Sophia Toubian

IS 291B

March 16, 2025

AI-Driven Metrics for Information Translation in Science Communication

Scientific communication plays a pivotal role in bridging the gap between specialized knowledge and public understanding. However, when scientific information is translated for general audiences, it often undergoes significant modifications that can lead to distortions, oversimplifications, and misinterpretations. This challenge is compounded by the growing role of artificial intelligence (AI) in shaping how information is categorized, disseminated, and understood. As Yuval Noah Harari cautions, AI is increasingly determining "what is considered true knowledge," raising concerns about the reliability and objectivity of automated systems in knowledge mediation (Harari, 2022). At the same time, researchers such as Timnit Gebru highlight the limitations of AI's linguistic comprehension, emphasizing that current models function as "stochastic parrots" that predict word sequences rather than grasping the contextual depth of language (Bender & Gebru, 2021). These critiques are echoed by Kate Crawford, who argues that AI systems reinforce structural biases, disproportionately amplifying dominant narratives while marginalizing alternative perspectives (Crawford, 2016). Given these concerns, this paper explores how AI-driven methods can be utilized to assess scientific translation while ensuring accuracy, accessibility, and contextual fidelity.

The transformation of scientific information occurs through multiple processes, including translation, transmutation, and transposition. Translation refers to the adaptation of complex scientific language into more accessible formats without fundamentally altering the core meaning. Transmutation, however, involves deeper ideological, commercial, or political influences that reshape the original content, often resulting in misleading narratives.

Transposition entails the recontextualization of scientific knowledge into new discursive environments, which can lead to misinterpretation or unintended implications. Crawford underscores the dangers of these shifts, warning that AI-driven systems tend to reinforce hegemonic frameworks rather than critically evaluating the accuracy of knowledge dissemination (Crawford, 2016). Consequently, the role of AI in scientific translation must be scrutinized to ensure that it does not inadvertently distort public understanding.

This paper aims to address the following research questions:

1. Can AI help create a metric that can qualify and quantify the degree of information loss or distortion in scientific translation?
2. What computational methods (NLP, network analysis, readability scoring) are most effective in tracking these changes?
3. How do AI-driven assessments compare to expert human evaluations?

The primary objectives of this research are to:

1. Develop a tool for measuring information transformation
2. Apply the methodology using case studies
3. Create a foundation for broader application across scientific disciplines

Science communication faces several challenges that impact how information is translated. The first is the tension between precision and accessibility. Scientists must simplify ideas without losing critical accuracy. This balance is difficult to achieve and often results in oversimplification or overcomplication. The second has to do with framing and rhetoric. Different groups (media, corporations, activists) may intentionally frame information for strategic purposes. This can include: academic vs. public controversies and differences in their manifestations, intelligent design as a method of reduction in translation, and translation as a mechanism for controversy reduction but also possibly increase (e.g., misinformation, disinformation, purposely inflammatory social media posts)

AI can map how scientific ideas spread from academic publications to the media and social networks. This approach involves: mapping concept relationships, tracking concept evolution over time, and visualizing information flow between academic and public spheres. Through network analysis we can identify which institutions, media outlets, or influencers amplify distortions or maintain fidelity to original scientific concepts.

To illustrate these concerns in greater depth, consider the case study of epigenetics, a rapidly evolving field that explores the biochemical mechanisms regulating gene expression in response to environmental stimuli. Within the academic domain, epigenetics is rigorously studied through controlled experiments, molecular analyses, and computational modeling to understand how factors like diet, stress, and exposure to toxins can influence gene activity without altering the underlying DNA sequence. Researchers emphasize the complexity of these mechanisms, highlighting that epigenetic modifications, such as DNA methylation and histone modification, operate within a highly intricate and context-dependent framework.

