ChatGPT

Generative AI in Health Economics and Outcomes Research (HEOR): A Pharma-Focused Introduction

Introduction: Generative artificial intelligence (GenAI), particularly large language models (LLMs) like GPT-4, holds *significant promise* for Health Economics and Outcomes Research (HEOR) [1]. In the pharmaceutical industry, these AI models offer a transformative opportunity to streamline *time-consuming, labor-intensive tasks* – from systematic literature reviews and data extraction to economic modeling [2]. By automating or accelerating such workflows, GenAI can augment the productivity of HEOR teams and enable faster evidence generation for decision-making. This report provides a high-level overview of GenAI (with minimal technical jargon) tailored to a pharma audience. We discuss how to access LLMs beyond basic chatbots, the use of prompt engineering to get optimal outputs, crucial privacy and security considerations, and specific applications across key HEOR domains (literature reviews, real-world evidence analysis, and economic evaluations), backed by recent references.

Understanding Generative AI and LLMs in HEOR

GenAI refers to algorithms (notably *foundation models* like LLMs) that can produce human-like content – text, in the case of LLMs – after being trained on vast datasets. An LLM such as GPT-4 can **recognize, summarize, and generate** text that is contextually relevant, making it a versatile tool for any HEOR task involving large amounts of textual or data-driven information [3]. For example, an LLM can read through dozens of clinical study reports or real-world evidence documents and highlight key insights or even draft summaries. In HEOR, where analyses often depend on extensive literature, complex models, and detailed reporting, this capability is game-changing.

Chatbots vs. Direct Integration: Many are first exposed to LLMs via user-friendly chatbots (e.g. ChatGPT) that answer questions in natural language. Chatbot interfaces offer an *accessible entry point* to explore LLM capabilities [4]. However, relying solely on a chatbot website is not ideal for pharmaceutical applications. Beyond simple Q&A, *advanced use of LLMs* requires integrating them into your workflows via **Application Programming Interfaces (APIs)** or specialized platforms [4]. Using an API allows a developer or analyst to send data (e.g. a set of trial results or a manuscript) to the LLM and receive structured outputs programmatically, which is more scalable and flexible than copy-pasting text into a public chatbot. In practice, this means pharma companies can build custom tools (or use AI-powered software) where the LLM works behind the scenes – for example, automatically screening study abstracts or generating draft reports on command – rather than a human manually querying a bot. Embracing these *beyond-chatbot* approaches unlocks the full potential of GenAI in HEOR by enabling automation and integration with existing data systems [5].

Prompt Engineering: Crafting Effective Queries

One key to using LLMs effectively is **prompt engineering** – the art and science of writing inputs that guide the model to produce the desired output. A *prompt* is essentially the question or instruction you give the model, and phrasing it well is crucial [6]. In simple terms, rather than asking an open-ended question and hoping for the best, HEOR researchers can design prompts that set context, provide

examples, or break down a complex task into steps. This improves the accuracy and relevance of the AI's response.

For instance, a basic prompt might be: *"Extract the population of interest from this study abstract."* This zero-shot instruction can indeed make an LLM pull out a population description without any additional guidance [7]. But for more complex requests – say, *"Build a cost-effectiveness model for a new oncology drug"* – a naive one-line prompt will **not** yield a useful result [8]. In such cases, advanced techniques help: **few-shot prompting** (supplying a couple of examples of the desired output format or reasoning) or **chain-of-thought prompting** (asking the model to outline its reasoning step by step) can lead to much more coherent and accurate outputs [6] [8]. Prompt engineering essentially adds structure to the AI's task, which is often necessary in HEOR where multi-step reasoning or domain knowledge is required. By iteratively refining prompts – and even using techniques like *retrieval-augmented generation (RAG)* that feed the LLM with relevant external documents – analysts can coax LLMs to produce high-quality, targeted results rather than generic or incorrect answers [9] [10]. The takeaway for pharma teams is that getting value from GenAI may require developing new skills in formulating queries and instructions, akin to a new form of programming with natural language.

