet, G., Vink, G., Stoop, I., Wicherts, J. M., & Bouter, L. M. (2022). Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in The Netherlands. *PLOS ONE*, 17(2), e0263023. Retrieved from <https://dx.plos.org/10.1371/journal.pone.0263023> doi: 10.1371/journal.pone.0263023
- Gray, A. (2024). *ChatGPT "contamination": Estimating the prevalence of LLMs in the scholarly literature*. arXiv. Retrieved from <http://arxiv.org/abs/2403.16887> (arXiv:2403.16887 [cs]) doi: 10.48550/arXiv.2403.16887
- Hernández-Huerta, V. A. (2020). Disputed Elections in Presidential Democracies: Contexts of Electoral “Blackmail”. *The Journal of Politics*, 82(1), 89–103. Retrieved from <https://www.journals.uchicago.edu/doi/10.1086/705599> doi: 10.1086/705599
- Hesmondhalgh, D. (2018). *The Cultural Industries* (Fourth edition ed.). Los Angeles: SAGE Publications Ltd.
- Hitsuwari, J., Ueda, Y., Yun, W., & Nomura, M. (2023). Does human–AI collaboration lead to more creative art? Aesthetic evaluation of human-made and AI-generated haiku poetry. *Computers in Human Behavior*, 139. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S0747563222003223> doi: 10.1016/j.chb.2022.107502
- Hutson, J., & Schnellmann, A. (2023). The poetry of prompts: the collaborative role of generative artificial intelligence in the creation of poetry and the anxiety of machine influence. *Global Journal of Computer Science and Technology: D*, 23(1).
- Informa PLC. (2024). *Press Release*. Retrieved 2025-06-26, from <https://www.informa.com/globalassets/documents/investorrelations/2024/informa-plc---market-update.pdf>
- Jaszi, P. (2017). Toward a theory of copyright: The metamorphoses of “authorship”. In *Intellectual Property Law and History* (pp. 61–108). Routledge.
- Johnston, A. (Ed.). (1953). *Chapter 4. In: The legendary Mizners* (pp. 66). New York: Farrar, Straus and Giroux.
- Justia. (2023). *Andersen et al v. Stability AI Ltd. et al*. Retrieved 2025-06-26, from <https://docs.justia.com/cases/federal/districtcourts/california/candce/3:2023cv00201/407208/67>
- Köbis, N., & Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Ex-

- perimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, 114, 106553. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S0747563220303034> doi: 10.1016/j.chb.2020.106553
- Liang, W., Zhang, Y., Wu, Z., Lepp, H., Ji, W., Zhao, X., ... Zou, J. Y. (2024). *Mapping the increasing use of LLMs in scientific papers*. arXiv. Retrieved 2025-12-05, from <https://arxiv.org/abs/2404.01268> (Version Number: 1) doi: 10.48550/ARXIV.2404.01268
- Library of Congress. (2023). *Copyright registration guidance: Works containing material generated by artificial intelligence*. Retrieved 2025-06-26, from <https://www.federalregister.gov/documents/2023/03/16/2023-05321/copyright-registration-guidance-works-containing-material-generated-by-artificial-intelligence>
- Lucchi, N. (2024). ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems. *European Journal of Risk Regulation*, 15(3), 602–624. Retrieved from https://www.cambridge.org/core/product/identifier/S1867299X23000594/type/journal_article doi: 10.1017/err.2023.59
- Lunney Jr, G. S. (2001). The death of copyright: Digital technology, private copying, and the digital millennium copyright act. *Virginia Law Review*, 813–920.
- Myers, C. S. (2018). Plagiarism and copyright: Best practices for classroom education. *College & Undergraduate Libraries*, 25(1), 91–99. Retrieved 2025-12-05, from <https://www.tandfonline.com/doi/full/10.1080/10691316.2017.1391028> doi: 10.1080/10691316.2017.1391028
- Nicholls, A. R., Madigan, D. J., Duncan, L., Hallward, L., Lazuras, L., Bingham, K., & Fairs, L. R. W. (2020, September). Cheater, cheater, pumpkin eater: the Dark Triad, attitudes towards doping, and cheating behaviour among athletes. *European Journal of Sport Science*, 20(8), 1124–1130. Retrieved 2025-12-05, from <https://onlinelibrary.wiley.com/doi/10.1080/17461391.2019.1694079> doi: 10.1080/17461391.2019.1694079
- Ochoa, T. T. (2002). Origins and meanings of the public domain. *U. Dayton L. Rev*, 28, 215.
- OpenAI. (2024). *Europe Terms of Use*. Retrieved 2025-06-26, from <https://openai.com/policies/terms-of-use/>
- Panettieri, J. (2024). *Generative AI Lawsuits Timeline: Legal Cases vs. OpenAI, Microsoft, Anthropic, Nvidia, Perplexity, Intel and More*.
- Pila, J. (2010). Copyright and Its Categories of Original Works. *Oxford Journal of Legal Studies*, 30(2), 229–254. Retrieved from <https://academic.oup.com/ojls/article-lookup/doi/10.1093/ojls/gqq009> doi: 10.1093/ojls/gqq009
- Quintais, J. P. (2025). Generative AI, copyright and the AI Act. *Computer Law & Security Review*, 56, 106107. Retrieved 2025-12-05, from <https://linkinghub.elsevier.com/retrieve/pii/S0267364925000020> doi: 10.1016/j.clsr.2025.106107
- Resnikoff, P. (2024). *UMG, WMG, Sony Music File Litigation Against AI Music Services Suno and Udio for Massive Copyright Infringement*. Retrieved 2025-06-26, from <https://www.digitalmusicnews.com/2024/06/24/umg-wmgsony-litigation-ai-music-suno-udio/>
- Robertson, E. (2023). *I tried the AI novel-writing tool everyone hates, and it's better*