However, as scientific discoveries in epigenetics gain public attention, their representation often diverges from the nuanced academic discourse. In popular discussions, epigenetics is frequently oversimplified or misrepresented, leading to exaggerated claims that lifestyle choices can “rewrite DNA” or that specific behaviors can be directly inherited by future generations. While there is some basis for understanding how environmental factors may influence gene expression over time, such interpretations often neglect critical scientific limitations, such as the transient nature of many epigenetic modifications and the complex interplay between genetic and environmental influences. This distortion can result in the proliferation of misleading narratives, influencing public perceptions, policy decisions, and even personal health choices based on incomplete or misrepresented science.

Given these challenges, AI-based natural language processing (NLP) tools offer a potential solution by systematically tracking semantic shifts between academic and public representations of scientific concepts. NLP techniques can be used to analyze vast textual corpora, identifying how key concepts are retained, modified, or distorted across different communicative domains. However, as Gebru et al. argue, AI systems often struggle with linguistic nuance, particularly when interpreting subtle distinctions between scientific precision and popular simplification (Bender & Gebru, 2021). The challenge lies in ensuring that AI-driven analyses do not themselves introduce new misinterpretations, as algorithmic biases can inadvertently reinforce pre existing distortions rather than clarify scientific meaning. This limitation highlights the need for robust AI-driven evaluation frameworks that can effectively distinguish between necessary simplifications essential for public accessibility and problematic distortions that mislead audiences.

A potential solution to these challenges is the development of an Accessibility-Accuracy (A/A) Matrix, which serves as a systematic framework for evaluating how scientific concepts are translated for public consumption. This matrix incorporates multiple dimensions of scientific communication, allowing for a more structured assessment of how well scientific accuracy is maintained while ensuring that information remains comprehensible to non-experts. The A/A Matrix includes four primary components: accuracy assessment, accessibility measurement, complexity reduction analysis, and controversy tracking. Accuracy assessment evaluates how faithfully key scientific concepts are preserved when translated into popular discourse. NLP techniques, including semantic similarity analysis, can be employed to compare claims, relationships, and contextual meanings between original academic texts and their public adaptations, providing a quantitative assessment of how closely public representations align with established scientific knowledge. Accessibility measurement assesses the readability of public-facing scientific translations. Since scientific concepts often contain highly specialized terminology, AI tools can apply automated readability formulas (e.g., Flesch-Kincaid, SMOG) to measure linguistic complexity while also analyzing the prevalence of domain-specific jargon. An optimal balance ensures that content remains both scientifically rigorous and broadly comprehensible. Complexity reduction analysis identifies which aspects of the original research are preserved, reinterpreted, or omitted in public discussions. While simplification is necessary for effective science communication, excessive reductionism can lead to misrepresentation. Topic modeling and concept mapping techniques can highlight the trade-offs between accessibility and accuracy. Finally, controversy tracking reveals how neutral scientific findings may be reframed in polarizing ways when adapted for broader audiences, particularly in areas where ideological, political, or commercial interests intersect with scientific knowledge. Sentiment analysis and opinion mining can detect shifts in framing, exposing how scientific conclusions are repackaged into contentious narratives.

By integrating these four dimensions, AI models can be trained to systematically detect distortions in scientific translation. The resulting AI system can provide a quantifiable score for various media representations of scientific topics, offering an evidence-based approach to evaluating the integrity of science communication. However, the reliability of such an AI-driven system ultimately depends on the quality of the datasets used for training, underscoring the importance of human oversight to ensure contextual integrity. Beyond textual analysis, network analysis offers an additional methodological approach for understanding how scientific concepts evolve as they move through different media landscapes.

By tracing the flow of information from peer-reviewed studies to news articles, social media discussions, and opinion pieces, researchers can visualize how scientific ideas are framed and reframed over time. This approach enables the identification of key influencers in scientific dissemination, highlighting which sources contribute to accurate reporting and which are more likely to introduce distortions. However, as Harari (2022) warns, the increasing reliance on AI-driven content curation and algorithmic filtering presents new challenges for scientific communication. Many digital platforms prioritize engagement over accuracy, creating “epistemic bubbles” in which audiences are selectively exposed to information that aligns with their existing beliefs. This phenomenon can lead to the reinforcement of misinformation, making it more difficult for accurate scientific knowledge to reach a broad audience. Ensuring transparency in AI-driven knowledge dissemination is therefore critical for maintaining public trust in scientific communication.

