Pharmacoevidence

Improving patient lives by supporting evidence base for healthcare interventions

Gen AI powered evidence generation: Implementing Advanced RAG architecture for sensitive data in HEOR applications

ISPOR 2025 | Montreal

Mohali India| London UK | Canada

Panelist and Session Outline

MODERATOR

Nicola Waddell

Pharmacoevidence

Introduction

Dr Vakaramoko Diaby

Otsuka

Retrieval Augmented Generation (RAG)

PANELISTS

Sven Klijn

Bristol Myers Squibb

Generative AI tools & RAG for HEOR

Prof Howard Thom

University of Bristol

HTA perspective on usage of Generative AI

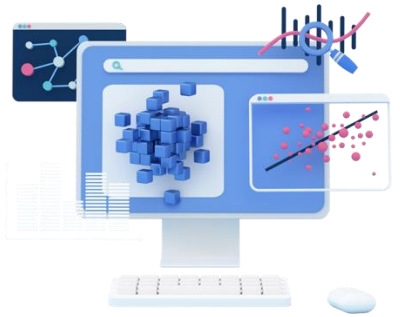
Increasing complexity & volume of data

Rising Complexity of Real-World Data

- Healthcare generates nearly 30% of the world's data[1]

Rapid Growth of Medical Literature

- The volume of healthcare data is projected to grow at a 36% Compound Annual Growth Rate (CAGR) by 2025 — faster than any other industry (Arcadia-2023) [1] [2]

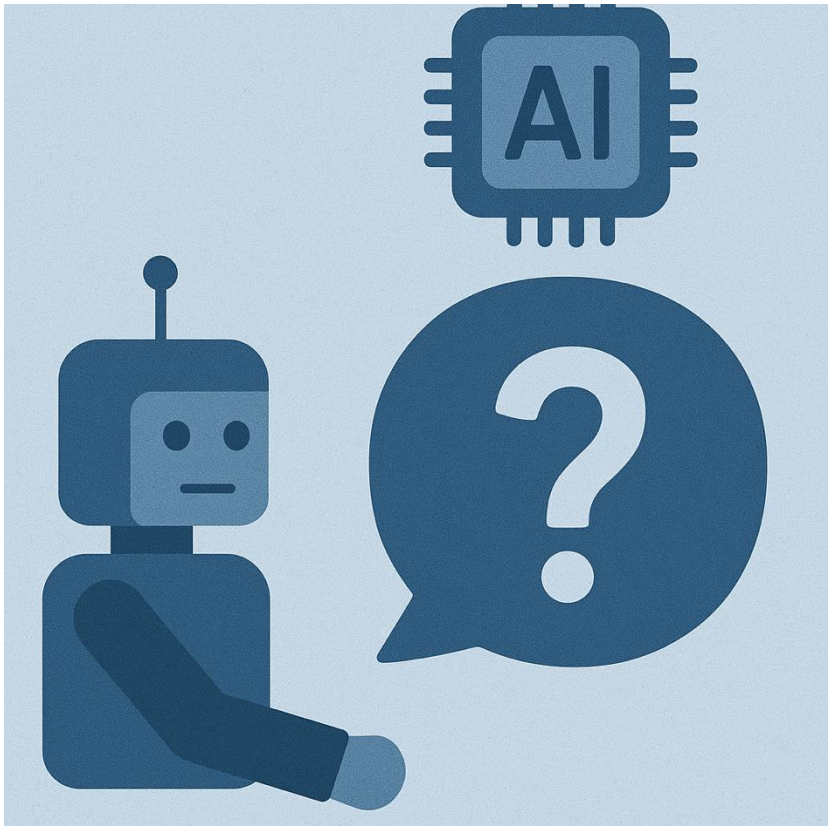


Secure & Reliable Evidence Generation using AI

- HTAs demand AI methods are used in a transparent, rigorous and trusted way (nice.org.uk) [3]

Ask the AI, But Ask Yourself First

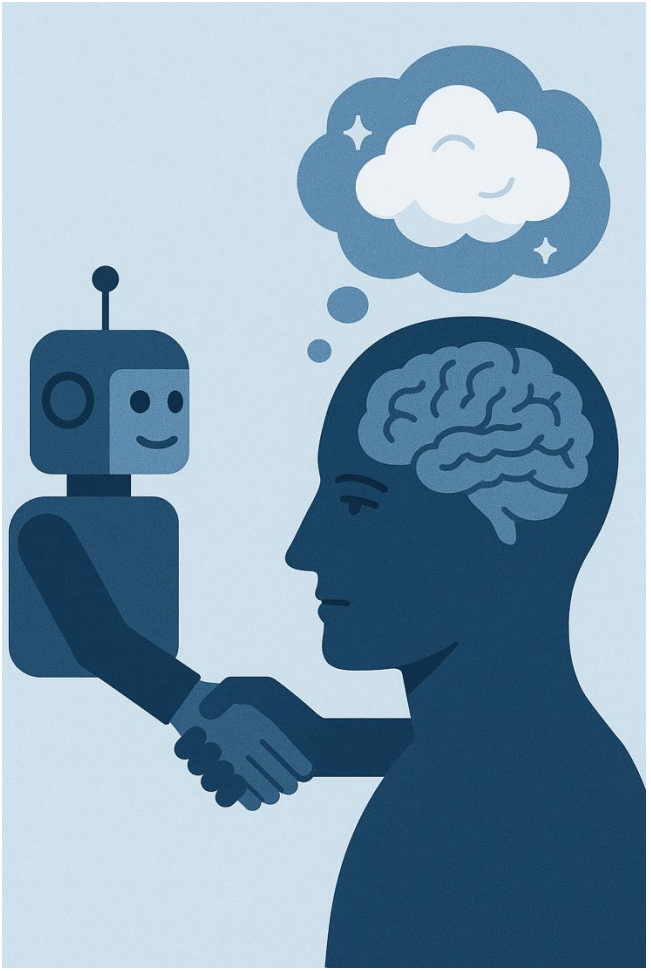
Do we really **need** Generative AI in HEOR, or is it just a buzzword?