Data Privacy and Security Considerations

When leveraging GenAI in a pharmaceutical context, **privacy and security** are paramount. HEOR often involves sensitive or proprietary information – clinical trial data, patient health records, unpublished economic models – that must be handled in compliance with regulations like the EU's GDPR and the US's HIPAA [11]. A critical risk of using public GenAI services (e.g. a cloud API without proper safeguards) is that confidential data could be inadvertently exposed or used to further train the model. Pharma organizations must therefore approach GenAI with a cautious strategy to protect data integrity and confidentiality.

There are a few strategies to mitigate these risks. One is to use **local or self-hosted LLM deployments**. In a local deployment, an open-source LLM (such as LLaMA or other medically-tuned models) is run on the organization's own secure servers, behind its firewall [12]. This way, any data processed by the model stays entirely in-house. The trade-off is that many of the most advanced LLMs (like GPT-4 or Google's models) are proprietary and not available for local use [13]. Organizations opting for this route need significant IT infrastructure and expertise to maintain the system, but they gain maximum control over security.

Another approach is to use **trusted cloud APIs with strict data agreements**. Some vendors allow enterprises to use their LLMs with guarantees that data won't be stored or will be isolated. For example, OpenAI offers a service where user data is not used to train models by default for business accounts. Even so, legal and IT teams should vet any third-party AI service for compliance with data protection standards. Companies should also consider techniques like encrypting or anonymizing data before sending it to an LLM, and only sending the minimum necessary information to get the job done.

In summary, implementing GenAI in pharma HEOR demands *robust security protocols* and thoughtful deployment choices [11]. Whether by **secure on-premises models** or **carefully governed cloud solutions**, the goal is to harness AI's power **without compromising patient privacy or corporate intellectual property**. Many early adopters combine measures – e.g. testing ideas with public models on dummy data, then shifting to a private model for real data – to balance innovation with compliance. As this field matures, we are also seeing the development of best practice frameworks to ensure safe AI use (for instance, guidelines that cover model selection, data handling, and audit trails for GenAI-assisted research).

Applications of GenAI Across Key HEOR Domains

Perhaps the most exciting aspect of GenAI in HEOR is the breadth of its potential uses. Below, we explore how LLM-driven tools are being applied in three core HEOR domains: **systematic literature reviews (evidence synthesis)**, **real-world evidence analysis**, and **health economic evaluations**. Each area illustrates how GenAI can enhance efficiency and insights, as well as the remaining need for human expertise and oversight.

Accelerating Systematic Literature Reviews and Evidence Synthesis

Conducting a systematic literature review (SLR) is traditionally a resource-intensive process: researchers must sift through hundreds or thousands of citations, apply inclusion/exclusion criteria, extract data, assess study quality, and synthesize findings. Generative AI has emerged as a powerful assistant in this domain. Foundation LLMs can be employed at multiple steps of an SLR workflow, effectively acting as an *AI research assistant*.

Citation Screening: One of the earliest demonstrated uses of LLMs in literature reviews is speeding up the screening of titles and abstracts. Rather than two human reviewers laboriously checking each record, an LLM can take on one of the reviewer roles. Recent research found that GPT-4, when used as an **additional reviewer**, achieved *almost perfect agreement* with human reviewers in abstract screening (Cohen's kappa > 0.9) [14]. In fact, the study concluded GPT-4 could potentially replace one human screener, reliably distinguishing included vs. excluded studies according to PRISMA-guided criteria [15]. This level of consistency suggests AI can dramatically reduce the manual workload in the initial triaging of studies.

Data Extraction and Summarization: Beyond screening, GenAI can assist in pulling out key details from full-text papers. By prompting an LLM to extract specific data points (e.g. patient population, intervention, outcomes), researchers can quickly build evidence tables. These models have shown they can reliably identify PICO elements (Population, Intervention, Comparator, Outcome) from unstructured text [16]. Impressively, LLMs are even being used to automate more complex evidence synthesis tasks. In one instance, GPT-4 was able to replicate data extraction across multiple network meta-analysis (NMA) studies with **over 99% accuracy**, correctly pulling relevant variables and results from published NMA datasets [17]. The AI essentially read through the studies and captured the needed numbers as accurately as a human expert in that case.

