Balancing precision and value with artificial intelligence: Value judgements in a machine learning-assisted literature analysis

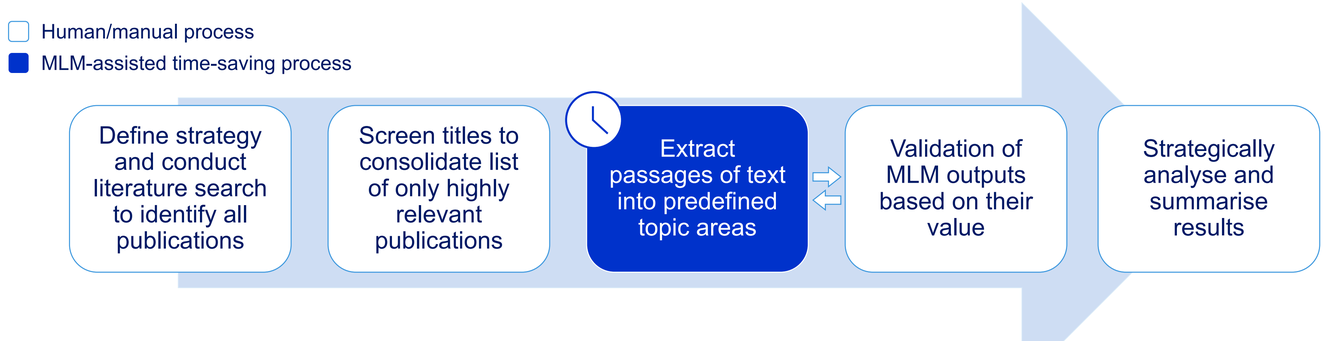
9

Kim Brown[a], Darren Fleetwood[b], Cat Hall-Jones[c], Katrina de Saram[c], and Gareth Morrell[b]

[a]Novo Nordisk Health Care AG, Zurich, Switzerland; [b]Madano, London, UK; [c]AXON Communications, London, UK

BACKGROUND & AIM

Figure 1: The process for MLM augmentation of scientific literature analyses.



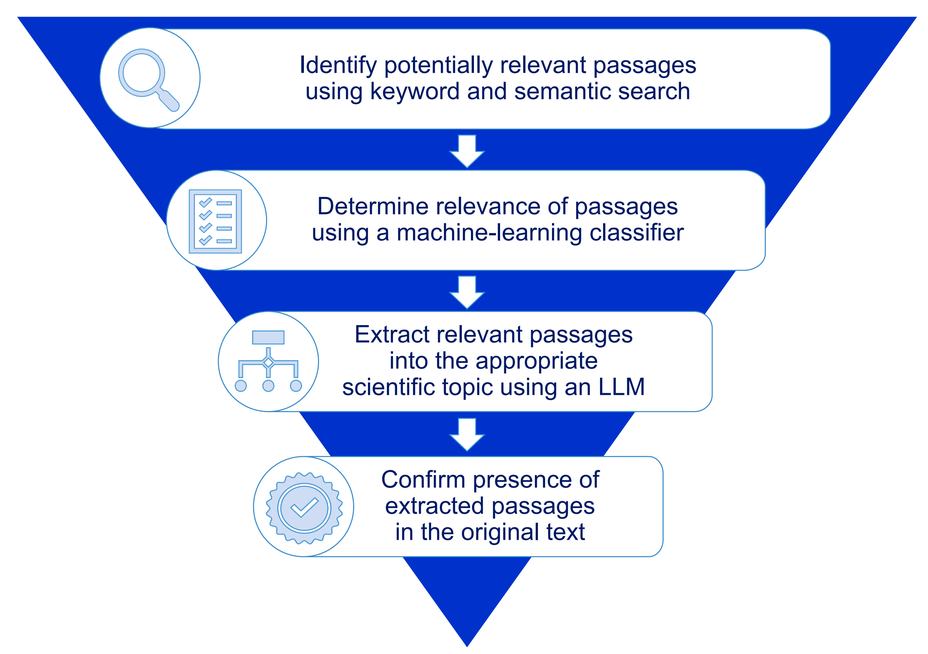
MLM, machine learning model.

- The ability of machine learning models (MLMs) to reduce manual workload by rapidly searching and extracting information from text has led to their application in scientific literature analyses.
- We developed and refined an MLM pipeline and assessed its usability to augment a literature analysis to support strategic publication planning in sickle cell disease **(Figure 1)**.

METHODS

- A literature search was performed to identify relevant publications (journal articles and congress abstracts) related to specific therapies in sickle cell disease between April 2022 and August 2023.
- An MLM pipeline was developed to identify and extract relevant passages of text from these publications into eight predefined topic areas: burden of disease, product descriptor, mechanism of action, efficacy, haemolysis, vaso-occlusive crises, safety, unmet needs.
- The pipeline integrated multiple methods, including a large language model (LLM), GPT-4-0613, to extract passages of text and verify them against the source publications to eliminate hallucinations **(Figure 2)**.
- The model was developed using one manually compiled data set (902 passages; 512 publications) and tested on a second data set (304 passages; 154 publications).
- MLM performance was assessed by recall, precision and F_1 score and compared with manual extraction by a medical writer.

Figure 2: The process of MLM data extraction.



LLM, large language model; MLM, machine learning model.

RESULTS

- For 154 publications, the MLM saved an estimated 90 hours that would have otherwise been required for a medical writer to extract data manually.
- Additional requests to extract data into new topic areas during the process would have resulted in greater time savings with the MLM vs manual extraction.
- On the test set, the MLM achieved high recall (98%) but moderate precision (72%), and the overall performance (F_1 score) was 83% **(Table 1).**

Table 1: Assessment of MLM performance.

Performance	Test set (n=304 passages)
Recall	98%
Precision	72%
F_1 score	83%

MLM, machine learning model.

Performance measures

- Precision = % of passages identified by the MLM that were relevant
- Recall = % of passages identified by the MLM among all relevant passages
- F_1 = overall performance

$$= 2 \times \frac{(\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})}$$

- Precision was limited by the MLM extracting passages that were relevant to the topic area but contained overly verbose or redundant information that lacked context.
- Approximately 20 hours were required for a medical writer to manually evaluate the most relevant passages based on their value and condense the verbose outputs before analysing and interpreting the results.

LIMITATIONS

- The evaluation of the MLM's performance was assessed retrospectively and was not predefined.
- The literature search was limited to one disease area.
- The latest LLM was used at the time (GPT-4-0613), but later models are typically more powerful and accurate.

CONCLUSIONS

- The MLM reliably identified and extracted relevant passages of text from publications and led to a substantial reduction in manual workload for data extraction. This gain in efficiency may allow for rapid re-analysis of new topic areas, providing greater flexibility in literature reviews.
- However, an effective literature review also requires subjective judgement and interpretation according to the known strategic priorities. As this analysis revealed, current MLMs struggle with the complexity and subjectivity of this task, requiring human input for value judgement.
- In summary, MLMs can result in substantial time saving for certain steps of a literature analysis; however, despite advancements in technology, they are unable to judge value and relevance sufficiently to produce meaningful outputs without human interpretation.

ACKNOWLEDGEMENTS
The MLM-assisted literature analysis and report was developed for and funded by Novo Nordisk Health Care AG, Zurich, Switzerland and performed by Madano and AXON Communications, London, UK.