Use of generative artificial intelligence to perform a publication gap analysis

Wing Yan Yeung[1], Brian Norman[2], Moamen Hammad[2], Eva Giannakouri[2], Patricia Badia Folgado[2], and Valerie Moss[2]
[1]Bayer Inc., Mississauga, Canada; [2]Prime, London, UK

No. 22

Explore research in more detail by scanning the QR code. A video describing this research can be accessed on the mobile-friendly website version of this poster.

Conclusions

- This hybrid human-AI approach provided an efficient and effective way of identifying phenotypic and genotypic differences in two very closely related diseases and identifying meaningful knowledge gaps in the literature.
- Generative artificial intelligence (genAI) was able to successfully perform a literature gap analysis with some human assistance.

Introduction

- Performing a comprehensive literature gap analysis is an important part of publication planning but can be a time-consuming process.
  - Traditional methods require extensive manual effort, designing search strategies, extracting relevant studies, summarizing findings, and identifying research gaps.

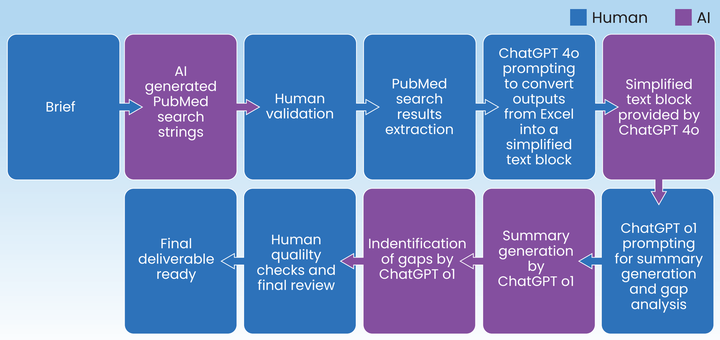
Objectives

- We identified an urgent educational need in the literature on understanding of genotypic and phenotypic differences between two closely related, albeit distinct, conditions in cardiovascular disease.
- An initial search identified >500 articles on the subject, which would be time-consuming to assess using a fully human approach.
- We explored the utility of genAI to determine the capability to efficiently and accurately perform a publication gap analysis, pull out insights and create an executive summary of the literature.

Methods

- An overview of the workflow is shown in **Figure 1**.

Figure 1. Workflow to create clinical training materials using genAI



AI, artificial intelligence

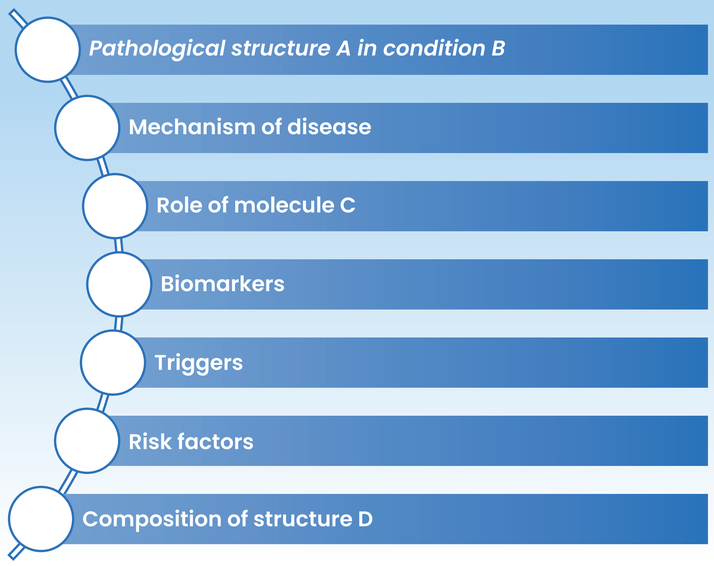
Acknowledgements: We would like to thank the Synapse Studio and Editor teams at Prime for their support with developing this poster. Generative AI (ChatGPT 4.0) was used for editorial support of the limitations section, which was further developed, reviewed, revised, edited, and updated by the authors.