Drafting and Analysis: Generative models can also support writing the review itself. They can summarize findings from a body of literature, draft narrative synthesis paragraphs, or even generate forest plots and statistical code for meta-analysis. For example, given a set of trial results, an LLM could be prompted to produce a concise summary comparing the outcomes, or to write R code that computes a meta-analytic estimate [18]. This can save valuable time in the analysis phase of an SLR. Some early projects have integrated LLMs to **justify exclusion reasons** for studies in a review or to perform risk-of-bias commentary based on study descriptions, tasks that typically require expert judgment [19].

Remaining Challenges: While these applications are promising, they are not without pitfalls. AI models sometimes **hallucinate** – generating plausible-sounding information that is incorrect or not actually stated in the source [20]. For instance, an LLM might confidently mislabel a study as a randomized trial when it was observational, or invent a reason for exclusion that isn't true. Inconsistent classification and extraction errors still occur, especially if the inclusion criteria or data fields are nuanced [20]. Thus, human oversight remains *essential*. The outputs of an LLM-assisted SLR must be validated by HEOR professionals to ensure scientific rigor [20]. Until standardized evaluation benchmarks for AI in literature

reviews are established, experts should double-check AI recommendations and maintain a role in final decision-making. Overall, the integration of GenAI can **accelerate evidence synthesis tremendously**, freeing up researchers to focus on interpretation and context, but it works best as a partnership where AI handles the heavy lifting under human guidance.

Enhancing Real-World Evidence (RWE) Generation

Real-world evidence studies make use of data from outside the controlled environment of clinical trials – for example, electronic health records (EHRs), insurance claims, registries, and even patient-reported data. These datasets pose a unique challenge: they are often **large, messy, and partially unstructured**. Here, generative AI – especially LLM-based natural language processing – can revolutionize how we extract insights from raw healthcare data.

Parsing Unstructured Clinical Text: A huge portion of valuable real-world data is buried in free-text form (think doctor's notes, discharge summaries, pathology reports). Traditional analytics might ignore these rich text fields or require months of manual abstraction. LLMs can dramatically speed up this process by reading and interpreting unstructured text. For example, generative models have been applied to pull out specific information like biomarker statuses from pathology reports or oncology notes that are not coded elsewhere [21]. In doing so, the AI can turn narrative text into structured data for analysis. One study showed that a foundation model could accurately extract details such as whether a patient underwent biomarker testing from clinical notes, which would otherwise require manual chart review [21]. This reduces human effort and the risk of oversight in RWE data preparation.

Scale and Speed: With LLMs, handling *volume* becomes easier. They can rapidly summarize a patient's medical history from multiple notes, or classify thousands of reports by content (e.g., find all patients who mention a certain symptom). Such capabilities let researchers operate at a scale not feasible through manual effort alone, potentially identifying real-world patient cohorts or outcomes faster than before [22] [23]. As a result, assembling an RWE dataset for an outcomes study – say, all patients with disease X who had outcome Y – can be accelerated using GenAI to triage and organize raw data sources.

Advanced Techniques – Fine-Tuning and Domain-Specific Models: It's worth noting that general-purpose LLMs may not always perform perfectly on highly specialized medical texts out-of-the-box. In fact, one experiment found that a baseline generative model struggled to map free-text descriptions of diagnoses to standard medical codes, achieving *below 50% accuracy* in that task [21]. However, that model had not been fine-tuned for the domain. By **fine-tuning** (further training) LLMs on biomedical text or by using domain-specific LLMs (models like **GatorTron** or **NYUTron** that are pretrained on clinical data), performance improves substantially [24]. These specialized models better grasp clinical jargon and context, resulting in more reliable extraction of information from EHR notes. Prompting strategies also matter here – carefully wording queries for the medical context can help general models like GPT-4 yield more accurate results [24].