Source: AI generated image

Ask the AI, But Ask Yourself First

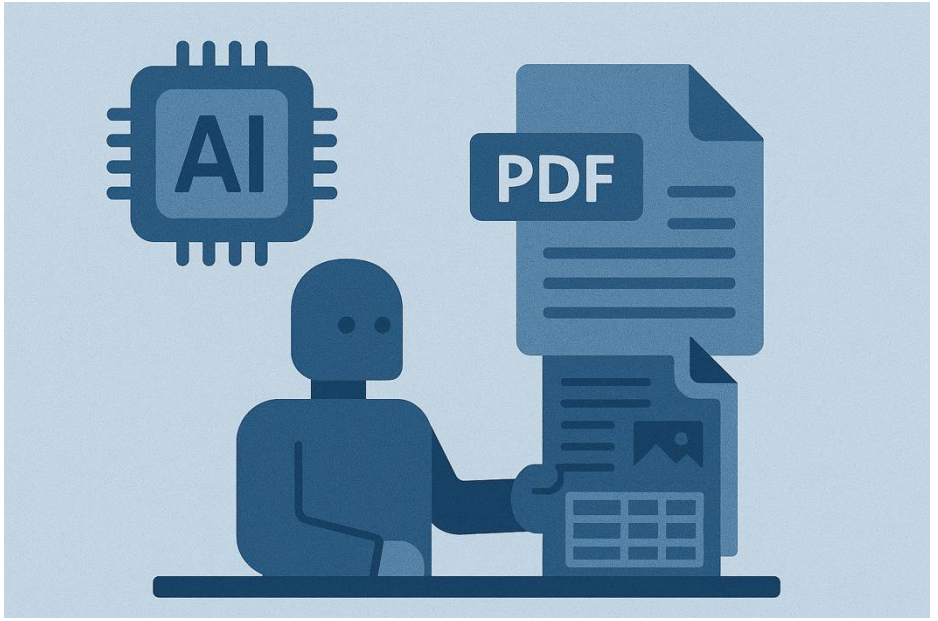
Can we **trust** the answers AI give us? Has your AI ever **made up** a publication title or quoted a journal that doesn't exist (**hallucinations**)?



Source: AI generated image

Ask the AI, But Ask Yourself First

Can Gen AI **handle** the **complex** real-world documents we deal with—PDFs, tables, scanned pages, even images?



Source: AI generated image

Ask the AI, But Ask Yourself First

Is my data **safe**? Can we use Gen AI without **risk** to sensitive information?

Source: AI generated image

www.ispor.org

ISPOR

Improving healthcare decisions

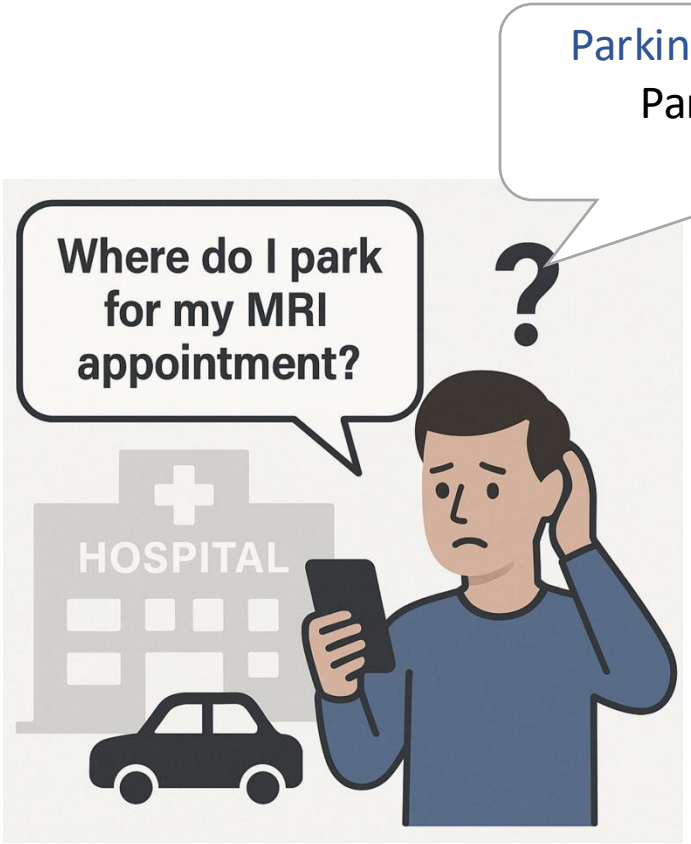
Retrieval-Augmented Generation (RAG): Making GenAI grounded, reliable, and actionable

Karam Diaby, PhD | ISPOR Montreal Panel 2025

Disclaimer

The opinions expressed on the following slides are solely those of the presenter and not those of Otsuka Pharmaceutical Development & Commercialization, Inc.

Patients attending MRI appointments often face challenges in navigating hospital parking systems



Option 1: Patient leverages an LLM to answer his question

Where do I park for my MRI appointment at Hospital X?

You can park on the east wing of hospital X, Parking lot A.

⚠ Note: This answer is generated based on general training data and could be prone to inaccuracies or hallucinations.

- High chance of hallucination – LLM not trained on the hospital's specific information.
- Information might not be traceable or verifiable.
- Potentially inaccurate or outdated, lacking explicit evidence references.

Option 2: Leverage the power of LLM on the internal, curated data, of the health system hospital X belongs to using RAG

Where do I park for my MRI appointment at hospital X?

