ChatGPT

How AI and Advanced Analytics Are Transforming Health Technology Assessment

Health technology assessment (HTA) is experiencing a paradigm shift thanks to cutting-edge artificial intelligence (AI) and analytic techniques. Recent developments show that **generative AI** and **machine learning (ML)** can dramatically enhance how we evaluate health technologies, from simulating complex committee deliberations to extracting insights from dense reports. Below, we explore key areas where these technologies are optimizing HTA processes, improving efficiency, accuracy, and the depth of analyses:

- **Simulating HTA Deliberations:** Using generative AI (large language models) to mimic multi-stakeholder HTA committee discussions and anticipate evidence gaps.
- **Automating Evidence Extraction:** Leveraging AI and natural language processing (NLP) to extract data from HTA reports across agencies, enabling faster comparative research.
- **Predicting Health Outcomes with ML:** Comparing machine learning approaches to traditional statistical methods for predicting patient outcomes (e.g., EQ-5D quality-of-life scores).
- **Enhancing Economic Models:** Applying AI to refine health economic model structures, adding clinical nuance to improve credibility and decision-making.

Each of these innovations exemplifies how AI-driven tools are streamlining HTA workflows while enriching the insights available to decision-makers.

Generative AI in HTA Committee Simulations

One groundbreaking application is the use of **generative AI to simulate HTA committee meetings** and sponsor interactions. By deploying multi-agent large language models (LLMs) that role-play different stakeholders (e.g. clinical experts, health economists, patient representatives), researchers have been able to generate realistic **clarification questions** and discussions mirroring those in actual HTA reviews [1] [2]. For example, a recent study created a role-based LLM framework representing a NICE (UK) appraisal committee – including agents for the HTA secretariat, clinical and economic reviewers, patient/public voices, etc. The AI-driven "virtual committee" was fed a drug's submission data and **successfully produced 104 pertinent questions** (spanning clinical efficacy, economic modeling, and textual clarity) that an expert panel judged to be 80–85% concordant with real questions raised by NICE [2]. Notably, even the remaining 15–20% of AI-generated queries, while not matching historical examples, were deemed relevant and evidence-based [3]. This indicates that **LLMs can closely emulate complex HTA deliberations**, uncovering potential evidence gaps and lines of inquiry that HTA bodies might pursue. Such simulations can greatly **enhance sponsor preparedness** for committee meetings and streamline the review process [4]. They offer a cost-effective, scalable way to "pressure test" evidence packages before actual submission, providing early insight into what tough questions may arise.

Beyond generating questions, AI-driven virtual committees demonstrate how **LLMs can foster robust debate** on a technology's merits and uncertainties. In one case, researchers used GPT-4 to simulate a full NICE appraisal discussion for a cancer therapy, assigning each LLM "committee member" a persona with specific expertise and even probabilistic biases [5] [6]. The motivation was to address the challenge that convening expert panels or advisory boards is time-consuming and expensive, whereas an AI-simulated discussion can rapidly iterate on different scenarios [7]. Early findings showed that these virtual committees could **summarize evidence and surface key issues** similar to human panels [8]. While AI deliberations are not (yet) a substitute for real experts, they illustrate the potential of generative AI to **augment human decision-making –** for instance, by highlighting contentious points or consensus areas in HTA evaluations. As LLM capabilities advance, we may see them used to support training of HTA professionals or to facilitate broader stakeholder engagement by simulating patient and clinician perspectives in technology assessments.

AI-Powered Data Extraction from HTA Reports

HTA agencies worldwide produce extensive reports evaluating clinical and economic evidence, but mining these documents for research or cross-country comparison is labor-intensive. **Artificial intelligence is now being harnessed to automatically extract and synthesize data from HTA reports**, greatly improving the efficiency of evidence reviews. A recent study demonstrated this by applying multiple NLP techniques to **NICE assessment reports**, extracting 14 key attributes (e.g. study outcomes, comparative effectiveness conclusions, etc.) that researchers often need [9]. The approaches compared ranged from rule-based text mining to machine learning classifiers and a generative AI method using a state-of-the-art LLM (Anthropic's Claude model) [9]. Strikingly, the **LLM-based extraction outperformed the others**, achieving about *88–98% accuracy on 12 out of 14 data fields* [10]. Areas like the stated outcome of the relative effectiveness assessment or the specified comparator treatment were a bit more challenging (around 70% accuracy for the LLM), but overall the AI was highly effective at pulling structured insights from unstructured text [10]. In contrast, traditional rule-based NLP required extensive upfront coding and still struggled to capture certain context-specific details (like nuances at the specific medicine-indication level) [11]. The generative model did have some drawbacks **–** for example, reliance on a commercial AI service and some reproducibility issues when re-running extractions **–** yet it enabled the creation of updatable data visualizations and comparative graphs that **would have been very difficult to assemble manually** [12] [13]. This proof-of-concept shows that **AI can automate evidence extraction from HTA documents**, paving the way for near-real-time monitoring of how different agencies evaluate therapies.