To operationalize these ideas, this paper proposes a methodology for systematically analyzing how epigenetics is represented in public discourse. This involves constructing two distinct textual corpora: an academic corpus comprising peer-reviewed scientific papers, conference proceedings, and academic books focusing on epigenetics, and a public corpus consisting of news articles, blogs, social media posts, and popular science publications discussing epigenetics. Using an AI-driven approach, this study seeks to compare semantic differences between academic and public representations of epigenetics, identify which sources are most prone to distort or simplify scientific information, and trace the evolution of specific epigenetic concepts as they transition from academic literature to mainstream discourse.

The analysis follows a four-stage process. The first stage involves corpus creation and preprocessing, in which both academic and public texts related to epigenetics are collected, cleaned, and prepared for computational analysis. The second stage applies NLP techniques such as word embeddings to track semantic shifts, topic modeling to identify key themes, and sentiment analysis to assess changes in framing. In the third stage, network analysis is used to map concept relationships across academic and public sources, visualizing how information flows between different media outlets and stakeholders. Finally, in the fourth stage, the A/A Matrix is applied to score public representations of epigenetics based on accuracy, accessibility, complexity reduction, and controversy, enabling the identification of common translation patterns and distortions.

While this paper focuses on epigenetics, the proposed methodology has broader applications across various scientific disciplines. Several case studies demonstrate the versatility of this approach. Maslow's Hierarchy of Needs, originally developed as a psychological framework, has been widely adapted for business marketing, education, and self-help, often in ways that oversimplify its theoretical foundations. The Theory of Evolution by Natural Selection, a cornerstone of biology, is frequently reinterpreted in educational, religious, and ideological debates, leading to persistent public misconceptions. Game Theory, initially a mathematical framework for competitive and cooperative strategies, is now commonly applied in economics, business, and even popular culture, sometimes deviating significantly from its original theoretical constructs. By applying AI-driven analysis to these and other scientific concepts, researchers can systematically assess how knowledge is communicated, identify points of distortion, and develop strategies for improving the accuracy and accessibility of public science communication. Ultimately, the goal is to foster a more informed public discourse, ensuring that scientific knowledge remains both rigorous and widely comprehensible.

Despite the powerful analytical capabilities that AI offers for measuring scientific translation, several ethical considerations require careful attention. A fundamental question concerns who defines what constitutes an “accurate” translation, as determining whether a simplification represents a legitimate adaptation or a problematic distortion involves subjective judgments that AI systems may not be equipped to make independently. Additionally, AI systems trained on existing scientific communications risk reproducing biases already present in those communications. The inherent tension between making content accessible and maintaining scientific accuracy presents another challenge, as AI metrics must avoid privileging one quality at the expense of the other. Perhaps most concerning is the potential for misuse, where AI-based accuracy scores could be weaponized to dismiss valid but controversial scientific perspectives or to systematically elevate certain viewpoints over others in public discourse.

Another case study that highlights the stakes of AI-driven scientific translation is climate science. The field of climate research has long been subject to misinformation and politically motivated distortions, often exacerbated by selective reporting and the framing strategies of various stakeholders. AI-driven sentiment analysis can help detect ideological biases in climate discourse by categorizing how different media sources frame scientific findings. Yet, as Crawford argues, AI systems themselves are not neutral arbiters of truth; rather, they reflect the priorities and biases of those who design and train them (Crawford, 2016). This raises questions about whether AI can truly serve as an impartial tool for evaluating scientific translation or if it will instead perpetuate pre-existing biases within knowledge ecosystems.