Retrieval: Parking Docs

Grounded answer

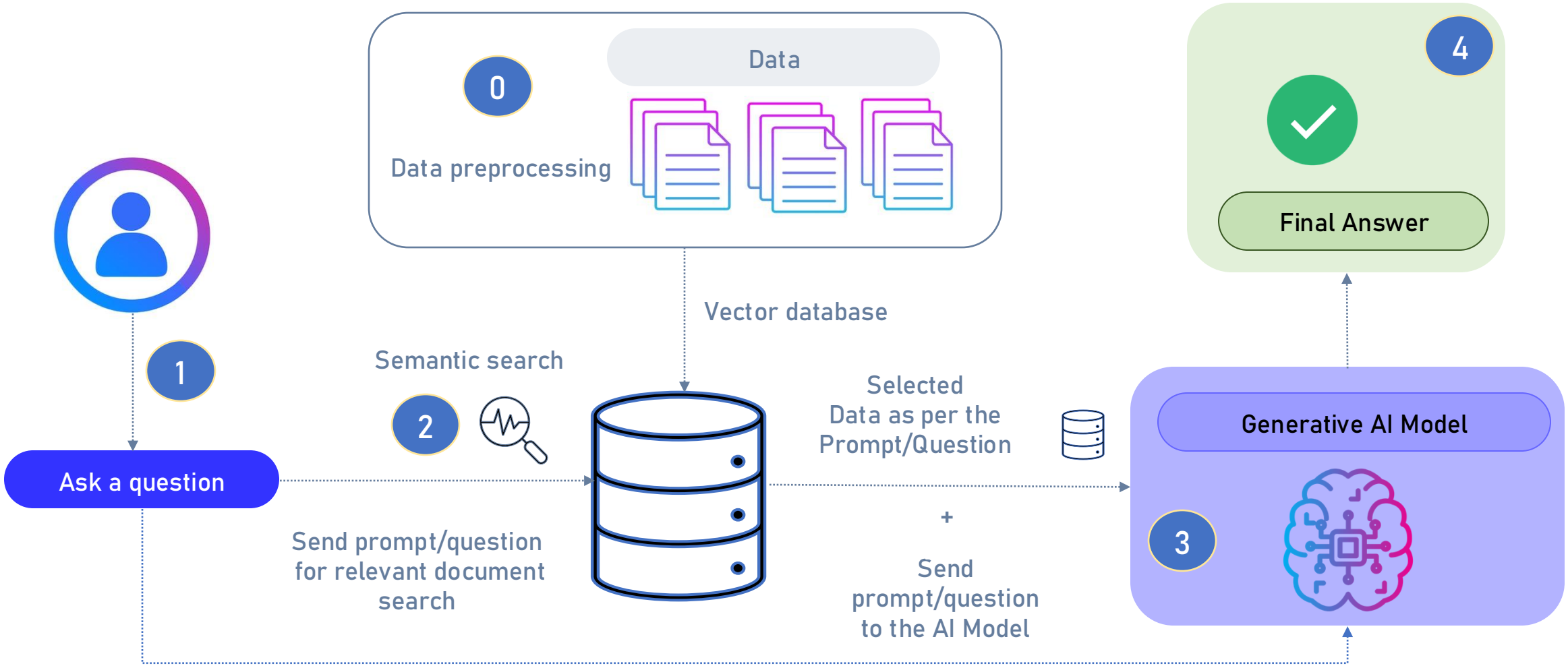
It is like prompting before you prompt

The generator (LLM) processes the fused context and generates a personalized response

"For your MRI appointment tomorrow at 10 AM, we recommend Parking Lot A, located at 1155 Centre street which is closest to the MRI facility. There are currently 20 available spots, and reserved parking for MRI patients is available. Please arrive 15 minutes early to ensure parking availability. Parking fees are $5 per hour, and validation is available at the MRI reception desk."

- ✓Accurate
- ✓Contextual
- ✓Trusted

RAG doesn't search with keywords, it searches with meaning



RAG ensures responsible use GenAI in HEOR

High bar for HTA evidentiary requirements to support	Transparency, reproducibility, accuracy
RAG handles different data types	Structured/unstructured data
Secure internal deployment with full audit trails	On-premise LLM deployment Service agreements with LLM providers
Traceable outputs compliant with regulatory/HTA standards	Linkage to source documents

RAG turns GenAI into a trusted tool, not just a good storyteller

Grounded AI that meets the rigor of HEOR standards.

Improves reproducibility of AI-generated evidence.

Supports regulatory, audit, and HTA transparency needs.

Amplifies expertise without replacing it.

Thank You!

RAG for Access, HEOR and HTA

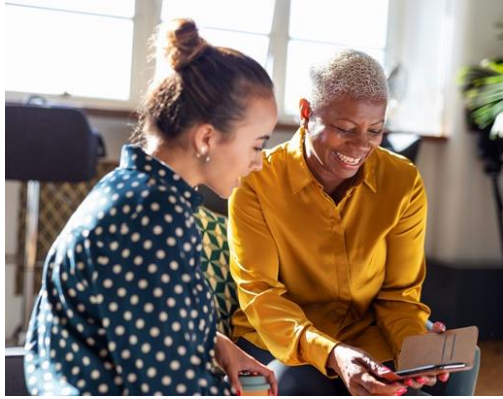
ISPOR Montreal

Sven Klijn

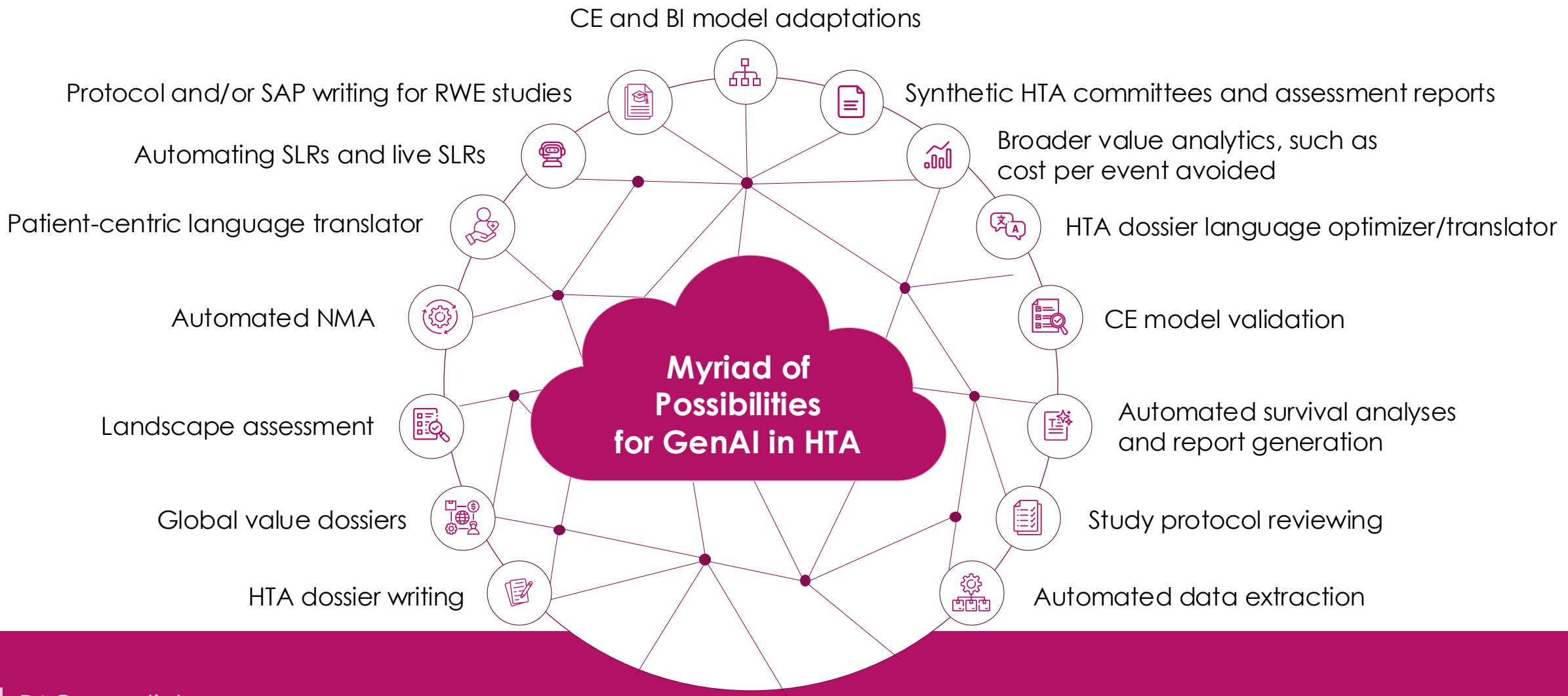
May 15th, 2025

Disclaimer

The opinions expressed in this presentation and on the following slides are solely those of the presenter and do not necessarily represent opinions of Bristol Myers Squibb.

