Leveraging GenAI for Literature Reviews

BACKGROUND

Scientific literature review is critical for most functional groups with Pharmaceutical and life science companies. Medical affairs organizations reply on scientific literature review for the following reasons:

- Evidence-based Decision Making: Using robust scientific evidence to make evidence-based recommendations and strategic planning.
- Regulatory Compliance: Presenting comprehensive evidence to regulatory agencies for approval.
- Safety Surveillance: Post-marketing surveillance is critical for understanding long-term side effects after a drug is launched.
- Identification of Knowledge Gaps: Identifying where data might be lacking for future research
- Maintaining Therapeutic Expertise: Staying abreast of latest development and ensuring MSLs can provide accurate and up-to-date information to KOLs and HCPs.
- Supporting Healthcare Professionals: Providing latest scientific information for a drug to HCPs so that the most recent and relevant data are at their fingertips.
- Competitive Intelligence: Obtaining insights into competitor's research to inform market change and adjust strategies.
- Scientific Communication: Ensuring communications with the medical community are accurate, relevant, and comprehensive.
- Scientific Exchange and Engagement: Helping meaningful scientific exchanges with KOLs and HCPs.
- Supports Value Proposition: Framing the value proposition of a drug regarding efficacy, safety, and positioning relative to other therapies. This is critical for payer discussions and market access strategies.

However, literature review currently is still a manual process – subject matter experts enter key words, surface relevant publications and then read those publication one at a time. To make matters worse, the body of scientific literature is growing rapidly. For example, there are 34 million publications in PubMed, with thousands of new ones added on a daily basis.

Leveraging generative AI for literature reviews can be transformative. AI can rapidly scan vast amounts of literature, identify and extract relevant data, and present it in a digestable way. Unfortunately,

current large language models such as GPT4 have severe limitations when applied to healthcare and life sciences domains and to literature review specifically:

1. LLMs can not process large amount of scientific data due to their token limitations.
2. Hallucination is a well known and common problem for all LLMs where wrong answers or insights are provided with high conviction.
3. Because of hallunication, transparency is required to verify generative output. LLMs routinely give wrong citations or even make up fake citations.
4. LLMs are trained on public data up until September 2021. Giving how fast the medical research is progressing, this is a serious limitation.

METHODOLOGY

We set out to address all the above limitations on our platform. Because of our early collaboration with OpenAI, we were able to train GPT models (both 3.5 and 4) with life science specific data; incorporate expert-in-the-loop for reinforced learning through human feedback (RLHF), and scale the LLMs to analyze the entire scientific literature corpus from multiple sources. More importantly, this "sum of all pieces" approach eliminates hallunications and provides citations for every piece of the generative content, providing much needed transparency.

VALIDATION AND RESULTS

We took a scientific approach to validate the AI-based literature review. For the same dataset (Pubmed and PMC) and same queries (16 queries total), we created the AI-based analysis and compared it side by side with the standard manual curation by experts. Then we analyzed the results for concordance including specificity (false positives) and sensitivity (false negatives). For those queries, we routinely saw the AI platform achieved 95% and above in sensitivity and specificity, outperforming experts for every query. Research has previously shown that human experts perform at 70% accuracy at best for manual scientific literature review. In addition, the speed of obtaining insights was shortened from days/weeks to minutes, not to mention surfacing critical insights that are routinely missed by manual curation.

CONCLUSION AND APPLICATION

- Generative AI can automate literature review from the entire corpus of scientific literature, even multiple data sources.
- Hallucination by LLMs can be eliminated by using a complex approach combining generative AI, NLP and human experts.
- Any types of generative summary has to have citations for creating transparency and trust for generative AI.
- Validation is possible and a must for any generative AI use cases.
- The time savings from days/weeks to minutes enables faster decision making.
- By automating a time consuming and error prone process, more comprehensive and highly accurate insights can be obtained to meet the objectives of medical affairs organizations (see background).

For more information visit Huma.ai.