Journal of Medical Economics

ISSN: 1369-6998 (Print) 1941-837X (Online) Journal homepage: www.tandfonline.com/journals/ijme20

Artificial intelligence in key pricing, reimbursement, and market access (PRMA) processes: better, faster, cheaper—can you really pick two?

Eva Susanne Dietrich

To cite this article: Eva Susanne Dietrich (2025) Artificial intelligence in key pricing, reimbursement, and market access (PRMA) processes: better, faster, cheaper—can you really pick two?, Journal of Medical Economics, 28:1, 586-595, DOI: 10.1080/13696998.2025.2488154

To link to this article: https://doi.org/10.1080/13696998.2025.2488154

Published online: 17 Apr 2025.

JOURNAL OF MEDICAL ECONOMICS
2025, VOL. 28, NO. 1, 586–595
https://doi.org/10.1080/13696998.2025.2488154
Article /2488154

Taylor & Francis
Taylor & Francis Group

PERSPECTIVE

Artificial intelligence in key pricing, reimbursement, and market access (PRMA) processes: better, faster, cheaper—can you really pick two?

Eva Susanne Dietrich

Institute for Evidence-based Positioning in the Healthcare Sector, Bonn, Germany

ABSTRACT

The rapid evolution of large language models (LLMs) and machine learning (ML) presents both significant opportunities and challenges for market access processes. These sophisticated AI systems, built on transformer architectures and extensive datasets, offer potential to forecast claims and decisions of health technology assessment (HTA) agencies and streamline processes, such as systematic literature reviews and HTA submissions. Furthermore, the analysis of real-world data—also for deriving causal relationships—is being discussed intensively. Despite notable advancements, their adoption in key PRMA processes is still limited at present, with only a small fraction of submissions to HTA bodies incorporating AI. Key barriers include stringent transparency requirements, the necessity of explainability and human oversight in data analyses, and the highly sensitive nature of text drafting—especially in cases where reimbursement decisions or pricing negotiations balance on a knife's edge. These requirements are often not met due to the immaturity of many AI applications, which still lack the necessary precision, reliability, and contextual understanding. Moreover, AI-generated evidence has yet to prove its validity before it can supplement or replace traditional study designs, such as randomized controlled trials (RCTs), which are critical for HTA decisions. Additionally, the environmental and financial costs of training LLMs require careful assessment. This paper explores various current AI applications, their limitations, and future prospects in key PRMA processes from a German perspective while also considering the broader implications of the EU Health Technology Assessment Regulation (HTAR). It concludes that while AI hold transformative potential, its integration into workflows must be approached cautiously, with incremental adoption, and close collaboration between industry, HTA agencies, and academia. Demonstrating robust, unbiased comparative evidence—showcasing superior performance and cost savings over traditional methods—could accelerate the adoption process.

ARTICLE HISTORY
Received 22 January 2025
Revised 29 March 2025
Accepted 31 March 2025

KEYWORDS
Artificial intelligence; drugs; pricing; reimbursement; market access; large language models; LLM; health technology assessment; HTA; AMNOG

JEL CLASSIFICATION CODES
L65; I18; H51

Introduction

Research and experimentation with artificial intelligence (AI) in healthcare date back to the 1960s, starting with ELIZA, a natural language communication program developed by J. Weizenbaum[1]. Early AI programs emerged that were structurally similar to those of today. Following several periods of stagnation, often referred to as "AI winters"[2], the last five years have witnessed a remarkable resurgence of AI enthusiasm. Today, AI is an omnipresent topic at congresses, training programs, and within pharmaceutical companies, consultancies, and HTA agencies.

Attractive market

One of the key drivers of the AI hype is the profitability of the healthcare market. Over the past few years, healthcare has emerged as one of the fastest-growing industries worldwide[3]. At the same time, AI holds the promise of improving efficiency, thereby further increasing profit margins in this already attractive sector. Additionally, many computer scientists, programmers, and software developers are now exploring the opportunities in this market. Among them are experts in computational linguistics, a field that saw significant research during the 1960s, 1970s, and 1980s, as well as graduates in business and economics.

In 2023, health startups ranked among the top three sectors for venture capital funding in Europe, alongside energy and transportation[4]. More than half of German startups in 2023 focused on AI, a trend mirrored in countries like France, Italy, Austria, and Spain[5,6]. Furthermore, many of Europe's most promising startup hubs are closely tied to universities[7] fostering collaboration at the intersection of industry and research, particularly in areas, such as information retrieval.

Enormous fascination

Moreover, the fascination with AI is undeniable. Many people recall the magic of their first encounter with tools like ChatGPT, Elicit, or similar AI systems—producing beautifully

CONTACT Eva Susanne Dietrich info@dietrich.healthcare Institute for Evidence-based Positioning in the Healthcare Sector, c/o University of Bonn, Pharmaceutical Institute, Clinical Pharmacy, An der Immenburg 4, Bonn, D-53121, Germany

worded, grammatically flawless texts. However, this initial excitement is often followed by disillusionment when further scrutiny reveals that some statements are hallucinated or that provided links are inaccurate[8–10].

AI in value assessments and pricing of drugs

This raises critical questions: How effective is AI in key pricing, reimbursement, and market access (PRMA) processes now that computational power is no longer a major limitation? What potential applications are envisioned for these models? Is AI already surpassing humans in terms of being better, faster, or cheaper—or even all three? These goals, long thought to be mutually exclusive since NASA's 1992 "Faster, Better, Cheaper" principle, now raise a new question: how many, if any, can AI truly deliver?

Against this backdrop, this paper examines this question from a German, decision-maker-focused perspective, while also broadening the scope to include the EU Health Technology Assessment Regulation (HTAR) and the workflows surrounding dossier submissions from a manufacturer's point of view.

Short methodical introduction

The rapid pace of development in generative AI has led to a diverse landscape of models, each optimized for different tasks. From text generation and image synthesis to code completion and multimodal applications, the range of capabilities continues to expand. Understanding the strengths and limitations of different model types is key to making effective use of these technologies.

LLMs are advanced deep learning models designed to process and generate human language. They are a subset of generative AI, which refers to models capable of generating new content—such as text, images, or music—based on learned patterns from vast amounts of data.

Deep learning, a subset of machine learning, uses structures inspired by the human brain, called neural networks, to analyze large amounts of data and recognize patterns.

Machine learning itself forms the foundation of LLMs and neural language models (NLMs), enabling computers to learn from data and improve their results over time without explicit programming.

Neural networks, which underpin both LLMs and NLMs, consist of layers of interconnected nodes or "neurons" that process data and adjust their behavior based on experience. These networks allow models to identify relationships between inputs, such as relationships between words in sentences and their context, and generate meaningful outputs.