Beyond single-agency reports, AI is also breaking new ground in **cross-country HTA analysis**. Using a generative AI tool called *ValueGen.AI*, researchers processed multiple HTA reports on the same drug (tofacitinib for ulcerative colitis) from **three national agencies (NICE in England, HAS in France, and G-BA in Germany)** [14] [15]. The system (built on GPT-4 with retrieval augmentation) could parse documents in English, French, and German, automatically **identifying common and divergent critiques** across jurisdictions [16] [17]. For instance, all agencies raised concerns about trial design and comparators, while only NICE flagged issues in the economic model and safety profile [18] [19]. The AI synthesized these findings into concise summaries, demonstrating a powerful approach to navigate the complex global HTA landscape. By **leveraging generative AI to analyze multi-language HTA texts and pinpoint shared vs. unique points of contention**, stakeholders (manufacturers, payers, policymakers) can gain a panoramic view of evidence requirements and decision drivers in different markets. This kind of automation not only saves countless hours of manual review but also supports more **data-driven, harmonized decision-making** in health economics and outcomes research [17] [20].

Machine Learning vs Traditional Methods for Outcome Prediction

Another area where advanced analytics are making waves is in **predicting health outcomes and patient-reported measures**, a common need in economic modeling and HTA when data are missing. Traditionally, statisticians have used regression-based methods to predict metrics like the EQ-5D index (a standardized health-related quality-of-life score). Now, **machine learning approaches (e.g. tree-based algorithms and neural networks)** are being evaluated against these classical techniques to see which performs better in capturing patient health states. A comprehensive comparison was recently conducted using the Health Survey for England dataset, predicting EQ-5D-3L index values from patient characteristics [21] [22]. The study tested six modeling approaches: two traditional regressions (ordinary least squares and a logistic model for dimensions) versus four ML models (including **XGBoost** gradient-boosted trees and **neural networks**), and examined both direct prediction of the index and indirect prediction via EQ-5D dimensions [23]. The results were illuminating **– neural networks outperformed all other models** in terms of error and accuracy [24]. The best-performing network (a simple one-hidden-layer NN) achieved the lowest mean absolute error (~0.088) and correctly classified health state severity groups 73% of the time, edging out the next-best ML model (XGBoost) which had ~0.092 error and 72% accuracy [24]. In contrast, the traditional regression models had roughly similar error margins but notably **lower accuracy in distinguishing severe health states**, particularly underestimating issues related to pain and anxiety [24]. This suggests that **ML algorithms can capture complex, non-linear relationships in quality-of-life data** better than linear models, thereby providing more nuanced estimates of patient outcomes.

It's worth noting that the performance gap, while consistent, was not enormous – the tree-based ML models were "competitive" and regression wasn't far off in average error [24]. However, even modest improvements in predictive accuracy can be important in HTA, where **small differences in quality-adjusted life year (QALY) estimates or risk predictions might influence reimbursement decisions**. Moreover, the machine learning models brought other benefits: they identified key predictors of quality of life (such as employment status, presence of chronic conditions, self-rated health, and recent illness) using techniques like SHAP for explainability, highlighting that **clinical factors dominated demographics in driving EQ-5D outcomes** [25]. This kind of insight is valuable for policymakers aiming to improve patient outcomes **–** it directs attention to the most impactful health determinants. The upshot is that **hybridizing traditional health-economic analysis with ML can enhance both precision and understanding**. While regressions remain useful for their transparency, modern ML approaches (when properly validated) offer a powerful complement or alternative for outcome prediction, treatment effect modeling, and other areas of HTA where richer predictive performance is needed.