Beyond Text – Multimodal Data Integration: Emerging generative AI approaches are starting to combine text with other data modalities to enrich RWE. For instance, researchers have experimented with models that take both structured data (like lab results) and unstructured text (like clinician notes) as input to predict outcomes or utilization. A notable example integrated textual data and genomic information to forecast COVID-19 hospitalizations, outperforming traditional epidemiological tools [25]. This hints at a future where GenAI might fuse different real-world data streams – clinical notes, imaging reports, genomic sequences – to provide **deeper and more nuanced insights** for HEOR studies [25]. In pharmacoeconomics, such multimodal analysis could help correlate real-world patient characteristics with outcomes and costs in ways not previously possible.

Quality and Validation: As with literature reviews, the use of AI in RWE must be accompanied by rigorous quality checks. Real-world data can be noisy and any AI-extracted dataset should be validated against a subset reviewed by humans. Inconsistencies or errors need to be caught early to avoid skewing study results. Encouragingly, the trend is that as models get more specialized and techniques improve, the reliability of AI-curated RWE datasets is steadily increasing [26]. The end goal is to attain **high-quality RWE faster**, to inform value evidence and outcomes research without compromising accuracy. Until then, combining GenAI with human domain expertise is the safest route – the AI rapidly organizes and distills the data, and the human experts ensure it all makes sense.

Supporting Economic Evaluations and Modeling

Health economic evaluations – such as cost-effectiveness analyses, budget impact models, and other pharmacoeconomic models – are another cornerstone of HEOR. These models synthesize evidence on costs and outcomes to inform decisions about healthcare value (e.g., pricing, reimbursement, formulary inclusion). Building and updating economic models is traditionally a meticulous, expert-driven task. Generative AI is beginning to *reshape how economic models are developed and used*, offering assistance in everything from literature review for model inputs to automating parts of model construction.

Summarizing and Extracting Model Inputs: An early use of LLMs in this arena is reading through existing economic evaluation studies to pull out key parameters and assumptions. Imagine you have numerous published studies on the cost-effectiveness of treatments in diabetes – an LLM can quickly summarize the range of inputs (like utility values, costs of complications, etc.) and outcomes reported, saving analysts hours of manual review. LLMs can identify and extract relevant model parameters from text, such as a transition probability or a health state utility value buried in the methods section of a paper [3] [27]. By doing so, GenAI reduces the time spent sourcing data inputs for new models. It can also highlight the structure of a published model (e.g., the model has X health states, or uses a Markov cycle length of 3 months) by interpreting the study's descriptions.

Automating Model Replication and Development: More ambitiously, researchers are testing whether LLMs can *reconstruct* entire economic models from literature. In some proof-of-concept cases, AI tools have successfully recreated simplified versions of well-known models. For example, large language models have been used to replicate **multi-state partitioned survival models in oncology** and **infectious disease cost-effectiveness models**, by reading the published papers and automating tasks like mapping out the model structure, retrieving parameter values, and even generating analysis code [28]. A recent case study went further: a team used GPT-4 (via a platform called *ValueGen.AI*) to rebuild a *complex Markov model* for ulcerative colitis entirely from the text of a published article and a technical report [29] [30]. The AI was tasked with extracting all model components (health states, transition probabilities, costs, utilities, etc.) and assembling a working cost-effectiveness model without access to the original Excel or code.

Findings from the Case Study: The results of that experiment highlight both the current capabilities and limitations of GenAI in economic modeling. On one hand, GPT-4 accurately identified many clearly-described elements of the model. For instance, it correctly picked up the names and definitions of distinct health states from the technical report, and it extracted certain cost inputs (like surgical costs) with precision when they were presented in a structured way [30]. This shows that when model documentation is transparent (with tables, headings, and clear definitions), an LLM can parse it effectively. On the other hand, the AI struggled with more complex or implicit information. It **hallucinated some non-existent health states** (likely by misinterpreting narrative text) and had difficulty understanding transition logic that wasn't explicitly stated (e.g. conditional treatment sequences or when a hazard ratio implied a transition to death) [31] [32]. Similarly, scattered cost data (like drug costs mentioned across paragraphs) were not always captured, and some utility values were

misassigned due to ambiguous context [33]. Notably, reconstructing treatment pathways – the sequence of lines of therapy a patient could follow in the model – proved beyond the AI's grasp when the information was only in descriptive form [32]. These shortcomings underline that the AI is not infallible; complex modeling logic often requires the nuance and deep understanding that human health economists provide.