Maslow's Hierarchy of Needs provides yet another example of scientific translation in action. Originally developed as a psychological theory describing human motivation, Maslow's framework has been widely adapted into business strategies, educational curricula, and self-improvement narratives. In many of these adaptations, the hierarchical structure of the original model is oversimplified or misrepresented to serve commercial interests. AI-based linguistic analysis can be used to quantify these deviations, tracking how key psychological principles are reframed in different contexts. However, as Gebru et al. note, AI lacks the ability to critically evaluate conceptual integrity, often treating all linguistic patterns as equally valid regardless of their epistemic accuracy (Bender & Gebru, 2021). This further illustrates the limitations of AI-driven assessment models and the need for human oversight in scientific translation evaluations.

The potential risks of AI-driven translation metrics extend beyond misinterpretation and bias; they also involve broader ethical concerns regarding knowledge authority and censorship. If AI-generated accuracy scores become widely adopted, they could be leveraged to suppress alternative interpretations of scientific research, particularly in politically contentious areas such as public health and technological ethics. As Harari warns, unchecked AI-driven knowledge validation could lead to a scenario in which algorithmic decisions dictate the boundaries of accepted scientific discourse, sidelining human judgment in favor of automated authority (Harari, 2022). To mitigate these risks, AI-driven translation assessments must be designed with transparency, adaptability, and accountability at their core.

To address these challenges, a mixed-methods approach that combines computational analysis with qualitative human evaluation is necessary. While NLP and machine learning offer powerful tools for identifying patterns in scientific translation, they must be supplemented with interpretive methodologies, including discourse analysis, expert panel reviews, and ethnographic studies of public engagement with science. By integrating these approaches, researchers can develop a more holistic understanding of how scientific knowledge is translated, ensuring that AI serves as a tool for enhancing rather than distorting public understanding.

This paper has thus far explored the theoretical and methodological considerations surrounding AI-driven scientific translation, drawing on insights from Harari, Gebru, and Crawford to highlight both the challenges and opportunities inherent in these systems. AI has emerged as a crucial tool for bridging linguistic gaps in scientific communication, but its integration into the translation process introduces concerns regarding accuracy, bias, and ethical responsibility. As AI models continue to evolve, it is essential to examine how these systems navigate complex linguistic structures, adapt to domain-specific terminology, and address the sociopolitical implications of automated translation. By critically engaging with these issues, we can work toward a more equitable and informed scientific discourse that leverages AI's potential while mitigating its inherent limitations.

As we consider the broader implications of AI-driven scientific translation, it is crucial to examine how interdisciplinary research can inform better translation strategies. Scientific knowledge is often deeply contextual, requiring expertise from multiple fields to ensure accurate communication across languages and cultural boundaries. Fields such as linguistics, cognitive science, and information theory offer valuable insights into how humans process and interpret scientific information, providing essential guidance for refining AI-based translation models.

By integrating AI methodologies with human-centered research approaches, we can develop hybrid systems that combine computational efficiency with the nuanced judgment required for accurate knowledge dissemination. For instance, discourse analysis techniques from linguistics can help AI models recognize context-specific meaning shifts, reducing the likelihood of misinterpretation and ensuring that translations reflect the intended meaning of complex scientific texts. Cognitive science can further inform AI-driven translation by shedding light on how individuals perceive and internalize translated content, helping developers design models that enhance clarity without oversimplifying technical details. Similarly, advances in information theory can improve AI's ability to encode and retrieve scientific knowledge, enhancing both the accuracy and efficiency of translation processes.

Beyond interdisciplinary research, an ethical framework for AI-driven scientific translation must be established to safeguard against potential harms. The deployment of AI in translation is not a neutral process, it reflects the biases and priorities embedded within the training data and underlying algorithms. To address these concerns, transparency measures must be implemented, requiring AI developers to disclose how translation models are trained, the datasets they rely on, and any known biases that may affect their outputs. Without transparency, the scientific community risks unknowingly relying on AI-generated translations that reinforce systematic distortions or exclude critical perspectives.