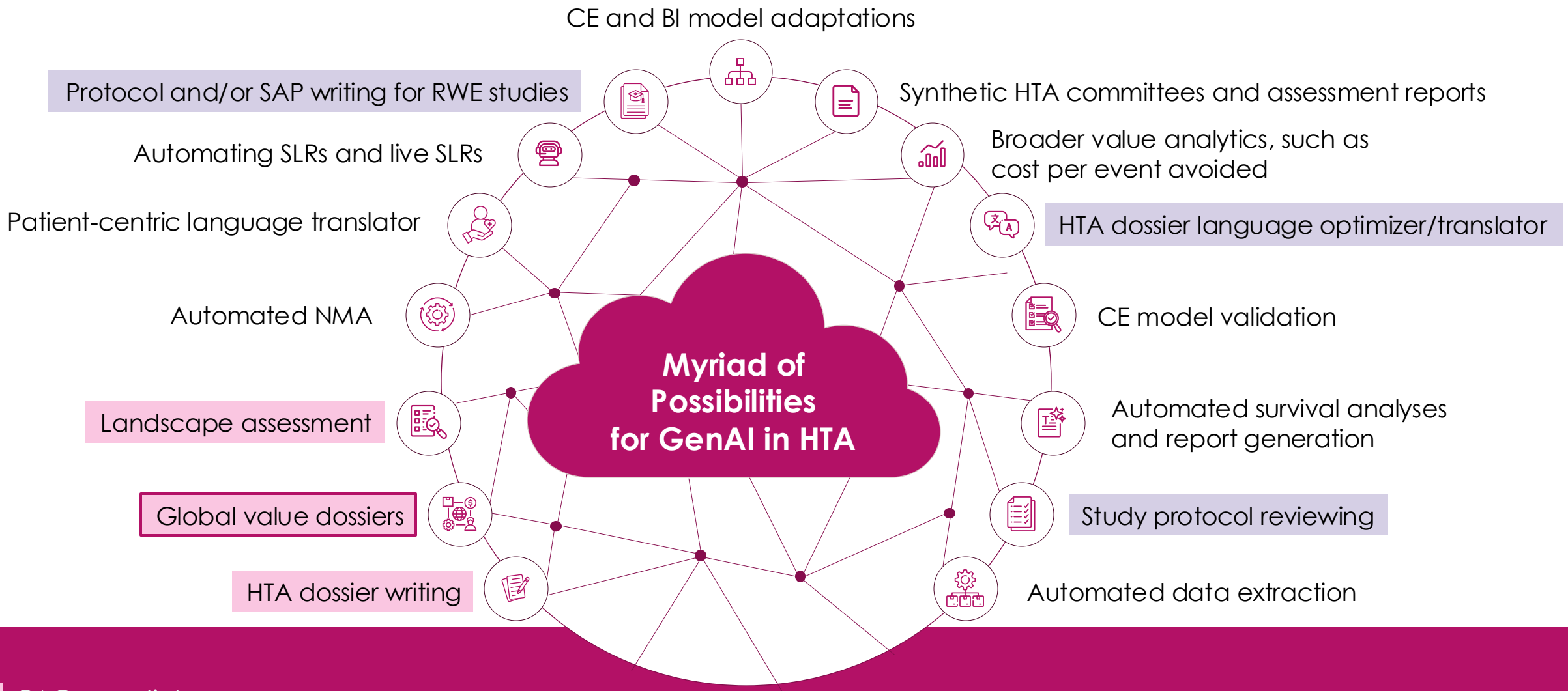





RAG essential

RAG of added value



RAG essential

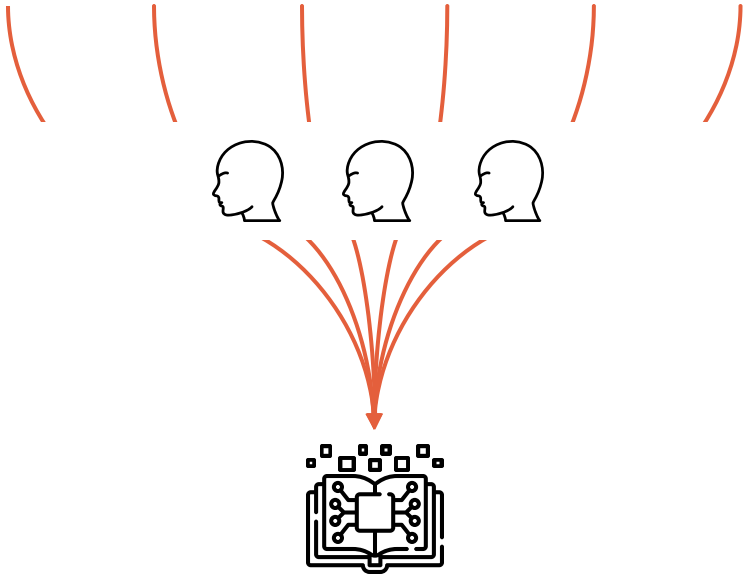
RAG of added value

Creation of value dossiers

Human only

Internal sources
Competitive intelligence
CSR
Health economic analyses

External sources
Treatment guidelines
Publications
Trial registries



Value Dossier

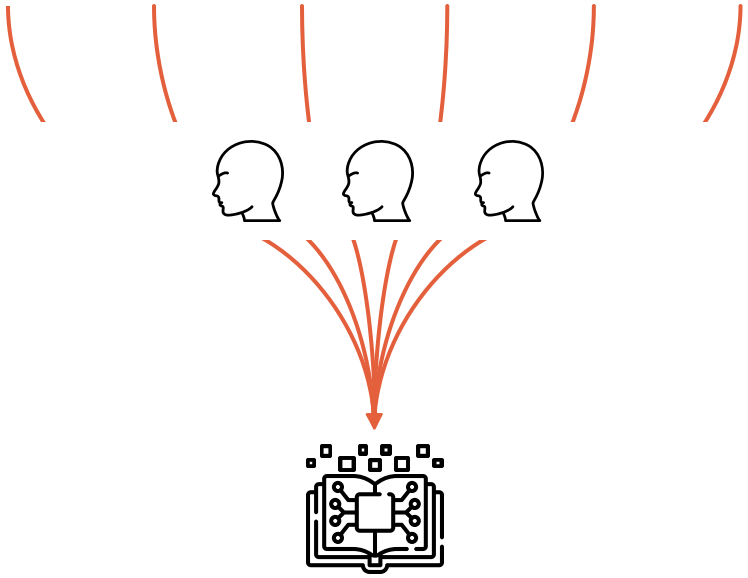
- Resource intensive process
- Challenge to keep up-to-date