Refining Health Economic Models with AI for Clinical Nuance

Health economic models (like cost-effectiveness models) are the backbone of HTA, but they often simplify clinical reality into a limited number of health states or assumptions to remain tractable. Advanced AI techniques are now helping **bridge the gap between necessary simplification and clinical nuance**. A novel approach introduced in 2025 uses a "hybrid intelligence" workflow – essentially combining human expertise with AI assistance – to **"re-expand" simplified model structures into more clinically detailed ones** [26] [27]. The idea is to start with a pared-down model (for example, a basic Markov model with a few health states) and let an AI system guide an exploration of what might be missing. Researchers achieved this by using LLMs with specialized prompt engineering and a graph-based retrieval augmented generation (GraphRAG) setup [28]. The AI would deconstruct the given model, analyze disease knowledge (from literature or guidelines), and suggest **potential additional health states or pathways** that could be relevant but were collapsed in the simplified version [28]. Through iterative dialogue and reasoning, the system proposed enriched model structures – for instance, adding in **treatment sequencing steps, complication pathways, or explicit quality-of-life trajectories** that the original model hadn't individually represented [29]. Human experts (health economists and clinicians) then reviewed these AI-suggested refinements to ensure they made sense medically and within the context of available data [30]. Importantly, this process not only expanded model fidelity but also helped clarify **why certain complexities were omitted** in the first place [31]. In one case, the AI identified that a "simplified" model for a chronic disease lacked an explicit post-complication state; adding it improved conversations with clinical stakeholders and made the model's assumptions more transparent [29]. In essence, the AI acted as a catalyst for modelers to **zoom in on clinical details** that could have policy significance, thereby making the model more credible to clinicians and HTA authorities.

The outcome of this approach was twofold. First, it **produced more refined economic models** that better mirrored real-world disease progression, without losing sight of feasibility. Second, it served as a communication aid – the enriched models were visually clearer and easier to explain to non-modelers, facilitating faster alignment with clinical experts and even preparing for tougher questions from payers [31]. By highlighting what had been **"deliberately excluded"** versus what was now added, the method turned model simplification into a more conscious, explainable choice [31]. This is crucial in HTA submissions, where transparency about model assumptions can influence an agency's trust in the results. The broader implication is that **AI-supported modeling can enhance both technical accuracy and stakeholder confidence**. As one summary put it, this hybrid intelligence strategy provides a scalable way for organizations to reconcile the need for operational simplicity with the demand for clinical credibility in their economic models [32]. Going forward, we can expect AI tools to become a standard part of model development – not to replace modelers, but to augment their ability to test alternatives, ensure no critical clinical detail is overlooked, and communicate complex models in a more digestible manner.

Boosting Efficiency, Accuracy, and Depth in HTA Research

Across these examples, a common theme is that **AI and advanced analytics are elevating the efficiency, accuracy, and depth of HTA activities**. The efficiency gains are perhaps the most immediately tangible: tasks that once took weeks of manual effort – reading hundreds of pages of HTA reports, mapping out potential questions, or tweaking model structures – can now be accelerated with AI-driven tools. For instance, an NLP/LLM system can scan and extract essential data points from dozens of assessment reports in minutes, enabling analysts to focus on interpretation rather than information retrieval [12] [17]. Generative AI can automate literature screening or evidence synthesis steps in systematic reviews, as highlighted by an ISPOR working group [33]. Even in model development, AI's ability to rapidly suggest and evaluate new scenarios means analysts can iterate faster and cover more ground than traditional methods allow.

In terms of **accuracy**, the introduction of sophisticated ML models and AI validation is helping to catch nuances that simpler approaches might miss. The EQ-5D prediction example illustrated that neural networks could better reflect patients' true health state distribution (especially on subjective domains like pain/anxiety) than linear models [24]. Similarly, an AI-simulated committee might surface a pertinent question about data uncertainty that a sponsor team hadn't considered, thus preventing a blind spot in the submission. AI isn't infallible, but when used judiciously alongside human expertise, it serves as a second pair of eyes – or rather, millions of eyes – cross-checking patterns and assumptions. In fact, **HTA bodies themselves acknowledge the potential for AI to reveal hidden patterns** in large datasets and generate novel insights; NICE's 2024 position statement on AI noted benefits like "processing and analyzing large datasets to reveal patterns and relationships that may not otherwise be readily apparent" [34]. By delegating certain complex pattern-recognition tasks to algorithms, analysts can achieve a more **comprehensive and accurate evidence base** for decision-making [34].