Word embeddings further enhance this capability by representing words in a numerical format that computers can better understand. In these embeddings, words are mapped into a multi-dimensional space where similar or related words are positioned closer together (e.g. "cat" and "dog"), while opposites are farther apart (e.g. "hot" and "cold"). This helps models recognize relationships between words and their meanings, concepts, and typical context.

NLMs are distinct from LLMs in their scale and capabilities. As earlier forms of AI-driven language models trained for specific tasks, they typically use smaller datasets and simpler architectures, which define the design and flow of data through the neural network. This simplicity makes NLMs less capable of handling complex language patterns for not predefined tasks or maintaining coherence over long passages of text.

LLMs, on the other hand, are trained on massive datasets containing billions of tokens. A token can be a word, part of a word, or even a single character, referring to the smallest meaningful unit in a text that is identified during text processing. By analyzing these tokens, LLMs learn to understand and produce language with a high degree of sophistication.

A key innovation that sets LLMs apart is the transformer architecture, which allows them to analyze and remember context from input data more effectively. Transformers enable LLMs to generate text by predicting the most probable next token based on previous input, thereby producing human-like responses. This architecture excels at understanding extended sentence structures or paragraphs by focusing on the relationships between words, even when they are far apart in the text.

The scalability and performance make LLMs suitable for complex tasks, such as reasoning, summarization, translation, and answering questions that require logical reasoning and contextual comprehension. A well-known example of a user interface that connects various LLMs is ChatGPT.

Reasoning models, on the other hand, are AI models specifically designed to support logical thinking, problem-solving, and decision-making. They go beyond merely mimicking language and can draw complex conclusions, for example, through mathematical or causal analyses.

Explainable AI (XAI) plays a crucial role in the development of LLMs, NLMs, and reasoning models. It focuses on making AI models and their results more transparent and understandable, ensuring that users can see why a model made a certain prediction or decision. This is essential for building trust, addressing legal concerns, and improving user confidence in AI systems[11–16].

Overview: potential roles of AI in market access

A wide variety of areas in which AI can be used are being discussed in the PRMA field. Following the chronology of the market access process, these are:

- Forecasting market opportunities and market development
- Replacing control arms or clinical trials through AI-driven analyses
- Preparing core/global value dossiers
- Predicting payer-relevant PICO or PICOS schemes (Population, Intervention, Comparator, Outcomes, Study design), e.g. in the context of the HTAR
- Identifying patterns in appraisal processes
- Estimating chances in the HTA and pricing process

- Compiling HTA submissions for pricing and reimbursement
- Evaluating HTA submissions (by HTA agencies)

For each of these areas, the AI solutions on the market have matured to different degrees, so some of these fields of application are considered separately in the following.

Review of the individual application options

Replacing control arms or clinical trials through AI-driven analyses

According to Bekryl Market Analysis, AI has the potential to save up to 70 billion USD in global drug development costs by 2028[17]. A report by BCG further estimates that AI could reduce the time and cost of drug development by 25–50%, depending on the prior level of knowledge about the target[18].

These savings can, in part, be attributed to AI's ability to enhance the efficiency of identifying active substances. For example, early progress has been demonstrated with an A2a adenosine receptor in solid tumors or the generation of dual-target compounds[19,20].

AI also offers significant potential in the clinical development phases. It can support the intelligent selection of trial participants, analyze individual metabolic profiles to tailor dosages, monitor study patients effectively, optimize clinical trial designs, and identify safety-relevant signals[18,21,22].

Various regulatory authorities are currently assessing AI's potential and developing draft guidance for integrating AI-based models into pivotal trials[22–26]. Key concepts include leveraging digital twins to evaluate the bioactivity, chemical, and pharmacological properties of new drugs, as well as utilizing virtual control arms and big data analytics[24,27].

While these studies will later serve as a foundation for the HTA process, AI-driven approaches are also being explored to mitigate the challenge of diverging HTA requirements regarding PICOS elements, such as comparator therapies, endpoints, and study designs, aiming to reduce the need for multiple tailored studies. Additionally, AI is increasingly being considered as a tool for processing non-randomized real-world evidence in HTA submissions.

However, these approaches are not met with enthusiasm everywhere[28,29]. The former head of the German Institute for Quality and Efficiency in Health Care (IQWiG) described virtual patients as a "grab bag of benefit assessment" designed to bypass randomized controlled trials (RCTs) and cited Mau et al.: "Inadequate preparation only shifts the necessary effort into the future since the lack of informative value must ultimately be compensated for by corresponding verification studies"[30,31].

This comment highlights a broader underlying issue—the ongoing debate over the necessity or potential replacement of RCTs. In Germany and beyond, benefit assessments are often required to adhere to evidence-based medicine standards. Integrating AI-generated insights into this framework may necessitate legislative changes, requiring broad political and stakeholder consensus.

Taking it one step further, even if some jurisdictions might accept low-level evidence from AI models, this does not guarantee that drug prices will meet pharmaceutical companies' expectations. Similar cases with non-randomized evidence have shown that pricing can profoundly diverge from industry hopes. An analysis of non-small cell lung cancer found that HTA evaluations based on single-arm trials with external controls, conducted in Germany, France, the UK, and Canada, led to either reimbursement denials or product withdrawals in ~50% of cases. In the remaining cases, significant rebates of 50–80% were required[32].

Big data-based analyses also face challenges due to the lack of representative datasets. Current datasets often suffer from known and unknown biases, limiting their applicability. Efforts to encourage population-wide participation in data sharing are advancing through legislative measures, such as the European Health Data Space (EHDS), Germany's Health Data Utilization Act (GDNG), and the Digital Act. However, these initiatives face hurdles, including data protection concerns, feasibility issues, and socio-political and environmental objections[33–35]. In Germany, for example, only half of the population is willing to share their health data for research purposes[36]. Moreover, the principle of data minimization, as enshrined in regulations like the General Data Protection Regulation (GDPR), conflicts with the vast data requirements of AI.

In conclusion, replacing clinical trials or control arms with AI-generated data may trigger similar internal deliberation processes among many HTA agencies as the acceptance of non-randomized clinical trials. It is not guaranteed that the specific characteristics of a dataset from one country can be applied to another, ensuring the transferability of results. The most critical aspects, however, are validity and transparency. Validity must still be demonstrated through comparisons with established methods, such as RCTs. Additionally, AI-generated results must be comprehensible, explainable, and transparent, with the underlying data, training datasets, and algorithms being accessible and easy to interpret.