Creation of value dossiers

Human only

Internal sources
Competitive intelligence
CSR
Health economic analyses

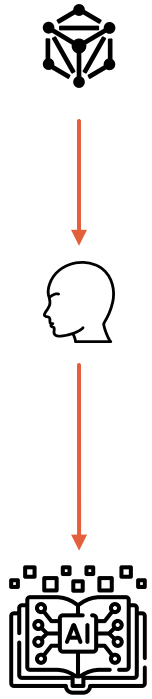
External sources
Treatment guidelines
Publications
Trial registries



Value Dossier

Human + GenAI

Large Language Model



Value Dossier

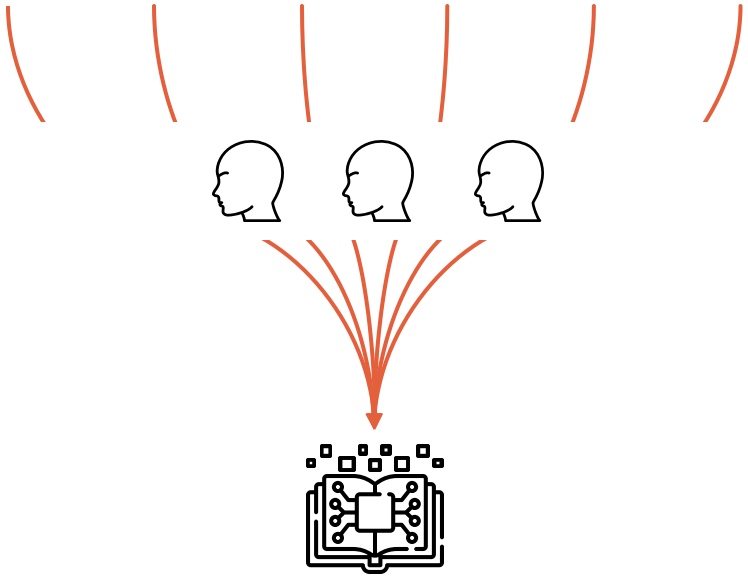
- Limited amount of domain-specific knowledge
- Information might be outdated
- Does not include proprietary information
- Prone to hallucinations

Creation of value dossiers

Human only

Internal sources
Competitive intelligence
CSR
Health economic analyses

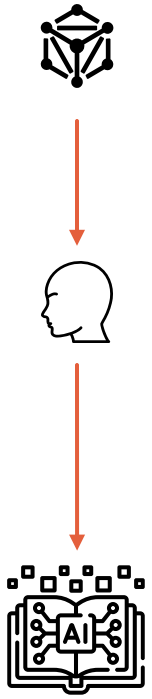
External sources
Treatment guidelines
Publications
Trial registries



Value Dossier

Human + GenAI

Large Language Model

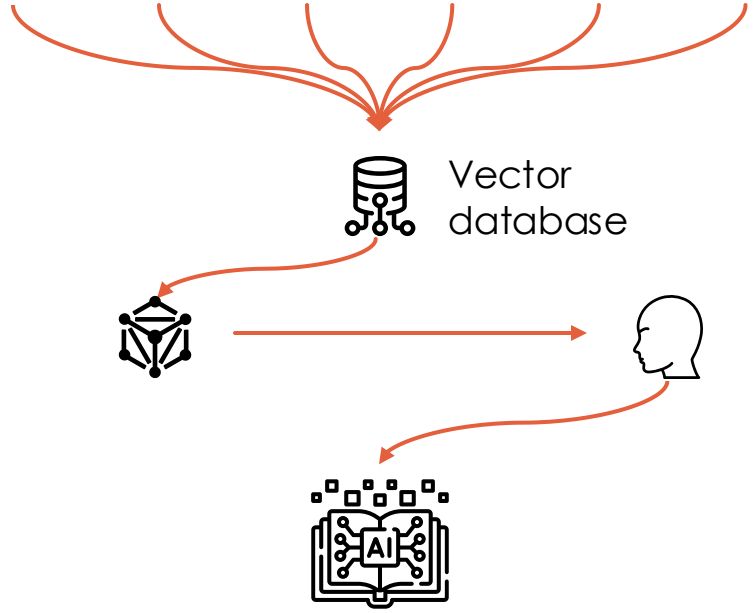


Value Dossier

Human + GenAI + RAG

Internal sources
Competitive intelligence
CSR
Health economic analyses

External sources
Treatment guidelines
Publications
Trial registries



Value Dossier

Limitations of RAG

High data quality remains essential

Design of the RAG solution impacts quality of the final result

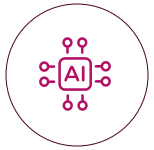
- Data ingestion: e.g., homogenization, chunking, embedding, database structure
- Retrieval: e.g., search method, re-ranking
- Integration: e.g., merging context with user prompt