Implications and Future Directions: Despite its limitations, the success in partially rebuilding a model suggests a **promising new use case** for generative AI in HEOR [34]. In the future, such tools could dramatically reduce the effort to adapt existing models for new settings (e.g., new countries or indications) by auto-extracting inputs and structure from literature and then allowing experts to refine them. It could also enhance transparency: if AI can check published models for consistency by attempting to rebuild them, it might reveal gaps or ambiguities in documentation. That points to an opportunity – as noted by the ulcerative colitis case study – to improve the standardization of how we document models so that both humans and AI can understand them better [35]. Looking ahead, we might see AI assisting with **validating models** (e.g. comparing model predictions to real-world data in an automated way), converting models between programming languages or platforms (imagine instantly turning an Excel model into an R script), or exploring scenario analyses much faster than before [36].

Crucially, in economic decision models, **human judgment remains indispensable**. When results are being used for important decisions like pricing or reimbursement submissions to health technology assessment (HTA) agencies, the bar for accuracy and transparency is very high [37]. Every assumption and equation in a model might be scrutinized. Therefore, current consensus is that GenAI should serve as a **powerful support tool** – handling groundwork and suggesting solutions – but *not a substitute* for the expertise of health economists in model design and validation [37]. Any AI-generated model components must be carefully reviewed, and the final model should be thoroughly validated (just as one would validate a human-built model). As evaluation frameworks for AI in modeling improve, we can increase trust in these outputs. But until then, AI is the junior analyst and the human remains the lead analyst, ensuring that the end product meets the required scientific and regulatory standards.

(Note: Another emerging application in HEOR worth mentioning is evidence communication. GenAI tools are beginning to assist in drafting value dossiers, HTA submissions, and reports. They can automatically generate well-structured documents or slide decks that compile evidence and follow specific templates – for example, creating a first draft of a global value dossier by pulling in efficacy, safety, and cost-effectiveness results in a standardized format. This can save considerable time for medical writers. AI-driven drafting can also help tailor communications to different audiences, producing both technical documentation and layperson summaries from the same data [38]. As with other uses, quality control is critical – any content generated for regulatory or payer audiences must be vetted line-by-line to ensure correctness [39]. Still, this is a promising area where GenAI might expedite the last mile *of HEOR: translating analysis into clear, impactful communications.)*

Challenges, Ethical Considerations, and Best Practices

While generative AI offers exciting opportunities for HEOR, it also introduces new challenges and risks that must be managed responsibly. **Model accuracy and "hallucinations"** are a primary concern – LLMs can produce information that looks plausible but is actually incorrect or fabricated [40]. In a scientific context, such mistakes can mislead conclusions if not caught. Researchers using GenAI should always cross-verify critical outputs (like an extracted data point or a summarized finding) against original sources. Developing a habit of skepticism toward AI output – "trust but verify" – is healthy in these early days.

Reproducibility is another issue. LLMs might not produce the exact same answer every time a prompt is run, due to their probabilistic nature. This variability can be problematic for scientific reporting, where ideally one would like to get consistent results or at least understand the margin of error. To improve reproducibility, it's recommended to **document the AI's parameters and prompt designs in detail** so others can follow the same process [41]. In fact, transparency in how GenAI is used is becoming a key part of *reporting guidelines*. A 2025 ISPOR working group introduced the **ELEVATE-GenAI framework**, which outlines structured guidelines for reporting studies that involved LLMs [42] [43]. The framework covers domains like describing the model used (and its version), the prompt strategy, how accuracy was measured, and how bias was assessed, among others [44]. The aim is to ensure that when AI is part of HEOR research, the published work clearly states how the AI contributed and how its outputs were validated, so that readers and decision-makers can trust the findings [45].