Compiling dossiers

The use of AI tools in dossier preparation—such as conducting systematic reviews, presenting data, and drafting texts—appears to be one of the more viable applications in the near term. However, it is crucial to distinguish between core value dossiers and HTA submissions, despite some overlap in terminology. HTA submissions, particularly those required under the HTAR, must comply with stringent guidelines and strict structural requirements. While these rigid frameworks could facilitate the development of AI systems tailored to HTA submissions, they also demand near-perfection. In contrast, for core value dossiers or routine texts, the Pareto principle may apply, allowing for significant time savings with an optimized level of effort. However, the preparation of dossiers, such as AMNOG submissions requires more than just high accuracy—it demands meticulously crafted and

strategically framed medical writing that precisely aligns with the mindset and decision-making framework of HTA agencies, especially in cases where the added benefit is not immediately evident.

As core value dossiers primarily serve as internal sources for various external purposes, they often do not require the same level of completeness and well-calibrated wording. This makes them a promising area for AI applications.

Regardless of the intended use, significant adaptations to existing models or the development of bespoke in-house solutions are necessary to handle the specific demands of these tasks[37].

Forecasting a PICOS scheme

Defining the PICOS requirements in national HTA processes or in the joint clinical assessment (JCA) is an essential step for preparing dossiers. In view of the significantly increased complexity and shortened timelines in centralized assessment, numerous companies have now developed AI to predict the differing PICOs of the scoping process. Exercises are partially presented. However, validations are still rare. Comparisons with manual scoping show that a lack of well-trained and robust AI tools still makes human involvement necessary[38]. As a consequence, consulting companies offer packages in which AI-generated PICOs are validated through surveys conducted with international experts.

Expert panels and consultations with relevant authorities and HTA agencies will therefore remain indispensable in defining and refining these criteria. However, it is questionable whether combining AI with expert interviews actually saves time and money compared to expert surveys alone.

Performing a systematic search

An examination of Prospero (https://www.crd.york.ac.uk/prospero/) reveals that new literature mapping tools and mega search engines are already being used sporadically for systematic searches[39]. However, their application remains selective and typically serves as a supplement rather than a replacement[39].

A Cochrane review of 196 reports on the use of ML and LLMs in systematic reviews in health research from February 2024 showed that LLMs were applied in 10 out of 13 review steps, most frequently for literature search (41%), study selection (38%), and data extraction (30%), with GPT being the most commonly used model. While LLM support in study selection and data extraction was largely considered promising, their application in literature search proved less reliable, with slightly more than half of the approaches rated as not promising. LLMs were rarely used for writing plain language summaries ($n = 1$) and publications ($n = 2$) or for quantitative analysis ($n = 2$). Notable gaps were seen in their use for deduplication, full-text retrieval, and assessing the certainty of evidence using the GRADE approach[40].

Scope of systematic reviews. The usefulness of AI in searching for relevant data largely depends on whether HTA agencies rely primarily on randomized evidence like Germany or France[41,42], or adopt a broader range of available data for assessments[43]. While RCTs can be quickly and easily identified using traditional methods, AI might be particularly helpful in identifying supplementary data. The larger the volume of data, the more complex the review, and the less experienced the review team, the greater the potential benefit in terms of time savings[44,45].

Developing a search strategy. However, broader search strategies designed to increase sensitivity often result in reduced efficiency gains, as they still require substantial human effort for screening. As the tools improved in identifying relevant articles, they also labeled more irrelevant ones as relevant, ultimately reducing time savings[44]. Furthermore, it is crucial to distinguish between searching for clinical evidence and searching for information on a disease or unmet need, as the completeness requirements differ. In the latter case, maintaining a fully comprehensive database and achieving 100% sensitivity may not be essential.

Screening hits. When screening search results, the volume of hits plays an important role. In German AMNOG dossiers, the number of hits identified in the relevant databases is typically in the single or double digits, allowing for an efficient and manageable screening process. If HTA agencies extend their scope beyond randomized and non-randomized comparative studies, the required effort might increase significantly.

Open-source tools that can differentiate RCTs from non-RCTs are already available and effective[46]. However, incorporating additional inclusion and exclusion criteria is often still a considerable challenge[47,48]. Reported AI accuracy in abstract reviews varies widely[49], ranging from 10% (for humanistic outcomes) to 99%[47,48,50–56]. From the perspective of the HTA agencies, the sensitivity of the full-text searches is ultimately the most important factor. All relevant studies must be identified and analyzed (100% sensitivity). An accuracy or sensitivity of $<50\%$ are therefore completely unacceptable[48,54] and human intervention remains essential to ensure completeness of the evidence base.

Furthermore, AI models must be trained, and while dual screening involving human reviewers and AI-assisted tools can moderately save time, accuracy decreases when the training set includes only a few eligible studies[48].

Documentation and transparency. Good scientific practice dictates that search and selection processes be meticulously documented and reproducible. In one project using custom classifiers to distinguish between human and animal studies, the British National Institute for Health and Care Excellence (NICE) found that "black box" AI elements conflicted with their transparency requirements. While pattern matching *via* GitHub was transparent, it introduced substantial manual effort and content challenges[57].

Data extraction and analysis. Data extraction is another area where AI faces limitations, especially when extracting data from tables[8]. While extracting overall survival (OS) data from publications has been achieved, the extraction of more

complex data—such as quality of life measures, specific domains, observation times, and population-specific results—remains challenging as a glance at solutions currently on offer shows (as of March 2025).

For analyzing and presenting original datasets, established software tools exist. However, the primary effort continues to lie in quality control, which still requires human oversight—both for the analysis of individual patient data and for the extraction of data from publications.

Interim conclusion on the use of LLMs in systematic reviews. So further advancements are needed before LLMs can fully automate systematic literature reviews[54,58,59]. Siw Waffenschmidt, head of the Information Management Department at IQWiG, concluded after a workshop with international information retrieval experts in 2024: "Do we have to throw away lexical search? Probably not"[60]. According to Waffenschmidt, "it is still a long way until GPT will be able to develop a full search strategy".

Medical writing

When it comes to writing texts, generative AI can be effective for creating simple summaries—such as outlining results, unmet needs, or general disease information. For example, AI might generate a core value dossier that includes figures on disease incidence, a brief discussion of disease burden with supporting references, and key guidelines—all easily sourced from the internet. A study found that AI system-generated summaries achieved informativeness scores that were relatively close to handwritten summaries. However, further refinement was needed to enhance the readability and coherence of the generated summaries[61].

Furthermore, major limitations are centered around AI's low capacity for critical evaluation and inaccuracy in source retrieval[62]. For instance, it will have problems (without extensive human input) to describe with robust figures the burden of disease and quantify the number of e.g. myeloma patients with relapsed or refractory myeloma in a country after at least three prior therapies and to underpin it with complete and relevant literature from primary sources. In such cases, retrieval-augmented generation (RAG) approaches might work by combining database searches, epidemiological modeling, and human refinement of AI-generated text.