Substantial hardware requirements

Navigation of intellectual property rights

Importance of RAG for Access, HEOR and HTA

Enables developing more robust evidence, faster

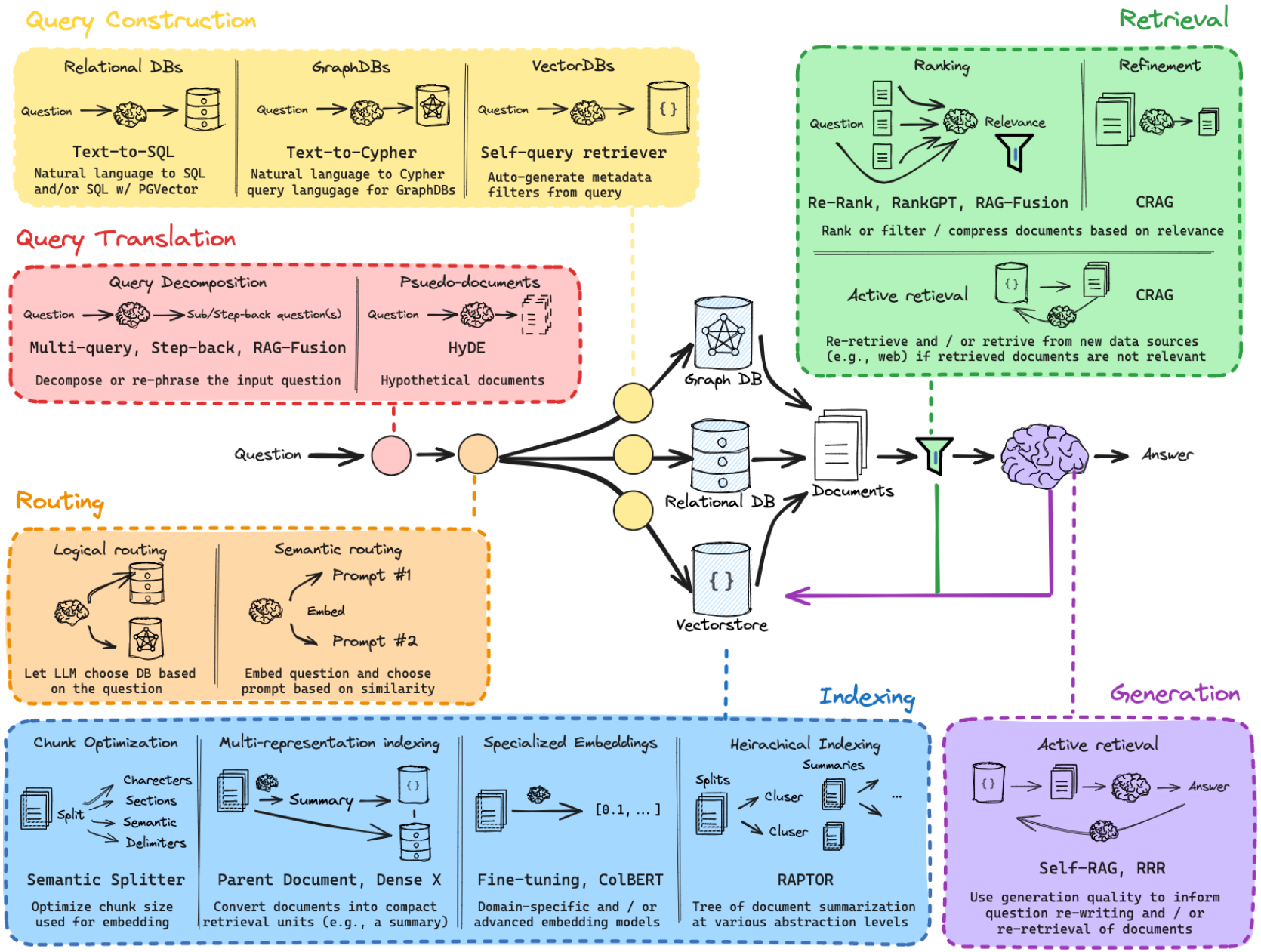


Enables use of GenAI in novel use cases

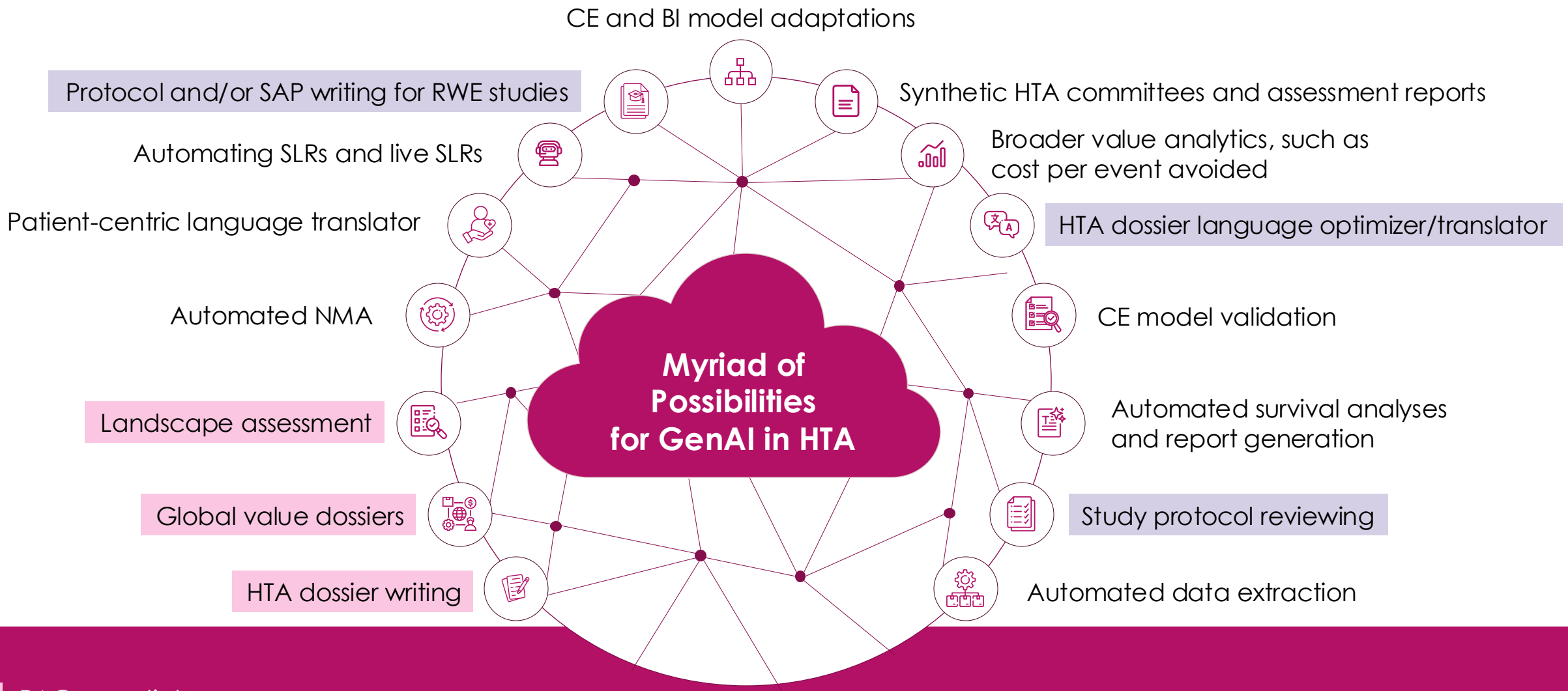
Can make existing GenAI use cases more robust

- Allows curation of data
- Transparency: traceability of sources
- Increases repeatability