Moreover, regulatory oversight may be necessary to prevent AI-generated translation metrics from being misused for censorship or manipulation. Harari warns that allowing AI to become the ultimate arbitrator of truth without accountability could undermine public trust in science and lead to increased polarization (Harari, 2022). If AI models disproportionately favor certain scientific narratives while excluding others, they could exacerbate existing disparities in knowledge access and create echo chambers that hinder intellectual progress. Thus, any AI system used for scientific translation must operate within a broader ethical and regulatory framework that prioritizes fairness, accountability, and human oversight.

The challenges of AI-driven scientific translation extend beyond academic discourse to the broader information ecosystem, including education, policy-making, and public engagement with science. In educational settings, AI-powered tools have the potential to help students and educators access scientific knowledge in more digestible formats, providing translations that adapt to different learning styles and linguistic backgrounds. However, without careful oversight, these tools could reinforce pre-existing biases, privileging certain interpretations over others or oversimplifying complex scientific concepts.

The reliance on AI-generated translations in educational materials must therefore be critically examined to ensure that students receive accurate and nuanced information rather than reductive summaries that fail to capture the depth of scientific discourse. Similarly, policymakers who depend on AI-generated summaries of scientific research must be cautious about potential distortions that could influence decision-making processes. If AI models prioritize certain types of research over others due to dataset imbalances or algorithmic biases, they could inadvertently shape policy discussions in ways that do not fully reflect the diversity of scientific thought. Public engagement with science also stands to be affected by AI-driven translation, as automated tools may determine how scientific findings are presented to non-expert audiences. Ensuring that AI-generated translations remain accurate, balanced, and contextually appropriate is essential for maintaining public trust in scientific communication.

The role of AI in scientific translation is thus a double-edged sword: while it has the potential to enhance access to knowledge, it also carries risks of misrepresentation, bias, and ethical concerns. As researchers continue to develop AI-driven translation frameworks, it is imperative to engage in interdisciplinary collaboration to ensure that these systems serve the public good. Crawford urges us to recognize that AI is not merely a technological tool but a socio-political construct shaped by human choices and values (Crawford, 2016). AI models are trained on data selected by human developers, meaning that the decisions made during data curation, model training, and system deployment influence the outputs these models generate. Without intentional safeguards, AI-driven translation systems could amplify structural inequalities in knowledge production, privileging dominant languages, scientific paradigms, or geopolitical interests. Addressing these challenges requires a commitment to interdisciplinary dialogue, bringing together AI researchers, linguists, ethicists, and domain experts to collaboratively design translation systems that align with scientific integrity and ethical responsibility.

AI-driven scientific translation presents both opportunities and challenges. While AI offers powerful tools for analyzing and tracking information shifts across languages and scientific disciplines, it remains limited in its ability to fully comprehend linguistic nuance and context. To address these limitations, hybrid models that integrate AI methodologies with human expertise must be developed, ensuring that automated translations are reviewed and refined by domain specialists. The A/A Matrix provides a structured approach to evaluating translation accuracy while ensuring that scientific integrity is maintained, but it must be accompanied by ongoing quality assessment processes that incorporate feedback from researchers and practitioners. Ethical considerations must also guide the deployment of AI in knowledge translation, ensuring that these systems promote transparency, accountability, and equitable access to scientific information. As AI continues to shape the landscape of scientific communication, it is essential to remain vigilant about its limitations and ethical implications, resisting the temptation to treat AI-generated translations as inherently objective or infallible. By critically engaging with the promises and pitfalls of AI-driven translation, we can work toward a more informed and scientifically literate society, one that benefits from both technological advancements and rigorous human oversight.

References

Bender, E. M., & Gebru, T. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT)*, 610–623. https://doi.org/10.1145/3442188.3445922

Crawford, K. (2016). *Artificial intelligence's white guy problem*. The New York Times. Retrieved from https://www.nytimes.com

Harari, Y. N. (2022). *21 Lessons for the 21st Century*. Spiegel & Grau.