Moreover, AI lacks the capability to organize content in a way that persuades payers of the urgency or necessity of a new therapy. Current AI has to be further trained and refined to better understand the nuances of HTA, to ensure their adaptability to different therapeutic areas and complex datasets and to take variations in HTA requirements across different countries into account. This task often requires input from external experts and key opinion leaders to refine and improve draft texts by experienced company employees. As Thomas Kaiser, head of IQWiG, noted with a touch of provocation: "Texts generated by AI are boring"[60].

Also in medical writing, the Pareto principle applies: When preparing a dossier, the 20% make the difference between a filled-in module template and a good dossier—between what an AI or a team of dedicated junior medical writers compiles and what is needed to convince the decision-makers. While further AI training with better selected training data could improve outcomes, this comes at a financial and—not to forget—ecological cost. Experienced professionals will still be required to thoroughly review AI outputs and revise them to meet the high standards in medicine that often exceed expectations in other areas, such as trade and industry[11].

Forecasting the appraisal and pricing processes

When simulating discussions and decisions of HTA agencies, it is crucial to clarify the expectations. AI might simulate typical formulations—such as those found in resolutions or recommendations[63]. It might predict the level of reimbursement[64] or the extent of added benefit[65]. However, predicting the achievable price and the key aspects of commercial arrangements remains a challenge. Still, it is one of the most exciting aspects of market access and a potential showcase field for AI.

In recent years, various automated solutions have been developed to estimate achievable drug prices. Pricing is inherently complex, as it depends on numerous parameters,

Key points.

Key area	Summary
AI in clinical trials	On the regulatory side, further developments, such as digital twins, virtual control arms, and big data analyses are being discussed and tested.
Dossier preparation	AI can support systematic reviews and routine writing tasks, but it may not yet fulfill the rigorous expectations for precision, transparency, and completeness in HTA contexts.
Data extraction limits	AI still struggles with complex data extraction.
PICOS forecasting	AI is not yet reliable enough for forecasting PICOS criteria, making expert validation a necessary additional step.
Systematic searches	The larger the volume of data, the more complex the review and the less experienced the review team, the greater the potential benefit in terms of time savings.
Pricing and reimbursement	AI can simulate reimbursement scenarios, but precise price predictions remain difficult.
Evaluation of HTA submissions	Although AI may assist with selected preparatory tasks in HTA, such as sample-based checks for literature completeness, the core appraisal process itself cannot be delegated to algorithms and continues to require human judgment.
Human expertise	Human experts remain essential for regulatory compliance and contextual evaluation.
Current AI use in HTA	AI adoption in HTA submissions is still very limited.
Environmental impact	High energy consumption raises questions about AI's threats to sustainability.
Future prospects	AI in market access is still in its infancy. AI experts cannot fully replace HTA experts, and HTA experts will need AI expertise in the future.

including drug efficacy, market competition, the market entry price, the number of potential patients, the drug's ability to improve life duration and quality of life, and negotiators' team spirit, soft skills, and form on the day. While AI can assist in identifying patterns, significant challenges remain. Algorithms trained on historical data often lack access to confidential negotiated prices, and the available data is frequently fragmented, requiring substantial manual effort to address gaps. Additionally, the omission of critical parameters—whether intentional or unintentional—can distort price predictions[66].

However, the crucial question is how realistic the prices in the advertised AI solutions and databases are. With the exception of Germany, negotiated prices are confidential in most countries, making referencing in AMNOG price negotiations a significant challenge for both health insurance companies and the German negotiating teams on the manufacturers' side in the past. It is therefore essential to critically question what data is actually stored in the databases. Even consulting firms with extensive experience in submission processes face limitations as strict confidentiality agreements surrounding individual projects may prevent sharing or reusing negotiation data for AI models employed across different projects.

Predicting appraisal outcomes with the help of AI might be easier. In recent years, several methods have been tested in different countries, which show that there is still a need for further development to achieve higher accuracy[63,64,67]. A German product developed to forecast AMNOG assessment results achieved a hit rate of 88%[65]. The developers attributed its reduced accuracy to the limited dataset of only several hundred cases[68]. They noted that human behavior involves too many unknown and unrecorded influencing factors[69]. Other experts emphasize that a lack of modularity in AI systems may hinder their ability to address numerous nuanced cases[37].

In summary, also here, AI applications have yet to demonstrate, through comparative studies, that they surpass expert surveys in terms of accuracy and cost-effectiveness.

Guidance on AI in dossier submissions

NICE has recently published a position paper on the use of AI in evidence generation[70]. However, previous searches reveal, that clear guidance on the use of AI for conducting SLR is lacking in most HTA agencies[71,72]. NICE recommends a screening technique using ML to identify relevant papers earlier. The Scottish Medicines Consortium (SMC) references the NICE methods. The National Centre for Pharmacoeconomics (NCPE; Ireland) only acknowledges the potential of ML algorithms for SLR[71]. IQWiG has taken up the ongoing discussions about AI in its General Methods 7.0, stating that "machine learning approaches (e.g. prioritization, application of classifiers) can be tested and used to support study selection." The method paper also emphasizes that "if available, validated study filters (e.g. for RCTs and systematic reviews) or validated classifiers from machine learning (e.g. RCT classifier) are used"[73].

Evaluating HTA submissions

Humans are still better in understanding the context[57]. As stated by the German Ethics Council in a position paper, human judgment cannot be fully technically replaced. Current AI systems lack the necessary capabilities to comprehend meaning, intentionality, and references to an extra-linguistic reality[12]. Moreover, HTA agencies aim for the most accurate and responsible actions over extended periods while ensuring a coherent practice framework[12]. The German Ethics Council also emphasizes that moral responsibility can only be ascribed to natural persons. This involves evaluating actions as right or wrong based on ethical principles. Therefore, responsibility cannot be directly attributed to machine systems but must lie with the individuals behind these systems in various roles, potentially within an institutional framework[12,14–16,74].

Furthermore, in benefit decisions, the focus seldom lies on questions that can be answered through simple searches, e.g. using retrieval-augmented generation. Instead, expertise is required; as Aaron Tay notes, experts "have deep or at least unusual expertise with ideas or points that aren't available on the web" and "academic search typically requires deep exploration not quick answers"[8].

While it is conceivable that individual steps in the assessment process may be delegated to AI in the future, the overall appraisal process will remain under the purview of human decision-makers in the long term.

Conclusion

Status quo

By the time this article is read, ongoing advancements in AI may have already addressed some of the challenges discussed above. This field is evolving rapidly, and any current snapshot risks becoming outdated within months.

Nonetheless, as of now, AI has made significant strides in fields, such as radiological diagnostics even though some legal and performance-related issues remain unresolved[75].