A RAG is not a RAG is not a RAG

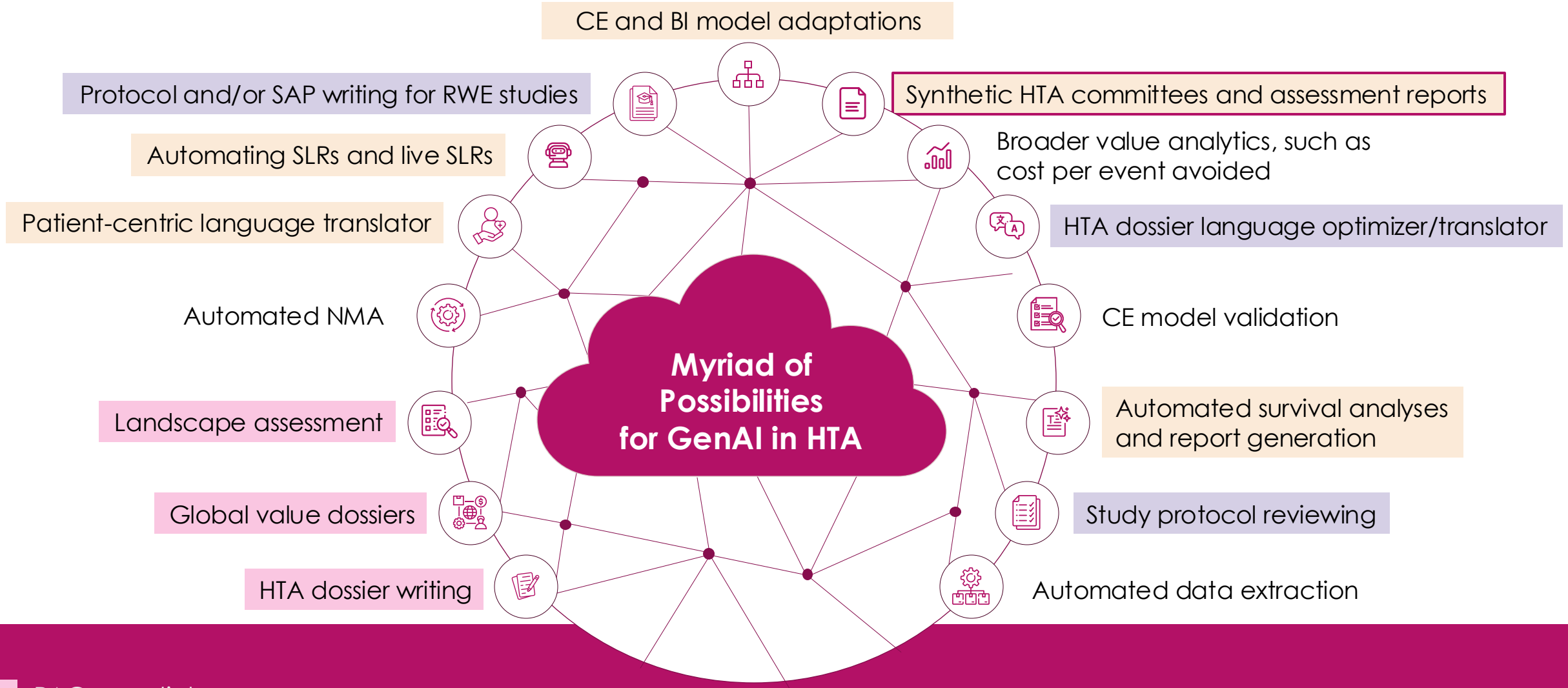


Source: Ababou TH (2025), MIT License, https://github.com/bragai/brag-langchain



RAG essential

RAG of added value

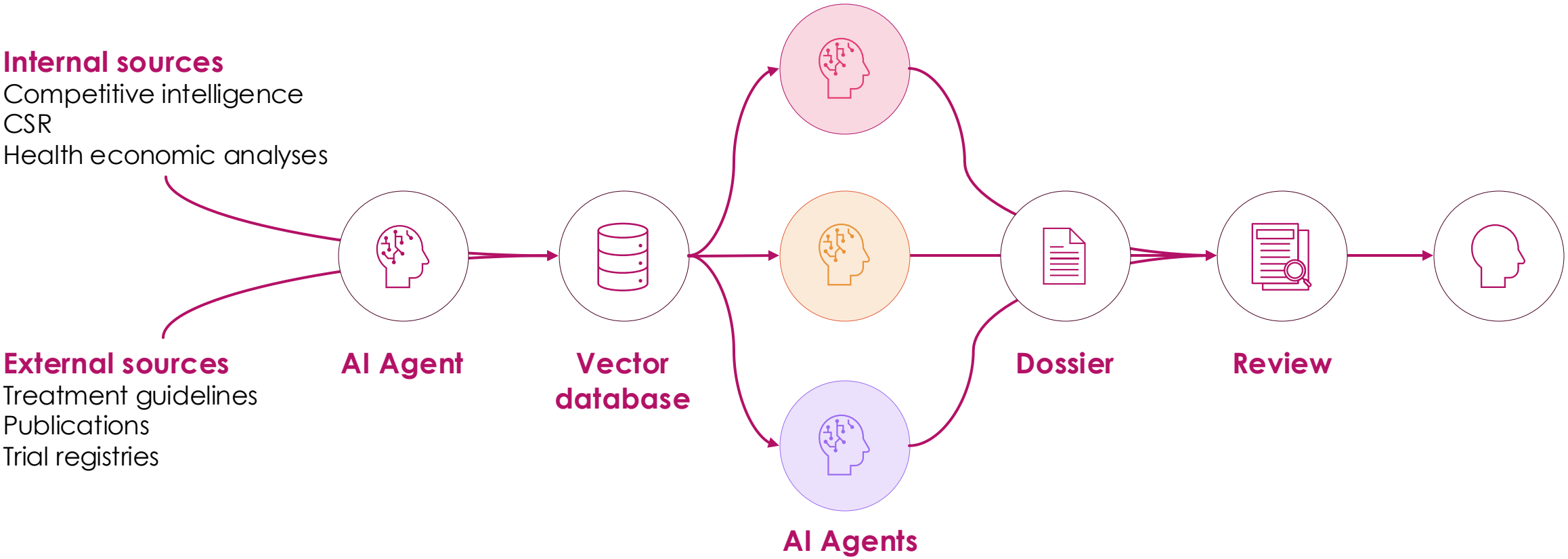


RAG essential

RAG of added value

RAG innovation

A look into the future? Combining RAG with agentic approaches



Thank you

University of BRISTOL

Gen AI powered evidence generation: Implementing Advanced RAG architecture for sensitive data in HEOR applications

15th May 2025

Howard Thom, University of Bristol

bristol.ac.uk

Disclaimer: The views presented are my own and do not represent the position of either the Bristol NICE Technology Assessment Group (TAG) or NICE in general.

NICE AI Position Statement

- NICE recognizes that artificial intelligence (AI), including generative AI will increasingly inform evidence considered in health technology assessments (HTAs).

- ‘The use of artificial intelligence (AI) methods, from relatively well-established machine learning approaches to newer and more complex generative AI, offers several potential benefits for this process’.

- The use of AI methods must be transparent, explainable, and ethically sound and used only when suitable and it’s use adds value.

- AI should augment human decision-making, not replace it.