Perhaps most interestingly, AI is deepening the **analytical depth** of HTA. It enables exploration of "what if" scenarios and subtle considerations that resource or time constraints used to force aside. With generative models, one can simulate **deeper discussions** or broader stakeholder input around a technology (e.g., including patient perspectives in an AI-driven simulation to ensure those views are considered). With NLP, one can aggregate data from many sources, potentially identifying *systematic differences* in how agencies evaluate evidence or what outcomes matter across contexts. With model refinement tools, one can examine the impact of adding greater clinical granularity and thereby understand which simplifications are harmless and which might be altering conclusions. All these contribute to an HTA process that is richer in insight – not just faster or cheaper. For example, after automating the extraction of numerous HTA reports and visualizing the results, researchers were able to derive **policy-relevant insights that would have been difficult to see through manual review alone** [12] [35]. By processing more information and highlighting non-obvious connections, AI can help HTA analysts and decision-makers probe questions like: Are there consistent evidence gaps across similar appraisals? How might real-world evidence alter a model's outcome? What patient subgroups drive most of the uncertainty or value? In short, advanced computational techniques are equipping us to ask – and answer – deeper questions in health economics and outcomes research.

Challenges and the Path Forward

While the promise of AI in HTA is enormous, it comes with **important caveats and responsibilities**. Both researchers and regulators stress that human oversight and methodological rigor are paramount when incorporating AI-driven findings. Generative AI, for instance, can sometimes produce outputs that are *persuasive but incorrect*, or reflect biases present in training data. **HTA agencies are proceeding cautiously** – NICE has explicitly warned of risks like algorithmic bias, reduced transparency, and the danger of diminishing expert oversight if one over-relies on AI [36]. In its guidance, NICE advises that AI-generated evidence should typically *augment, not replace*, traditional evidence and expert analysis [36]. Any submission utilizing AI methods is expected to **justify their use, provide thorough validation (e.g. sensitivity analyses, comparisons with conventional results), and clearly explain methods and assumptions in an accessible way** [37] [38]. This means if a company uses a machine learning model to indirectly estimate a treatment effect or a cost parameter, they should cross-verify that with established approaches and be transparent about the AI's workings. Tools to improve explainability – such as LIME or SHAP for model outputs, or checklists like ISPOR's PALISADE for ML in HEOR – are increasingly recommended to accompany any AI-driven analysis [39] [40]. The emphasis is on maintaining confidence that decisions are evidence-based and understandable to all stakeholders, even as the methods for generating evidence evolve.

Moreover, the field is recognizing the need for **standards and best practices** for AI in HTA. An ISPOR expert group recently highlighted the necessity for guidelines on responsibly integrating generative AI into HTA workflows [41]. They pointed out challenges around scientific validity, bias, and ethics – for example, if an LLM assists in literature review, how do we ensure it doesn't inadvertently exclude critical studies or inject any favoritism? They concluded that generative AI should support human efforts, *not* replace them, underscoring that **human experts remain the final arbiters of quality and relevance** in HTA assessments [42] [43]. Similarly, researchers are encouraged to stay educated about the latest AI capabilities and limitations, and to continuously evaluate how these tools perform in practice [44]. As these technologies mature, we can expect HTA bodies to update their methodology guides, much like how NICE issued a position statement early to shape the appropriate use of AI in evidence generation [45] [36]. This proactive stance is crucial – it ensures innovation can flourish under a framework that safeguards scientific integrity.

In summary, the convergence of AI and HTA heralds a new era where we can tackle assessments with greater scale and insight than ever before. From **faster evidence synthesis to more sophisticated predictive modeling and transparent decision models**, the benefits are multifaceted. As one report noted, generative AI and similar technologies *"hold promise for transforming HTA by improving efficiency and accuracy in evidence generation"*, provided we apply **careful oversight and clear guidelines** as this field evolves [46] . For pharma and HEOR professionals, the takeaway is to embrace these advanced tools as a means to enhance – not supplant – the expert-driven processes that underpin HTA. By doing so responsibly, we can achieve **richer, more robust assessments** that ultimately lead to better-informed healthcare decisions and improved patient outcomes. The future of HTA is not just about doing the same assessments faster; it's about doing them smarter, with AI as a valuable partner in our quest for value in healthcare.