In contrast, its application in the context of health technology assessment (HTA) is still limited. Recent studies by a large international consulting firm reveal that LLMs are rarely utilized in health technology assessment (HTA) submissions so far. Only 7 out of 5,000 English-language documents submitted to NICE, the Canadian Agency for Drugs and Technologies in Health (CADTH), or the Australian Therapeutic Goods Administration (TGA) indicated that AI may have been employed for support[39]. A broader analysis based on HTA reports published between 2012 and 2023 identified 11 reports (four from the UK, six from Canada, one from Italy) where AI/ML methods were mentioned[76].

One likely explanation is that, while many providers offer AI-based solutions, the technologies themselves are still immature and rely on practical projects to mature further. Pharmaceutical companies, in turn, appear open to pilot projects. At the very least, some report off the record that they are experimenting with test runs.

Claims

The list of requirements that official authorities impose on AI for market access is extensive. In contrast to consumer-facing AI applications like apps or entertainment tools, AI systems used in market access for healthcare operate within a highly regulated environment with significant legal and ethical considerations. These applications have the potential to directly impact the lives of millions of people, necessitating a higher level of scrutiny and stricter requirements for safety, efficacy, and ethical use[14,16].

Therefore, capable and informed humans should always be involved[59,70]. To address this, a societal debate is needed to determine when "human action" is essential, when "human oversight" (weak AI) is sufficient, and when autonomous action (strong AI) may be appropriate. This discussion is critical for shaping AI regulations and ethical considerations[13].

Furthermore, decisions based on AI systems must be both comprehensible and explainable. Moreover, the processes and reasoning behind these decisions must be traceable[11,70,77]. Key aspects, such as transparency, verifiability, explainability, and controllability are also focus areas of ISO SC 42, a subcommittee of the International Organization for Standardization. Despite being established already in 2017, this committee had not yet published definitive results by the time of writing this text[78,79].

In addition, AI systems must avoid discriminatory outcomes. Since learning systems heavily rely on input data, inadequate datasets or flawed conceptual frameworks can lead to biased and discriminatory results[16,77,80].

Providers of AI technologies must substantiate their claims[81]. Performance benchmarks comparing AI outputs to human work already exist to some degree for some potential tasks but are primarily focused on English and Chinese. This creates challenges for languages like German, especially when generating texts in the national language is a requirement[37].

Furthermore, the promised savings often appear to be estimates with an advertising character rather than validated values from systematic measurements. To provide an example, a company promised to reduce the cost of preparing an HTA submission by 70%. However, these projections are based on an assumed workload of 4,500–5,500 h, when experienced consultancies often complete such tasks in <2,000 h. Yao et al. found in a systematic review of AI-based automation tools, that the following times were not reported: "the investigators needed to be trained on how to use an AI tool if they use the tool for the first time, the time to prepare the training datasets for AI tool, and the time to summarize the results from the training sets and the final screening by the AI tool (including both title and abstract screening stage and full-text screening stage)"[44]. This places the advertised savings into perspective and highlights potential overestimations in such claims.

And finally, in times of climate crisis and the Corporate Sustainability Reporting Directive (CSRD—2022/2464), an essential point should not be lost sight of, which will also be addressed by the European Commission in the future AI regulation[82]: the enormous energy consumption of AI. Google reported that training its PaLM language model consumed energy equivalent to that used by 300 U.S. households in a year[83] and every ChatGPT request costs ten times as much energy as a Google search[84]. This also highlights the importance of addressing the environmental impact of training and using LLMs[34].

Promising fields of application

In drug discovery, AI may simplify and enhance efficiency by identifying and validating new targets for therapeutic development. Furthermore, AI presents a significant opportunity to assist companies in identifying patient populations that could benefit from new therapies. By leveraging trial data and other sources, AI has the potential to streamline the process of pinpointing eligible groups[85].

As an initial application in HTA submissions, AI could be used to tackle existing challenges in traditional study evaluations rather than attempting to replace entire RCTs with AI-driven data analyses. Issues, such as handling missing data require practical solutions. However, even in this preliminary approach, collaboration with statisticians and evidence-based medicine (EBM) experts from HTA agencies is essential to ensure methodological rigor and regulatory acceptance.

AI systems could also generate frequently asked questions (FAQs) to aid in preparing for hearings and discussions with decision-makers and payers.

When it comes to pricing, AI offers some opportunities to estimate potential price ranges—at least until pricing becomes entirely confidential, as requested by some pharmaceutical manufacturers.

Future prospects

Looking ahead, the application of AI in market access will likely rely on a trial-and-error approach. Ultimately, the decision to implement AI will not rest with technology developers, CEOs, or CTOs, but with the people in the workplace who evaluate and determine what works best for their teams[33].

In the end, good and reasonably priced technologies are the ones that will prevail. Reflecting on the 1990s, references in publications were entered manually. Today, we use affordable—or even free—reference management tools that can automatically download PDFs from the Internet at the click of a button. These tools simplify quality control, minimize errors, save time, and ensure interoperability.

However, considering the limitations of current language models—such as issues of scientific validity and reliability, risk of bias, regulatory and ethical considerations, appropriateness of datasets, and generally lack of mature tools[22,59,80,85]—it remains uncertain whether we are on the cusp of another "AI winter" or whether a real breakthrough in the application of AI in market access is just around the corner.

Transparency

Declaration of funding

No funding was received to produce this article.

Declaration of financial/other relationships

No potential conflict of interest was reported by the author.

A reviewer on this manuscript has disclosed that they are an advisor for AminoChain LLC. Peer reviewers on this manuscript have no other relevant financial relationships or otherwise to disclose.

Acknowledgements

The author thanks Reinhold A. Mainz for reviewing the manuscript.

References

[1] Bergemann R. The evolution of AI in evidence generation: exploring levels of automation and explainability: World Evidence, Pricing and Access Congress. Amsterdam; 2024.

[2] Floridi L. AI and its new winter: from myths to realities. Philos Technol. 2020;33(1):1–3. doi: 10.1007/s13347-020-00396-6.