Bias and Fairness: LLMs learn from vast datasets that inevitably contain human biases (gender, racial, etc.) and gaps in knowledge. If not careful, using an LLM could propagate biases – for example, an AI summary of literature might underrepresent studies from certain regions or produce answers that are skewed by more frequently published data. It's important to be aware of this risk. Techniques to mitigate bias include fine-tuning models on more diverse or balanced datasets and explicitly checking AI outputs for signs of unfair bias. Some of the reporting frameworks explicitly call out the need to evaluate *fairness and bias* as part of any AI-driven analysis [44]. In pharma HEOR, maintaining scientific objectivity and equity (e.g., in evidence consideration) is crucial, so any AI assistance must be critically appraised for these aspects.

Ethical and Regulatory Compliance: Beyond privacy (discussed earlier), there are broader ethical questions. For example, if an AI model assists heavily in writing a manuscript or report, how should it be credited or acknowledged? Who is accountable for errors the AI makes – the tool provider or the researchers who used it? The consensus so far is that responsibility lies with the humans using the AI, and transparency is key: disclose when and how AI was used in the research or writing process. From a regulatory perspective, agencies are beginning to pay attention. HTA bodies and journals may start expecting declarations of AI assistance and even reviewing AI-generated analyses with extra scrutiny. It's wise for pharma companies to stay abreast of any guidelines from authorities (like NICE's statements or ISPOR working groups) on the acceptable use of AI in submissions.

Finally, **intellectual property (IP)** considerations arise: LLMs might output text that inadvertently resembles the training data, which could be copyrighted material. Using GenAI to draft a report requires ensuring the content doesn't plagiarize existing sources. Most well-known LLMs have measures to avoid verbatim regurgitation of training text, but it's another reason human review is needed – to make sure any references are properly cited and no unseen copyright issues are present.

Best Practices: To harness GenAI in HEOR effectively while mitigating risks, experts recommend a few best practices. These include: - *Start small and pilot:* Test AI on non-critical tasks first (e.g., summarizing a few papers) and validate its outputs thoroughly. Learn its failure modes in your context. - *Human in the loop:* Never let the AI work unchecked. Use it to augment, not replace, human analysts. Have humans review and approve any AI-generated result that will be used in decision-making or published. - *Maintain transparency:* Keep a clear record of how AI was used – e.g., save your prompts, note which version of a model was used, and document any parameters or settings. This is invaluable for internal quality control and external reporting [46] [45]. - *Secure your data:* Follow IT and compliance guidelines when handling data with AI. If using cloud services, ensure data is de-identified if possible and that vendor agreements are in place. If using local models, keep them updated and secure just like any other software containing sensitive information. - *Educate and train staff:* Build competency in AI tools among HEOR teams. Training in prompt engineering and basic understanding of AI limitations will empower teams to use GenAI effectively and avoid misuse.

By approaching GenAI adoption in HEOR with these safeguards, pharma organizations can innovate confidently and responsibly.

Conclusion and Future Outlook

Generative AI, and LLMs in particular, are poised to become integral allies in the work of health economists and outcomes researchers. For pharmaceutical companies, the ability of GenAI to rapidly synthesize information, draft analyses, and support complex decision models can translate into **faster insights and more efficient evidence generation**. This can ultimately aid quicker decision-making on drug development, market access, and value demonstration. Early applications in literature review, RWE analytics, and economic modeling already show how, when used wisely, GenAI can save time and augment human expertise without sacrificing quality [47] [3].

Looking ahead, experts foresee even more transformative potential. We may soon have **living systematic reviews** or economic models that are continuously updated as new data emerge, powered by LLMs that constantly ingest and summarize the latest evidence [48]. Patients and clinicians could benefit from HEOR findings translated into accessible language by AI, narrowing the gap between technical research and practical understanding [48]. And within pharma companies, HEOR teams might use GenAI-driven platforms to instantly explore "what-if" scenarios in models or to generate real-time analyses during payer negotiations.

