

The rapid growth of Artificial Intelligence (AI) has resulted in an overwhelming volume of research papers, technical reports, blog posts, and other AI-related documents being published across the globe. While the majority of scientific AI papers are written in English, a substantial amount of other content, such as industrial reports, blog posts, and policy documents, is produced in various other languages. Therefore, developing machine translation methods for AI documents is a critical and timely need, not only to break down language barriers but also to ensure equitable access to cutting edge research and technological advancements. However, traditional translation tools are typically not designed to handle domain-specific, multimodal data in AI documents. They also perform poorly in low-resource languages with limited training data. This creates significant obstacles to global collaboration and innovation in AI across linguistic and cultural boundaries.

The LEXICAI project aims to develop a new generation of translation tools specifically tailored for AI documents. Our system is designed to understand cross-modal context, linking information from text, figures, equations, tables, and code snippets, to improve translation accuracy. It also integrates human-in-the-loop feedback from domain experts and linguists to ensure readability, correctness, and terminology consistency. To make the system widely accessible, we focus on developing lightweight models that can operate in resource-constrained environments. Our approach supports the translation of both major languages and underrepresented ones to ensure equitable access to AI knowledge and promote linguistic diversity in global research. We will contribute a large-scale AI terminology database to support consistent and reliable translation across languages. LEXICAI will break down linguistic and technical barriers in AI communication, enabling researchers, innovators, and broader public audiences to access, understand, and contribute to cutting-edge AI developments, fostering a more connected and equitable AI ecosystem. Our project will advance machine translation for complex AI content while providing open-source translation tools and a terminology database to support a more connected, multilingual, and inclusive AI future.