[3] Demidenko O. Top industries for startups [cited 2025 Mar 29]. Available from: https://geomotiv.com/blog/top-industries-for-start-ups-in-2022-2023/

[4] Dealroom.co. Europe; 2024 [cited 2025 Mar 29]. Available from: https://dealroom.co/guides/europe

[5] Kollmann T, Hirschfeld A, Gilde J, et al. Deutscher-Startup-Monitor-2023. Berlin; 2023 [cited 2025 Mar 29]. Available from: https://deutscherstartupmonitor.de/wp-content/uploads/2023/09/Deutscher-Startup-Monitor-2023.pdf

[6] Brother DE. Wir haben die Start-up-Zentren Europas ausgewertet; 2023 [cited 2025 Mar 29]. Available from: https://www.brother.de/ecopro/start-up-zentren-europas

[7] Thornhill J. Top start-up hubs reinvent the model for European tech sector needs. Financial Times; 2024 [cited 2025 Mar 29]. Available from: https://www.ft.com/content/d30a4ebc-21e5-4c49-9ac0-b930e87c30ac

[8] Tay A. How Q&A systems based on large language models (e.g. GPT4) will change things if they become the dominant search paradigm—9 implications for libraries | Aaron Tay's Musings about librarianship; 2023 [cited 2025 Mar 29]. Available from: https://musingsaboutlibrarianship.blogspot.com/2023/03/how-q-systems-based-on-large-language.html

[9] Liu NF, Zhang T, Liang P. Evaluating verifiability in generative search engines. Stanford; 2023 [cited 2025 Mar 29]. Available from: http://arxiv.org/pdf/2304.09848

[10] Pride D, Cancellieri M, Knoth P, et al. CORE-GPT: combining open access research and large language models for credible, trustworthy question answering. In: Alonso O, Cousijn H, Silvello G, editors. Linking theory and practice of digital libraries. Cham (Switzerland): Springer Nature; 2023. p. 146–159.

[11] Bundesregierung. Die Stellungnahmen zu Handlungsfeld 1 "Forschung in Deutschland und Europa stärken, um Innovationstreiber zu sein". Berlin; 2018.

[12] Deutscher Ethikrat. Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz. Berlin; 2023 [cited 2025 Mar 29]. Available from: https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf

[13] DIN e.V., DKE German Commission for Electrical, Electronic and Information Technologies of DIN and VDE. German standardization road map on artificial intelligence. Berlin; Frankfurt; 2020 [cited 2025 Mar 29]. Available from: https://www.dke.de/resource/blob/2017010/99bc6d952073ca88f52c0ae4a8c351a8/nr-ki-english—download-data.pdf

[14] European Union. Regulation—EU—2024/1689—EN—EUR-Lex; 2024 [cited 2025 Mar 29]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

[15] Organisation for Economic Co-operation and Development. OECD AI principles overview; 2024 [cited 2025 Mar 29]. Available from: https://oecd.ai/en/ai-principles

[16] United Nations Educational, Scientific and Cultural Organization. Ethics of artificial intelligence; 2025 [cited 2025 Mar 29]. Available from: https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

[17] Bekryl Market Analysts. Artificial intelligence in drug discovery market size analysis and opportunity, 2018–2028; 2018 [cited 2025 Mar 29]. Available from: https://bekryl.com/industry-trends/ai-artificial-intelligence-in-drug-discovery-market-size-analysis

[18] Budde K. Arzneimittel mit KI entwickeln: von der Idee bis zur Zulassung Anwendungen, Potenziale und Herausforderungen. München; 2024 [cited 2025 Mar 29]. Available from: https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG6_WP_Wirkstoffforschung_2024.pdf

[19] Evotec. Pressemitteilung: evotec und Excientia geben klinische Studie mit neuartigem Immunonkologie-Wirkstoff bekannt. Hamburg; 2021 [cited 2025 Mar 29]. Available from: https://www.evotec.com/uploads/download-files/eb3a903bce483c6fd3b422071b2be2e8.pdf

[20] Srinivasan S, Bajorath J. Generation of dual-target compounds using a transformer chemical language model. Cell Rep Phys Sci. 2024;5(11):102255. doi: 10.1016/j.xcrp.2024.102255.

[21] Liddicoat JE, Lenarczyk G, Aboy M, et al. A policy framework for leveraging generative AI to address enduring challenges in clinical trials. NPJ Digit Med. 2025;8(1):33. doi: 10.1038/s41746-025-01440-5.

[22] International Coalition of Medicines Regulatory Authorities. Horizon scanning assessment report—artificial intelligence. Amsterdam; 2021 [cited 2025 Mar 29]. Available from: https://www.icmra.info/drupal/sites/default/files/2021-08/horizon_scanning_report_artificial_intelligence.pdf

[23] U.S. Food and Drug Administration. AI Initiatives at NCTR. NaN [cited 2025 Mar 29]. Available from: https://www.fda.gov/about-fda/nctr-research-focus-areas/artificial-intelligence

[24] U.S. National Science Foundation. Foundations for Digital Twins as Catalyzers of Biomedical Technological Innovation (FDT-BioTech); 2024 [cited 2025 Mar 29]. Available from: https://www.nsf.gov/funding/opportunities/fdt-biotech-foundations-digital-twins-catalyzers-biomedical

[25] U.S Food and Drug Administration. Artificial intelligence & medical products: how CBER, CDER, CDRH, and OCP are working together. New Hampshire; 2024 [cited 2025 Mar 29]. Available from: https://www.fda.gov/media/177030/download?attachment

[26] European Medicines Agency. Draft reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle. Amsterdam; 2023 [cited 2025 Mar 29]. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/draft-reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf

[27] D'Amico S, Dall'Olio D, Sala C, et al. Synthetic data generation by artificial intelligence to accelerate research and precision medicine in hematology. J Clin Oncol Clin Cancer Inform. 2023;7:e2300021. doi: 10.1200/CCI.23.00021.

[28] Heinz S, Kumari C, Ataide J, et al. P22 Current and future use of RWE in HTA decision-making: payers view globally. Value Health. 2024;27(12):S6. doi: 10.1016/j.jval.2024.10.028.

[29] Heinz S, Bondal E, Kumari C, et al. HTA141 Artificial intelligence: is it a new era in payer decision making? Value Health. 2024;27(12):S381. doi: 10.1016/j.jval.2024.10.1966.

[30] Mau J, Netter P, Nowak H, et al. Biometrische Aspekte der Planung und Durchführung nicht-randomisierter vergleichender klinischer Prüfungen [Biometric aspects of planning and carrying out non-randomized comparative clinical tests]. Dtsch Med Wochenschr. 1986;111(41):1569–1573. doi: 10.1055/s-2008-1068674.