Link: https://www.nice.org.uk/about/what-we-do/our-research-work/use-of-ai-in-evidence-generation--nice-position-statement

NICE AI Position Statement

NICE Guidance on Responsible AI Use in HTA Evidence Generation

Systematic Reviews & Evidence Synthesis

Automating study screening, classification, automate data extraction, and to generate the code required to synthesize extracted data for meta-analysis support

Clinical evidence

Trial design, summarization, synthetic data and generate external control arms, predictions and reducing bias

Real-world data and analysis

Natural language processing (NLP) for to generate structured data from unstructured real-world data, efficient selection of relevant populations and observations, support estimation of comparative treatment effects

Economic Modelling

To generate insights into cost drivers and health outcomes, automate the construction and calibration of new economic models, and generation of model reports, cross-validation of existing economic models

Link: https://www.nice.org.uk/about/what-we-do/our-research-work/use-of-ai-in-evidence-generation--nice-position-statement

NICE AI Position Statement

Do's

- Engage NICE early to align planned AI methods with guidance.
- Apply AI only when it adds demonstrable value.
- Keep human reviewers in the loop; AI assists, never replaces.
- Disclose algorithms, data sources and validation tools.
- Perform and document validation, bias and security tests.
- Comply with UK AI rules, licenses and IP requirements.

Don'ts

- Hide risks, biases or model limitations.
- Provide black-box outputs without clear explanation.
- Selectively report or manipulate data to inflate performance.
- Skip periodic monitoring and re-validation.
- Ignore cybersecurity safeguards against attacks.
- Share sensitive data in ways that violate privacy or consent requirements

Link: https://www.nice.org.uk/about/what-we-do/our-research-work/use-of-ai-in-evidence-generation--nice-position-statement

CDA AI Position Statement

NICE's position statement was adapted, with minor refinements to align with the organization's context and Canada's HTA environment. Artificial intelligence technologies have the potential to revolutionize health care systems. Any potential benefits of using AI methods must be balanced against anticipated and/or known risks

Two Key Sections

Section 1: Potential Uses of AI Methods in HTA

- Systematic Review and Evidence Synthesis
- Clinical Evidence (Trial design and optimization, Identification of patient cohorts using EHRs (via NLP), Generating synthetic control arms)
- Real-World Data and Analysis
- Cost-Effectiveness Evidence

Section 2: User Responsibilities When Using AI Methods in Evidence Generation and Reporting

- Engage early with CDA-AMC
- Align with Canada's Voluntary Code of Conduct (Canadian AIDA)
- User's responsibility to ensure compliance with licensing agreements, explainability of AI methods, increase transparency of their application

Link: Position_Statement_AI_Renumbered.pdf

AI for HTA – cross agency key takeaways

NICE (UK), CDA/CADTH (Canada) and EMA (EU) all recognize AI's use across the HTA evidence lifecycle

- Engage regulators early by seeking pre-submission advice on planned AI methods.
- Adopt AI only when it adds demonstrable value beyond established, interpretable approaches.
- Maintain transparency and explainability by fully disclosing models, data flows, validation and oversight.
- Keep a human in the loop so AI assists—but never replaces—human reviewers.
- Apply robust risk management with bias checks, cybersecurity safeguards and strict IP/licence compliance.

RAG Components mapped to align HTA Guidance

RAG Component	What RAG Does	CDA Position (Canada)	NICE Position (UK)
Retrieval	Uses trusted sources (e.g., internal databases, real-world data) before generating answers	Encourages AI use for literature searches, evidence identification, and data extraction (via LLMs and search tools)	Supports ML and LLMs for generating search strategies, screening records, and extracting structured data from unstructured sources
Augmentation	Uses retrieved data to enhance, not replace, generative outputs	Emphasizes augmentation of human decisions; AI should support, not replace expert review	Reinforces "human-in-the-loop" principle; AI must complement, not substitute, expert evaluation
Generation	Produces readable summaries, reports, or recommendations based on retrieved evidence	Supports using LLMs to summarize clinical data, reports, dossier, and create lay-friendly outputs	Endorses use of LLM prompts to draft meta-analyses, model scripts, and executive summaries
Transparency	Clearly shows what sources were used and how outputs were derived	Requires full documentation of AI tools used, assumptions, and methods; discourages "black-box" models	Calls for declared use of AI, detailed rationale, and use of tools like ELEVATE-AI, TRIPOD+LLM for validation
Validation	Ensures outputs are verified by humans; performance is tested	Recommends sensitivity analysis, external validation, and reporting of known risks	Demands model validation, simulations, and triangulation of causal inference methods

ELEVATE: ELEVATE-AI LLMs Framework: An Evaluation Framework for Use of Large Language Models in HEOR: an ISPOR Working Group Report | Semantic Scholar
TRIPOD: The TRIPOD-LLM reporting guideline for studies using large language models

References

NICE Position statement – Use of AI in evidence generation: NICE position statement | Our research work | What we do | About | NICE

CDA position statement – Position_Statement_AI_Renumbered.pdf

Padula WV, Kreif N, Vanness DJ et al. (2022) Machine learning methods in health economics and outcomes research—the PALISADE checklist: a good practices report of an ISPOR task force. Value In Health 25(7): 1063–80

ELEVATE: ELEVATE-AI LLMs Framework: An Evaluation Framework for Use of Large Language Models in HEOR: an ISPOR Working Group Report | Semantic Scholar

TRIPOD: The TRIPOD-LLM reporting guideline for studies using large language models

Zemplényi A, Tachkov K, Balkanyi L et al. (2023) Recommendations to overcome barriers to the use of artificial intelligence-driven evidence in health technology assessment. Frontiers in Public Health 11: 1088121

Fleurence RL, Bian J, Wang X et al. (2024) Generative AI for health technology assessment: opportunities, challenges, and policy considerations [online; accessed 17 July 2024]

Closing the Loop: How RAG Solves Key HEOR Challenges

❓ Do we really need Generative AI in HEOR, or is it just a buzzword?

✅ Yes—**Generative AI,** particularly **when combined with RAG**, provides real value by accelerating evidence synthesis, automating documentation, and enhancing decision-making in HTA and HEOR workflows

❓ Can we trust the answers AI gives us?

✅ Yes – when **Gen AI is used in combination with RAG** - responses are anchored in verified HEOR sources, improving factual accuracy and auditability.

❓ Can Gen AI handle the complex real-world documents we deal with—PDFs, tables, scanned pages, even images?

✅ Yes - **RAG can process complex document types** including structured tables, scanned PDFs, and images making it suitable for HEOR documentation workflows.

❓ Is my data safe? Can we use Gen AI without risking sensitive information?

✅ Yes - with secure deployments, **RAG uses only your approved content**—keeping sensitive HEOR data private and compliant.

Thanks.

Questions ???