However, realizing this potential requires navigating the challenges outlined: ensuring data privacy, maintaining rigorous validation, and embedding ethical guardrails. The next few years will be critical for establishing **best practices and industry standards** for GenAI in HEOR [49]. Initiatives like ISPOR's working groups and emerging case studies will guide the community on how to balance innovation with scientific rigor. If done responsibly, the payoff is substantial – a future where HEOR activities are not only more efficient, but also more dynamic and insightful, ultimately supporting the delivery of effective, value-based healthcare.

In conclusion, **generative AI is not a magic wand, but it is a powerful new tool** in the HEOR toolbox. Pharma organizations that learn to wield this tool adeptly (with appropriate caution) stand to enhance their evidence generation and value demonstration capabilities. By combining the strengths of AI – speed, scale, and pattern recognition – with the irreplaceable expertise of human researchers, the HEOR field can evolve to meet the demands of an increasingly data-rich, fast-paced healthcare environment. Embracing GenAI in a thoughtful way today will prepare teams for a future where such technology is ubiquitous. The goal is to *work smarter*, not harder, and let generative AI handle the grunt work while HEOR professionals focus on strategy, interpretation, and delivering insights that improve health outcomes.

References:

1. Fleurence RL *et al.* (2025). *ELEVATE-GenAI: Reporting Guidelines for the Use of Large Language Models in HEOR*. **Value in Health, 28**(11):1611-1625. [1] [43]

2. Reason T *et al.* (2025). *Using Generative AI in Health Economics and Outcomes Research: A Primer on Techniques and Breakthroughs*. **PharmacoEconomics - Open, 9**:501–517. [2] [50]

3. Landschaft A *et al.* (2024). *Implementation and evaluation of an additional GPT-4-based reviewer in PRISMA-based medical systematic literature reviews*. **Int J Med Inform, 189**:105531. [14] [15]

4. Maxis Clinical Sciences (2023). *What Generative AI Means for HEOR: A Practical Taxonomy of Tools and Applications* (blog). [51] [21]

5. Value Analytics Labs (2025). *Can Generative AI Rebuild Health Economic Models?* (Ulcerative Colitis case study blog). [30] [52]

6. Peng C *et al.* (2023). *A study of generative large language model for medical research and healthcare*. **npj Digital Medicine, 6**(1):210. [21] [24]

7. Reason T *et al.* (2024). *Artificial Intelligence to Automate Network Meta-Analyses: Four Case Studies to Evaluate the Potential of LLMs*. **PharmacoEconomics - Open, 8**(2):205-220. [17]

8. ISPOR Generative AI in HTA Working Group (2024). *Generative AI for Health Technology Assessment: Opportunities, Challenges, and Policy Considerations*. **Value in Health** (Working Group Report). [37] [48]

---

[1] [42] [43] [44] [45] [46] ELEVATE-GenAI: Reporting Guidelines for the Use of Large Language Models in Health Economics and Outcomes Research: An ISPOR Working Group Report - PubMed
https://pubmed.ncbi.nlm.nih.gov/40653157/

[2] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [40] [41] [48] [49] [50] Using Generative Artificial Intelligence in Health Economics and Outcomes Research: A Primer on Techniques and Breakthroughs | PharmacoEconomics - Open
https://link.springer.com/article/10.1007/s41669-025-00580-4

[3] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [36] [37] [38] [39] [47] [51] What Generative AI Means For HEOR I Blog
https://maxisclinical.com/insights/blog/what-generative-ai-means-for-heor-a-practical-taxonomy-of-tools-and-applications/

[14] [15] Implementation and evaluation of an additional GPT-4-based reviewer in PRISMA-based medical systematic literature reviews - ScienceDirect
https://www.sciencedirect.com/science/article/pii/S1386505624001941

[29] [30] [31] [32] [33] [34] [35] [52] Can Generative AI Rebuild Health Economic Models? Insights from an Ulcerative Colitis Case Study – Value Analytics Labs
https://www.valueanalyticslabs.com/can-generative-ai-rebuild-health-economic-models-5270/