[31] Windeler J. Der digitale Zwilling und andere Wundertüten der Nutzenbewertung; 2022 [cited 2025 Mar 29]. Available from: https://observer-gesundheit.de/der-digitale-zwilling-und-andere-wundertueten-der-nutzenbewertung/
[32] Tomeczkowski J, Dietrich ES. What is an externally controlled trial worth in clinical-and cost-effectiveness HTA systems?: World Evidence, Pricing and Access Congress. Amsterdam; 2024.
[33] Neff G. Künstliche Intelligenz: "Eines der größten sozialen Experimente aller Zeiten"; 2023 [cited 2025 Mar 29]. Available from: https://netzpolitik.org/2023/kuenstliche-intelligenz-eines-der-groessten-sozialen-experimente-aller-zeiten/?via=nl
[34] Bender EM, Gebru T, McMillan-Major A, et al. On the dangers of stochastic parrots. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. Proceedings; 2021 Mar 3–10. Virtual Event Canada; 2021. p. 610–623.
[35] World Health Organization. Regulatory considerations on artificial intelligence for health; 2023 [cited 2025 Mar 29]. Available from: https://iris.who.int/bitstream/handle/10665/373421/9789240078871-eng.pdf?sequence=1
[36] Weisband Y, Schachinger A, Thun S, et al. GIHF-AI Studie 2023. Vertrauen in die Nutzung von Gesundheitsdaten—ein Vergleich zwischen Deutschland und Israel. Berlin; 2024 [cited 2025 Mar 29]. Available from: https://gihf-ai.eu/wp-content/uploads/2024/01/20240118_GIHF-AI_Survey_DE_WEB.pdf
[37] Löser A, Tresp V, Hoffart J, et al. Große Sprachmodelle. München: Grundlagen, Potenziale und Herausforderungen für die Forschung; 2023. doi: 10.48669/pls2023-3.
[38] Jaiswal H, Rolska-Wójcik P, Delaitre-Bonnin C. HTA108 Artificial intelligence tools for PICO prediction: a new reality or a future dream? Value Health. 2024;27(12):S374. doi: 10.1016/j.jval.2024.10.1933.
[39] Halmos T, Health Technology Assessment International. AI augmented HTA submissions and decisions: evolving trends and real-world use cases: evolving trends and real-world use cases; 2023 [cited 2025 Mar 29]. Available from: https://www.youtube.com/watch?v=3B0BgipsyWY
[40] Lieberum J-L, Töws M, Metzendorf M-I, et al. Large language models for conducting systematic reviews: on the rise, but not yet ready for use-a scoping review. J Clin Epidemiol. 2025;181:111746. doi: 10.1016/j.jclinepi.2025.111746.
[41] Heidbrede T, Basic E, Bussilliat P, et al. HTA258 Health technology assessments of single-arm clinical trials in Germany and France. Value Health. 2022;25(12):S347. doi: 10.1016/j.jval.2022.09.1715.
[42] Baschet L, Jarne A, Monnereau M, et al. SA39 Assessment of indirect comparisons by the HAS Transparency Commission (TC) in France: systematic overview and insights. Value Health. 2024;27(12):S621. doi: 10.1016/j.jval.2024.10.3120.
[43] Lambert V, Arondekar B, Doan J, et al. HPR152 A comparison of the acceptability of evidence in the new joint clinical assessment and nine European Union countries. Value Health. 2024;27(12):S305. doi: 10.1016/j.jval.2024.10.1578.
[44] Yao X, Kumar MV, Su E, et al. Evaluating the efficacy of artificial intelligence tools for the automation of systematic reviews in cancer research: a systematic review. Cancer Epidemiol. 2024;88:102511. doi: 10.1016/j.canep.2023.102511.
[45] Rawal A, Ashwworth L, Luedke H, et al. Developing and testing AI-generated PICOS summaries to aid in literature reviews: World Evidence, Pricing and Access Congress. Amsterdam; 2025.
[46] Marshall IJ, Noel-Storr A, Kuiper J, et al. Machine learning for identifying randomized controlled trials: an evaluation and practitioner's guide. Res Synth Methods. 2018;9(4):602–614. doi: 10.1002/jrsm.1287.
[47] Miles G, Giles L, Kerr BC, et al. MSR201 Optimising performance of generative artificial intelligence (GenAI) in systematic literature review (SLR) screening using PICOS criteria. Value Health. 2024;27(12):S478. doi: 10.1016/j.jval.2024.10.2435.
[48] Ng J, Szydlowski N, Gill M, et al. MSR72 An evaluation of the rayyan artificial intelligence tool for systematic literature review screening. Value Health. 2024;27(6):S273. doi: 10.1016/j.jval.2024.03.1505.
[49] Hardy EJ, Jenkins AR, Ross J, et al. MSR229 Accuracy and efficiency of automated or artificial intelligence tools in systematic literature reviews: a systematic literature review. Value Health. 2024;27(12):S483. doi: 10.1016/j.jval.2024.10.2462.
[50] Attri S, Kaur R, Singh B, et al. MSR57 Transforming systematic literature reviews: unleashing the potential of GPT-4: a cutting-edge large language model, to elevate research synthesis. Value Health. 2024;27(6):S270. doi: 10.1016/j.jval.2024.03.1490.
[51] Gultyaev D, Hummel N, Koller L, et al. MSR219 Can artificial intelligence (AI) accurately screen abstracts in systematic literature reviews? Value Health. 2024;27(12):S481. doi: 10.1016/j.jval.2024.10.2452.
[52] Guo E, Gupta M, Deng J, et al. Automated paper screening for clinical reviews using large language models: data analysis study. J Med Internet Res. 2024;26:e48996. doi: 10.2196/48996.
[53] Kaur R, Rai P, Attri S, et al. MSR15 Revolutionizing systematic literature reviews: harnessing the power of large language model (GPT-4) for enhanced research synthesis. Value Health. 2024;27(6):S262. doi: 10.1016/j.jval.2024.03.1448.
[54] Khraisha Q, Put S, Kappenberg J, et al. Can large language models replace humans in systematic reviews? Evaluating GPT-4's efficacy in screening and extracting data from peer-reviewed and grey literature in multiple languages. Res Synth Methods. 2024;15(4):616–626. doi: 10.1002/jrsm.1715.
[55] Sertel A, Samur S, Yildirim I, et al. MSR44 How accurate are large language models for abstract screening in systematic literature reviews? Value Health. 2024;27(6):S267. doi: 10.1016/j.jval.2024.03.1477.
[56] Rai P, Pandey S, Attri S, et al. MSR78 Advancing systematic literature reviews: a comparative analysis of large language models (Claude Sonnet 3.5, Gemini Flash 1.5, and GPT-4) in the automation era of generative AI. Value Health. 2024;27(12):S453. doi: 10.1016/j.jval.2024.10.2312.
[57] Finnegan A. Upskilling and exploring avenues for information specialists in collaboration with data scientists: Information Retrieval Meeting (IRM 2024). Cologne; 2024 [cited 2025 Mar 29]. Available from: https://www.iqwig.de/veranstaltungen/workshop-4-finnegan-upskilling.pdf
[58] Wallace BC, Noel-Storr A, Marshall IJ, et al. Identifying reports of randomized controlled trials (RCTs) via a hybrid machine learning and crowdsourcing approach. J Am Med Inform Assoc. 2017;24(6):1165–1168. doi: 10.1093/jamia/ocx053.
[59] Fleurence RL, Bian J, Wang X, et al. Generative artificial intelligence for health technology assessment: opportunities, challenges, and policy considerations: an ISPOR Working Group Report. Value Health. 2025;28(2):175–183. doi: 10.1016/j.jval.2024.10.3846.
[60] Institute for Quality and Efficiency in Health Care. Information Retrieval Meeting (IRM 2024): verbal statement. Köln; 2024 [cited 2025 Mar 29]. Available from: https://www.iqwig.de/veranstaltungen/information-retrieval-meeting/
[61] Langston O, Ashford B. Automated summarization of multiple document abstracts and contents using large language models. TechRxiv. 2024:1–9. doi: 10.36227/techrxiv.172262754.45577350/v1.
[62] Ai B, Chanwanpen C, Kramer K, et al. AI for science advice: an evaluation for the Government Office for Science. London; 2024 [cited 2025 Mar 29]. Available from: https://www.ucl.ac.uk/steapp/sites/steapp/files/9_final_report_-_ai_for_science_advice.pdf
[63] Reason T, Klijn S, Gimblett A, et al. P46 Can large language models simulate HTA committee discussions? Findings and challenges from a case study in neoadjuvant treatment of resectable non-small cell lung cancer. Value Health. 2024;27(6):S11. doi: 10.1016/j.jval.2024.03.072.
[64] Wang Y, Toumi M, François C, et al. HTA74 Could artificial intelligence support prediction of reimbursement decisions in Scotland? A pilot project. Value Health. 2022;25(7):S517. doi: 10.1016/j.jval.2022.04.1206.
[65] Kauth C, Gantner TD. Der G-BA Algorithmus, geknackt: neue Digitale Welt—Einblicke in die Zukunftswerkstatt der "Künstlichen