- than I expected. Retrieved 2025-06-26, from <https://www.theverge.com/2023/5/24/23732252/sudowritestory-engine-ai-generated-cyberpunk-novella>
- Schilit, H. M., & Perler, J. (2010). *Financial Shenanigans* (3rd ed.). McGraw-Hill.
- Schoening, S. (2023). *Machine Translation Explained: Types, Use Cases, and Best Practices*. Retrieved 2025-06-26, from <https://phrase.com/blog/posts/machine-translation/>,
- Spink, J., Moyer, D. C., Park, H., & Heinonen, J. A. (2013). Defining the types of counterfeiters, counterfeiting, and offender organizations. *Crime Science*, 2(1), 8. Retrieved from <https://crimesciencejournal.biomedcentral.com/articles/10.1186/2193-7680-2-8> doi: 10.1186/2193-7680-2-8
- Stanford University IT. (2024). *Dangers of Deepfake: What to Watch For*. Retrieved 2025-06-26, from <https://uit.stanford.edu/news/dangersdeepfake-what-watch>
- Temple, K. (2003). *Scandal Nation: Law and Authorship in Britain, 1750–1832*. Ithaca, NY: Cornell University Press.
- Tumblr. (2023). *The AI Writers Workshop*. Retrieved 2025-06-26, from <https://robotlit.com/post/717956168948023296/the-electric-sea-an-ai-generated-experiment>
- U.S. Copyright Office. (1976). *The Lifecycle of Copyright*. Retrieved 2025-06-26, from <https://www.copyright.gov/history/copyrightexhibit/lifecycle/>
- U.S. Copyright Office. (2020). *The Digital Millennium Copyright*. Retrieved 2025-06-26, from <https://www.copyright.gov/dmca/>
- U.S. Copyright Office. (2024). *Copyright Law of the United States (Title 17)*. Retrieved 2025-06-26, from <https://www.copyright.gov/title17/>
- U.S. Department of Justice. (2022). *CRM 1500-1999*. Retrieved 2025-06-26, from <https://www.justice.gov/archives/jm/criminal-resource-manual1852-copyright-infringement-penalties-17-usc-506a-and-18-usc-2319>
- Varian, H. R. (2005). Copying and Copyright. *Journal of Economic Perspectives*, 19(2), 121–138. Retrieved 2025-12-05, from <https://pubs.aeaweb.org/doi/10.1257/0895330054048768> doi: 10.1257/0895330054048768
- Vieira, L. N., O'Hagan, M., & O'Sullivan, C. (2021). Understanding the societal impacts of machine translation: a critical review of the literature on medical and legal use cases. *Information, Communication & Society*, 24(11), 1515–1532. Retrieved from <https://www.tandfonline.com/doi/full/10.1080/1369118X.2020.1776370> doi: 10.1080/1369118X.2020.1776370
- Wach, K., Duong, C. D., Ejdys, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., ... Ziemia, E. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review*, 11(2), 7–30. Retrieved 2025-12-05, from <https://eber.uek.krakow.pl/index.php/eber/article/view/2113> doi: 10.15678/EBER.2023.110201
- WIPO. (1979). *Berne Convention for the Protection of Literary and Artistic Works*. Retrieved 2025-06-26, from <https://www.wipo.int/wipolex/en/treaties/textdetails/12214>
- WIPO. (1996). *Copyright*. Retrieved 2025-06-26, from <https://www.wipo.int/copyright/en/>
- Woody, K. E. (2020). The new insider trading. *Ariz. St. LJ*, 52.
- Young, L. (2024). *What are the new and emerging trends in academic misconduct?* Retrieved 2025-06-26, from <https://www.turnitin.com/blog/what-are-the-new-andemerging-plagiarism-trends>

- Yusuf, A., Pervin, N., & Román-González, M. (2024). Generative AI and the future of higher education: a threat to academic integrity or reformation? Evidence from multicultural perspectives. *International Journal of Educational Technology in Higher Education*, 21(1), 21. Retrieved from <https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-024-00453-6> doi: 10.1186/s41239-024-00453-6
- Zdravkova, K. (2023). Evolution of academic dishonesty in computer science courses. In *9th International Conference on Higher Education Advances (HEAd'23)* (pp. 421–428). Universitat Politècnica de València. Retrieved from <http://ocs.editorial.upv.es/index.php/HEAD/HEAd23/paper/view/16081> doi: 10.4995/HEAd23.2023.16081
- Zdravkova, K., & Ilijoski, B. (2025). Preventing Academic Dishonesty Originating from Large Language Models. In C. Bădică, M. Gušev, A. Iftene, M. Ivanović, Y. Manolopoulos, & S. Xinogalos (Eds.), *Advances in ICT Research in the Balkans* (Vol. 2391, pp. 118–132). Cham: Springer Nature Switzerland. Retrieved from https://link.springer.com/10.1007/978-3-031-84093-7_9 (Series Title: Communications in Computer and Information Science) doi: 10.1007/978-3-031-84093-7_9