



Dataset Bias in AI: A Strategic Priority for Language Service Companies

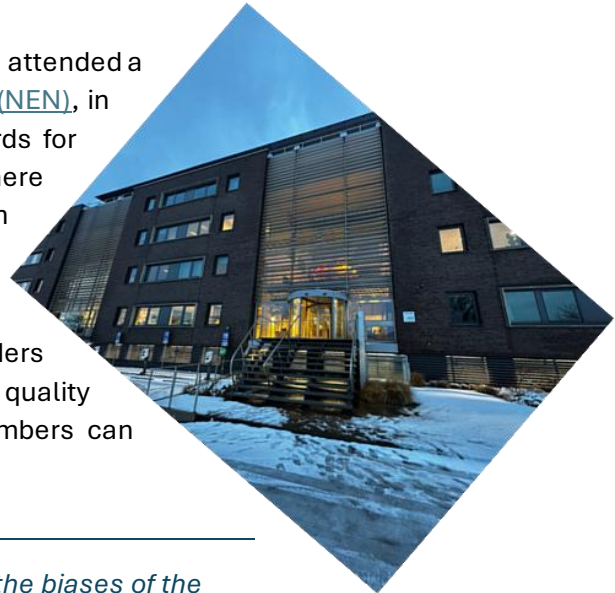
By Carina Balbo, ATC ISO Committee

Happy New Year to all our member companies! I, on behalf of all the ATC team, sincerely hope 2026 will be a year of growth.

As the ATC's ISO representative, I have taken on the task of researching and collaborating with ISO standards that are relevant to our industry. Briefly, I am a member of three mirror committees of the [British Standards Institution \(BSI\)](#): TS/1 (language and terminology), ART/1 (artificial intelligence) and IST/43 (information technology for learning, education and training). More on this in future blogs.

In early January, with my ART/1 committee member hat on, I attended a meeting at the [Royal Netherlands Standardization Institute \(NEN\)](#), in Delft, the Netherlands focused on establishing AI standards for **datasets** and **bias** intended to reinforce the EU AI Act. There were very interesting and stimulating discussions which certainly brought some heat into the room on one of the coldest days in 2026!

So, what were the key takeaways for language service providers and our daily operations? This article explores why dataset quality and bias matter, what risks they pose, and how ATC members can respond strategically.



AI systems inherit the characteristics and the biases of the datasets they are trained on.

Artificial intelligence is accelerating change across the translation and localisation industry. From MT-driven workflows to AI-enhanced CAT tools and speech-to-speech interpreting platforms, member companies are navigating a rapidly evolving technological landscape. Yet beneath this wave of innovation lies a factor that demands our collective attention: the quality and representativeness of the datasets powering these systems.

For language service companies, this is no longer an abstract technical concern. It is a business-critical issue that affects quality assurance, client trust, operational efficiency, and the long-term positioning of the industry.



The Quality Impact: Biased Data Creates Biased Output

Machine translation engines and language models learn from vast multilingual corpora. When these corpora are unbalanced or unrepresentative, the resulting systems reproduce those patterns.

Machine translation engines and language models train on massive multilingual corpora. When these datasets are unbalanced or fail to represent linguistic diversity, the systems perpetuate those limitations. The consequences show up in your workflows as gender stereotyping in professional contexts, cultural misinterpretations that undermine brand voice, flattened dialects in languages with rich regional variation, and inaccuracies in sensitive domains like legal, medical, or public sector communications.

These issues translate directly into higher post-editing costs, increased quality assurance overhead, reputational risk for your clients, and potential non-compliance with regulatory or accessibility requirements. Dataset bias becomes a quality risk that LSCs must actively manage rather than passively accept.

CAT Tools and MT Integrations: Dependant on Data Representativeness

Today's AI-powered CAT tools offer impressive capabilities including predictive suggestions, automated terminology extraction, and quality estimation. But these features depend entirely on the representativeness of their underlying data. When datasets are skewed, your tools may misinterpret domain-specific terminology, reinforce dominant linguistic patterns while marginalising minority varieties, produce inconsistent suggestions across projects, and underserve smaller or emerging languages.

This directly affects translator productivity, reviewer confidence, and the quality you deliver to clients.

Interpreting Technology: Where Bias Becomes an Equity Issue

Speech-to-text and real-time interpreting platforms rely on acoustic datasets that should reflect the full spectrum of human voices. Bias in these datasets manifests as poor recognition of non-standard accents, reduced accuracy for underrepresented languages, misinterpretation of culturally specific politeness strategies, and accessibility barriers in multilingual settings.

For interpreting providers, this transcends technical performance and becomes a matter of inclusion and equity.

The Strategic Business Case

As clients increasingly integrate AI into their content workflows, they expect LSCs to bring not only linguistic expertise but also AI literacy to the table. Companies that ignore dataset bias face mounting risks including misaligned client expectations around MT quality, inefficient post-editing workflows, lower translator satisfaction and retention, reduced competitiveness in tenders, and erosion of trust when clients demand transparency around AI use.

Conversely, LSCs who understand and actively mitigate bias can position themselves as trusted advisors in responsible AI adoption. This is a differentiator in a crowded market.



The Regulatory Dimension

The governance landscape is evolving rapidly. AI frameworks including the EU AI Act increasingly require transparency around data sources, documented bias mitigation strategies, human oversight mechanisms, and robust quality processes. For ATC members already operating within ISO-aligned frameworks, this shift represents an opportunity rather than a burden. Our industry's established commitment to quality, process rigor, and accountability can become a competitive advantage in the age of AI.

What Recent Industry Discussions Tell Us



Conversations at the Royal Netherlands Standardization Institute reinforced a critical point: as AI reshapes our sector, dataset integrity and bias have moved from technical considerations to strategic imperatives. For LSCs, ensuring high-quality, representative data directly influences not only the fairness and accuracy of AI-driven services, but also client trust, operational success, and regulatory compliance.

The message is clear: every LSC needs to address dataset quality and bias proactively within their workflows and service models.

Leading the Conversation

Our industry was built on cultural intelligence, linguistic expertise, and trust. AI doesn't diminish the importance of these foundations; it amplifies the need for them. By engaging proactively with dataset quality and bias, language service companies can strengthen their value proposition, protect clients from reputational and compliance risks, support linguists with transparent and fair workflows, influence the development of more inclusive AI tools, and lead the broader conversation on responsible AI in multilingual communication.

This is our moment to shape the future of our industry rather than simply respond to it.

How the ATC Can Support You

The Association of Translation Companies is committed to helping members navigate the opportunities and challenges of AI. Whether your organisation wants to build responsible AI workflows, understand how dataset bias affects your specific services, strengthen quality and governance frameworks, or educate clients on the realities of AI-enabled localisation, we're here to support that journey.

The challenges ahead are complex, but they're also opportunities to lead, differentiate, and advance together. The ATC is committed to ensuring our industry has a seat at the table as AI standards are developed and implemented. We're actively working to involve members and partners in shaping these frameworks, because the future of multilingual communication should be guided by those who understand it best. Join us in this important work.

Read more about ISO standards development for language services at

<https://atccertification.com/about-iso-standards/>