Language and Law / Linguagem e Direito, Vol. *12*(1), 2025, pp. 39-60
DOI: doi.org/10.21747/21833745/lanlaw12_1a2

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

Aneta Skorupa-Wulczynska [1] and Anna Zannina [2]

[1] *Chancellery of the Sejm*, aiskorupa@gmail.com
[2] *University 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-

[1]The 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 Láncos (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

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 PressOxford. 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 sensored 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/

Thellmann, 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.europarl.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/