Disclosures: Wing Yan Yeung is an employee of Bayer Inc. and owns shares of Bayer AG. Valerie Moss, Brian Norman, Moamen Hammad, Eva Giannakouri, and Patricia Badia Folgado are employees of Prime, London, UK. The research and interpretation provided here represent the views of the authors and not necessarily those of their employers.

Presented at the US Annual Meeting of the International Society for Medical Publications Professionals 2025, Washington DC, US, 11–14 May 2025

- We performed a PubMed literature search on eight specific topics (**Figure 2**) of predeveloped AI search strings.
  - The search was focused on articles published between 01/01/2018 and 09/19/2024.

Figure 2. Topics identified for the gap analysis

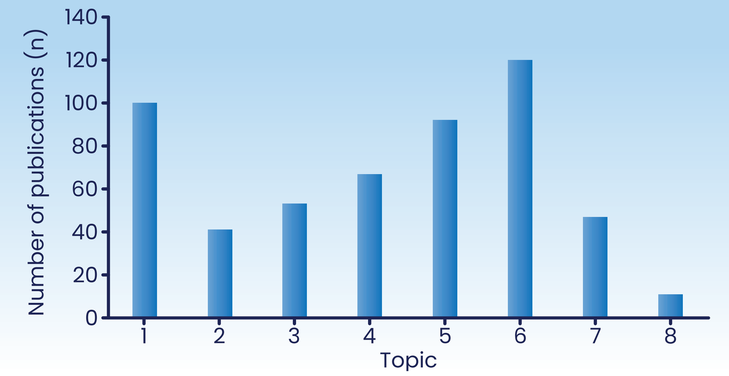


- Search results were exported into Excel.
- ChatGPT 4o was prompted to convert the outputs from Excel into a simplified text block per individual record, including title and abstract for each.
- We then initiated a new chat in ChatGPT o1, as its multi-step approach handles complex problems that need reasoning.
- As a final step, we prompted ChatGPT o1 to identify potential gaps in the literature provided.
- The literature summaries and gaps identified by ChatGPT o1 were assessed by a medical writer for quality and accuracy and were independently reviewed by a senior medical writer.
- The disparities in the volume of publications influenced the depth of analysis for each topic.

Results

- We observed variations in the number of publications across topics, ranging from 10 to 120 publications per topic (**Figure 3**).
- For topics with a high volume of publications, AI-assisted summarization and human review were needed for identifying key insights from large datasets.
- For topics with fewer publications, the gap analysis relied more on highlighting knowledge gaps and understanding why certain areas remained underexplored rather than extracting insights from a large dataset.

Figure 3. Number of publications per topic



AI, artificial intelligence; genAI, generative artificial intelligence.

- After spending some time refining the prompts, ChatGPT o1 generated the outputs quickly (less than a minute per prompt).
- The final output was approximately 30,000 words.
- This approach allowed humans to focus on higher-level interpretation, contextualization, and critical evaluation rather than manually sorting through large datasets.
- ChatGPT o1 successfully identified pertinent gaps in the literature.
  - In well-researched topics, where human-identified gaps have been previously discussed in systematic reviews and meta-analyses, AI-assisted gap analysis showed a high degree of concordance with known knowledge gaps.
  - For emerging or under-researched topics, where fewer publications were available, AI's role was more exploratory, identifying potential gaps based on patterns of missing data, underrepresented study designs, or inconsistencies across findings.
- Human oversight was essential as some AI hallucinations occurred, such as providing citations that did not correspond to the correct studies.
  - A medical writer cross-checked citations against the actual publications to confirm their validity, ensuring that no non-existent or misattributed studies were included.
  - In addition, humans provided contextual corrections, refining AI-generated summaries where wrong interpretations or clinical relevance was misrepresented.
- When the number of publications was high (>90) for a topic, ChatGPT o1 was unable to summarize all the content. As a result, we applied the same methodology to the remaining references and ChatGPT included the new information to the already generated summary.

Limitations

- The assumption of time spent per paper may not be accurate enough. Individual expertise, complexity of the papers, and level of analysis required may affect the actual time spent.
- The methodology primarily considers time spent, but other factors, such as accuracy and depth of analysis are harder to quantify and may not be fully captured in the comparison.