Intelligenz" Teil 6; 2023 [cited 2025 Mar 29]. Available from: https://www.con-nexi.de/Der-G-BA-Algorithmus%2C-geknackt__gXja.html

[66] Rauland M. Drug price prediction: where do machine learning and AI stand? 2022 [cited 2025 Mar 29]. Available from: https://pharmaboardroom.com/articles/drug-price-prediction-where-do-machine-learning-and-ai-stand/

[67] Wang Y, Tolley K, Francois C, et al. Machine learning-based models for prediction of innovative medicine reimbursement decisions in Scotland. J Epidemiol Popul Health. 2025;73(1):202802. doi: 10.1016/j.jeph.2024.202802.

[68] Gantner TD. Taking pharmaceuticals market access from guesswork to strategy and beyond: Pharma Trends 2025. Berlin; 2024.

[69] Gantner TD. Daten. Können. Heilen. Heute mutig für morgen. Einführung in das Thema: Jahrestreffen. Köln; 2024 [cited 2025 Mar 29]. Available from: https://youtu.be/losTIFXzmp0

[70] National Institute for Health and Care Excellence. Use of AI in evidence generation: NICE position statement [cited 2025 Mar 29]. Available from: https://www.nice.org.uk/about/what-we-do/our-research-work/use-of-ai-in-evidence-generation-nice-position-statement

[71] Singh-Mangat G, Sharma S, Bergemann R. OP69 are artificial-intelligence-based literature reviews accepted by health technology assessment bodies? Int J Technol Assess Health Care. 2024;40(S1):S30–S31. doi: 10.1017/S0266462324001284.

[72] Rtveladze K, Oikonomidi T, Ferizovic N, et al. HTA284 Use of artificial intelligence and machine learning (AI/ML) in systematic literature reviews (SLR): review of state-of-the-art health-technology assessment (HTA) and future directions. Value Health. 2024;27(12):S410. doi: 10.1016/j.jval.2024.10.2109.

[73] Institute for Quality and Efficiency in Health Care. General methods—version 7.0. Köln [cited 2025 Mar 29]. Available from: https://www.iqwig.de/methoden/general-methods_version-7-0.pdf

[74] Shaw B. Could human HTA committees be replaced with AI systems? 2025 [cited 2025 Mar 29]. Available from: https://pharmaboardroom.com/articles/could-human-hta-committees-be-replaced-with-ai-systems/

[75] Wigge P. KI-basierte Medizinprodukte in der Radiologie; 2023 [cited 2025 Mar 29]. Available from: https://gesundheitsrecht.blog/wp-content/uploads/2023/08/Wigge_Gesundheitsrecht.blog-Nr.-29-2023.pdf

[76] Szawara P, Zlateva J, Kotseva F, et al. HTA364 Can artificial intelligence and machine learning be used to demonstrate the value of a technology for HTA decision-making? Value Health. 2023;26(12):S390. doi: 10.1016/j.jval.2023.09.2047.

[77] Datenschutzkonferenz. Hambacher Erklärung zur Künstlichen Intelligenz: Entschließung der 97. Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder. Hambacher Schloss; 2019. [cited 2025 Mar 29]. Available from: https://www.datenschutzkonferenz-online.de/media/en/20190405_hambacher_erklaerung.pdf

[78] Zielke T. Is artificial intelligence ready for standardization? Systems, Software and Services Process Improvement: 27th European Conference, EuroSPI 2020. Proceedings; 2010 Sep 9–11; Düsseldorf, Germany; 2020. p. 259–274.

[79] International Organization for Standardization. ISO/IEC JTC 1/SC 42—artificial intelligence; 2024 [cited 2025 Mar 29]. Available from: https://www.iso.org/committee/6794475.html

[80] World Health Organization. Ethics and governance of artificial intelligence for health: large multi-modal models. WHO guidance. Geneva: World Health Organization; 2024.

[81] Wornow M, Fries J, Shah N. Large language models in healthcare: are we there yet? Stanford HAI; 2024 [cited 2025 Mar 29]. Available from: https://hai.stanford.edu/news/large-language-models-healthcare-are-we-there-yet

[82] European Commission. Artificial intelligence—questions and answers; 2024 [cited 2025 Mar 29]. Available from: https://ec.europa.eu/commission/presscorner/api/files/document/print/en/qanda_21_1683/QANDA_21_1683_DE.pdf

[83] Ananthaswamy A. In AI, is bigger always better? Nature. 2023;615(7951):202–205. doi: 10.1038/d41586-023-00641-w.

[84] Arbeitgebervereinigung für Unternehmen aus dem Bereich EDV und Kommunikationstechnologie e.V. ChatGPT-Stromverbrauch: das 10-fache von Google—AGEV; 2024 [cited 2025 Mar 29]. Available from: https://www.agev.de/nachrichten/it-telekommunikation/chatgpt-stromverbrauch-das-10-fache-von-google/

[85] Wellcome trust, Boston Consulting Group. Unlocking the potential of AI in drug discovery—current status, barriers and future opportunities. London; 2023 [cited 2025 Mar 29]. Available from: https://web-assets.bcg.com/86/e5/19d29e2246c7935e179db8257dd5/unlocking-the-potential-of-ai-in-drug-discovery-vf.pdf