



AIDESL

Automated data extraction for faster healthcare innovation

To speed up systematic literature reviews (SLRs) while improving their accuracy, the ITEA project AIDESL (Artificial Intelligence Data Extraction of Scientific Literature) will use artificial intelligence (AI) to automate data extraction from scientific literature, particularly in the healthcare domain.

Addressing the challenge

An SLR is a comprehensive, structured approach to gathering, screening, assessing, analysing and synthesising research literature. This is fundamental to identifying and evaluating clinical evidence to support the development and safe use of medical devices and pharmaceuticals and in informing health policy guidelines and academic research. However, traditional SLRs are complex, time-consuming and operationally intensive, resulting in long completion times, resource challenges and significant error rates – all compounded by a rapid growth in scientific publications over recent years. This impedes the cost-effective availability of new treatments and public access to timely medical advice from trusted health authorities.

Proposed solutions

Fundamentally, AIDESL seeks to fully automate data extraction from text, charts, tables and images in scientific literature using a no code/low code platform with a focus on accessibility and user-friendliness. This will be achieved by establishing comparative performance benchmarks for multiple generic and domain-specific large language models (LLMs) for SLRs and by testing a range of datasets, including publicly available SLRs and anonymised contributor datasets for SLRs representing different medical device product categories and therapeutic fields. This will allow academic and commercial research to identify bias risks, establish contextual longitudinal assessments of data in non-scientific or technical literature, and

identify and characterise commonalities and causal relationships across data sources. In addition to commercial software for data extraction, AIDESL will provide process benchmarking tools to the community as an open access starting point for standardised model evaluation, aiming to increase adoption

All are essential roles in the value chain for spurring cost-effective innovation, improving time to market and generating faster returns in an LLM market poised to grow from USD 40 billion in 2022 to USD 1.3 trillion over the next ten years. For users, the platform will offer benefits in terms of both efficiency and accuracy. The consortium therefore expects the results to reduce reviewer SLR completion times by as much as 80% over manual methods while improving LLM model performance and data/evidence quality by 7-25%. By focusing on user-



^ AIDESL will use AI to automate data extraction from scientific literature, particularly in the healthcare domain.

and expand the project's applicability to other forms of scientific and medical literature.

Projected results and impact

By leveraging AIDESL's results, researchers will be able to use LLMs to automatically detect adverse events that are longitudinal in scope, providing rapid patient safety data to physicians, regulatory bodies and manufacturers.

friendliness, AIDESL also aims to reduce the time to train intermediate staff on AI from three to five weeks to a matter of days. On a wider scale, the project anticipates a 25% reduction in the time taken for industry standard adoption, which helps enable AIDESL's most crucial long-term goal: improving the equitable development of new health innovations and knowledge for all.



Project start
October 2024

Project leader
Mike Gilbert, DistillerSR

Project website
<https://itea4.org/project/aidesl.html>

Project end
September 2027

Project email
michael.gilbert@distillersr.com



ITEA is the Eureka RD&I Cluster on software innovation, enabling a large international community of large industry, SMEs, start-ups, academia and customer organisations, to collaborate in funded projects that turn innovative ideas into new businesses, jobs, economic growth and benefits for society. ITEA is part of the Eureka Clusters Programme (ECP).

<https://itea4.org>