Sources:

1. Singh B. *et al.* (2025). *Simulating Sponsor-HTA Conversations: A Multi-Agent GenAI Framework for Reimbursement Processes* [1] [2] . *ISPOR Europe 2025 Presentation.*
2. Reason B. *et al.* (2024). *Can Large Language Models Simulate HTA Committee Discussions?* (Case study using GPT-4 for NICE appraisal) [5] [6] . *ISPOR Presentation.*
3. Versteeg JW. *et al.* (2025). *AI-Assisted Data Extraction from HTA Reports: Improving Comparative Research* [10] [12] . *ISPOR Europe 2025 Presentation.*
4. Ayer T. *et al.* (2024). *Generative AI to Identify Comparative Critiques in HTA Reports (ValueGen.AI case study)* [16] [17] . *Poster Presentation.*
5. Wang H-I. *et al.* (2025). *Predicting EQ-5D Index Scores: ML vs Statistical Methods* [24] . *ISPOR Europe 2025 Presentation.*
6. Srivastava T. *et al.* (2025). *Uncovering Clinical Nuance in Simplified HE Models through AI (Hybrid Intelligence)* [29] [32] . *ISPOR Europe 2025 Poster.*
7. ISPOR Working Group (Fleurence RL *et al*.) (2024). *Generative AI for HTA: Opportunities, Challenges, and Policy* [33] [42] . *Value in Health 28(2) Plain Language Summary.*
8. NICE (2024). *Position Statement on Use of AI in Evidence Generation* (summarized by BeaconOne) [34] [36] .

---

[1] [2] [3] [4] ISPOR - Simulating Sponsor-HTA Conversations: A Multi-Agent GenAI Framework for Reimbursement Processes
https://www.ispor.org/heor-resources/presentations-database/presentation-cti/ispor-europe-2025/artificial-intelligence-and-advanced-analytics-in-hta/simulating-sponsor-hta-conversations-a-multi-agent-genai-framework-for-reimbursement-processes

[5] [6] [7] [8] Welcome to the Bristol Myers Squibb PowerPoint template
https://www.ispor.org/docs/default-source/intl2024/ispor24reasonp46presentation138695-pdf.pdf?sfvrsn=a1ac4f14_0

[9] [10] [11] [12] [13] [35] ISPOR - Can AI-Assisted Data Extraction From HTA Reports Improve Comparative HTA Research: A Case Study on NICE Assessment Reports
https://www.ispor.org/heor-resources/presentations-database/presentation-cti/ispor-europe-2025/artificial-intelligence-and-advanced-analytics-in-hta/can-ai-assisted-data-extraction-from-hta-reports-improve-comparative-hta-research-a-case-study-on-nice-assessment-reports

[14] [15] [16] [17] [18] [19] [20] valueanalyticslabs.com
https://www.valueanalyticslabs.com/wp-content/uploads/2024/12/Generative-AI-to-Identify-Comparative-Critiques-in-HTA-Reports-from-Multiple-Countries.pdf

[21] [22] [23] [24] [25] ISPOR - Predicting EQ-5D Index Scores: A Comparison Study of Machine Learning and Statistical Methods on Health Survey for England Data
https://www.ispor.org/heor-resources/presentations-database/presentation-cti/ispor-europe-2025/artificial-intelligence-and-advanced-analytics-in-hta/predicting-eq-5d-index-scores-a-comparison-study-of-machine-learning-and-statistical-methods-on-health-survey-for-england-data

[26] [27] [28] [29] [30] [31] [32] Program - ISPOR Europe 2025
https://www.ispor.org/conferences-education/conferences/upcoming-conferences/ispor-europe-2025/program/program

33 41 42 43 44 46 Generative Artificial Intelligence for Health Technology Assessment: Opportunities, Challenges, and Policy Considerations: An ISPOR Working Group Report
https://www.ispor.org/publications/journals/value-in-health/plain-language-summary/Volume-28--Issue-2/Generative-Artificial-Intelligence-for-Health-Technology-Assessment--Opportunities--Challenges--and-Policy-Considerations--An-ISPOR-Working-Group-Report

34 36 37 38 39 40 45 NICE Opens Door to Use of AI in HTA Submission - BeaconOne Healthcare Partners
https://beacononehcp.com/2025/02/11/nice-opens-door-to-use-of-ai-in-hta-submission/