

OSCAIL - Open Science Communication through AI in EU Languages

Sheila Castilho,¹ Susanna Fiorini,² Lynne Bowker,³ Petr Motlicek,⁴ Joss Moorkens,¹
Lieve Macken,⁵ Dairazalia Sanchez-Cortes,⁴ Janne Pölönen,⁶ Sami Syrjämäki,⁶
Mikael Laakso,⁶ Mark Fishel,⁷ Anastasia Stasenko⁸

¹Dublin City University, ²OPERAS, ³Université Laval, ⁴Idiap Research Institute,
⁵Ghent University, ⁶Federation of Finnish Learned Societies, ⁷University of Tartu, ⁸pleias

Abstract

The Anglocentric nature of scholarly communication has many implications, such as limiting publication, discoverability and access from other language communities (even for major languages); putting minoritized languages at risk in the academic domain; and excluding many from peer review. The OSCAIL project addresses these challenges by exploring how machine translation (MT) enhanced by large language model (LLM)-based technologies can support access to scientific knowledge. Outputs will include evaluation datasets, protocols and best practices for MT in scholarly communication, and a prototype integration of MT tools into Open Journal Systems, the world's most widely used open-source scholarly publishing platform.

1 Introduction

The dominance of English in scholarly communication creates structural barriers to both participation in research and access to scientific knowledge. Researchers working in other languages often face limitations in publishing and disseminating their work internationally (Amano et al., 2023), while important research produced in languages other than English may remain less visible or accessible (Hannah et al., 2024). This linguistic imbalance also affects peer review processes and the discoverability of scientific outputs, and limits the accessibility of research for non-specialist audiences.

MT is already widely used informally by researchers, reviewers, and readers to access scientific content across languages (Fiorini, 2024). However, the effective and responsible integration of MT into scholarly communication workflows remains insufficiently explored (Ayeni et al., 2025). Previous studies and initiatives have shown that MT can help users understand the general meaning of scientific texts, but challenges remain regarding terminology, domain-specific accuracy, and the preservation of meaning in specialised contexts (Macken et al., 2024). In addition, ethical considerations related to transparency, governance, and the responsible use of artificial intelligence (AI) technologies (Moorkens et al., 2024) remain central to the adoption of automated translation tools in scholarly publishing.

The OSCAIL project (Open Science Communication through AI in EU Languages) aims to address these challenges by exploring how advances in MT and LLMs can support more inclusive multilingual communication in science. The project is funded under the CHIST-ERA ERA-NET Call 2025 “Science in Your Own Language”, with the following funding agencies: Research Ireland, Swiss National Science Foundation (SNSF), The Academy of Finland (AKA), French National Research Agency (ANR), Estonian Research Council (ETAg). OSCAIL will run for 36 months (1 May 2026–30 April 2029) and will bring together an interdisciplinary and multilingual consortium including:

- Dublin City University (Ireland, coordinator)
- OPERAS Research Infrastructure (France)
- Idiap Research Institute (Switzerland)
- Pleias (France)
- University of Ghent (Belgium)
- Université Laval (Canada)

- University of Tartu (Estonia)
- Federation of Finnish Learned Societies (Finland)

The consortium combines expertise in MT, natural language processing, LLM fine-tuning, translation evaluation, scholarly publishing infrastructures, and research policy.

Central to the project is the investigation of MT in three practical use cases inspired by real scholarly communication workflows. First, the project will explore the use of MT to support multilingual peer review, enabling reviewers to engage with manuscripts and author responses in their preferred working language while allowing authors to write and respond in the language with which they are most comfortable. Second, OSCAIL will investigate how MT can improve the discoverability of scientific publications, for example through the translation of metadata and key descriptive elements to support multilingual indexing and search. Third, the project will examine how MT and automatic summarisation can improve the accessibility of research outputs, including the translation and plain-language summarisation of scientific articles for broader audiences.

To support these goals, OSCAIL will gather domain-specific datasets (still under discussion) in project languages representing different levels of technological support according to the Digital Language Equality framework. These data will be used to develop, adapt, and evaluate MT systems, including approaches based on large language models, for the translation of scientific texts.

2 OSCAIL Objectives

OSCAIL pursues five main objectives:

Advance MT for scientific content by developing and benchmarking MT systems adapted to the linguistic and structural characteristics of scholarly texts.

- Support high-, mid-, and low-resource European languages through domain-adapted models and evaluation across diverse language pairs.
- Enable multilingual peer review workflows allowing authors and reviewers to work in their preferred languages without compromising the rigour of the review process.
- Develop evidence-based evaluation frameworks combining automatic and human evaluation tailored to different user roles in scholarly communication.
- Promote responsible and inclusive scholarly AI,

addressing ethical, legal, and governance considerations in the deployment of AI-based translation technologies.

3 Expected Results

As the project has recently been funded and just started, results are expected during the implementation phase. Planned outcomes include the creation of multilingual evaluation datasets and corpora for scientific MT, reproducible evaluation protocols and best-practice guidelines for MT deployment in scholarly communication, comparative analyses of MT systems across different scholarly use cases, and a prototype integration of MT functionalities into Open Journal Systems. The project will also produce ethical guidelines and recommendations to support responsible AI-driven translation in scholarly publishing and to promote more inclusive and multilingual access to scientific knowledge.

References

- Amano, Tatsuya, Valeria Ramírez-Castañeda, Violeta Berdejo-Espinola, Israel Borokini, Shawan Chowdhury, Marina Golivets, Juan David González-Trujillo, Flavia Montañó-Centellas, Kumar Paudel, Rachel Louise White, and Diogo Veríssimo. 2023. The manifold costs of being a non-native english speaker in science. *PLOS Biology*, 21(7):e3002184.
- Ayeni, Philips, Emanuel Kulczycki, and Lynne Bowker. 2025. Machine translation and scholarly publishing: A scoping review. *Canadian Journal of Information and Library Science*, 48(1):123–145.
- Fiorini, Susanna. 2024. Exploratory studies for the creation of a technology-aided collaborative translation service in open scholarly communication.
- Hannah, Kelsey, Neal R. Haddaway, Richard A. Fuller, and Tatsuya Amano. 2024. Language inclusion in ecological systematic reviews and maps: Barriers and perspectives. *Research Synthesis Methods*, 15(3):466–482.
- Macken, Lieve, Vanessa De Wilde, and Arda Tezcan. 2024. Machine translation for open scholarly communication: Examining the relationship between translation quality and reading effort. *Information*, 15(8):427.
- Moorkens, Joss, Sheila Castilho, Federico Gaspari, Antonio Toral, and Maja Popović. 2024. Proposal for a triple bottom line for translation automation and sustainability. *The Journal of Specialised Translation*, (41):2–25.