

# Collaborative framework on responsible AI in LLM-driven CDSS for precision oncology leveraging real-world patient data

Received: 21 June 2025

Accepted: 5 November 2025

Cite this article as: Mathes, S., Ferber, D., Dreyer, T. *et al.* Collaborative framework on responsible AI in LLM-driven CDSS for precision oncology leveraging real-world patient data. *npj Precis. Onc.* (2025). <https://doi.org/10.1038/s41698-025-01180-5>

Sonja Mathes, Dyke Ferber, Tobias Dreyer, Kai J. Borm, Luise Modersohn, Theresa Willem, Richard Dirven, Julien Vibert, Simon Kreutzfeldt, Raquel Perez-Lopez, Arsela Prelaj, Fredrik Strand, Richard D. Baird, Martin Boeker, Jakob Nikolas Kather, Maximilian Tschochohei & Jacqueline Lammert

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

# Collaborative Framework on Responsible AI in LLM-driven CDSS for Precision Oncology leveraging Real-World Patient Data

## Authors

Sonja Mathes<sup>1,2</sup>, Dyke Ferber<sup>3,4</sup>, Tobias Dreyer<sup>6,7</sup>, Kai J. Borm<sup>8,9</sup>, Luise Modersohn<sup>10</sup>, Theresa Willem<sup>11</sup>, Richard Dirven<sup>12,13,14</sup>, Julien Vibert<sup>15,14</sup>, Simon Kreutzfeldt<sup>16,14</sup>, Raquel Perez-Lopez<sup>17,14</sup>, Arsela Prelaj<sup>18,14</sup>, Fredrik Strand<sup>19,20</sup>, Richard D. Baird<sup>21,14</sup>, Martin Boeker<sup>10</sup>, Jakob Nikolas Kather<sup>3,4,5,14</sup>, Maximilian Tschochohei<sup>10, 23</sup>, Jacqueline Lammert<sup>6,7,10, 24</sup>

1 Technical University of Munich, School of Medicine and Health, Department of Dermatology, Munich, Germany

2 Institute for History, Theory and Ethics of Medicine, University of Mainz Medical Center, Mainz, Germany

3 Else Kroener Fresenius Center for Digital Health, Technical University Dresden, Dresden, Germany

4 National Center for Tumor Diseases, Heidelberg University Hospital, Heidelberg, Germany; Department of Medical Oncology, Heidelberg University Hospital, Heidelberg, Germany

5 Department of Medicine I, University Hospital Dresden, Dresden, Germany

6 Technical University of Munich, School of Medicine and Health, Department of Gynecology and Center for Hereditary Breast and Ovarian Cancer, Munich, Germany

7 German Cancer Consortium (DKTK), partner site Munich, a partnership between DKFZ and University Hospital of the Technical University of Munich (TUM), Germany

8 Technical University of Munich, School of Medicine and Health, Department of Radiation Oncology, Munich, Germany

9 Bavarian Center for Cancer Research (BZKF), Technical University of Munich partner site, Germany

10 Technical University of Munich, School of Medicine and Health, Institute of Artificial Intelligence in Medicine and Healthcare, Munich, Germany

11 Technical University of Munich, School of Medicine and Health, Institute for History, Theory and Ethics of Medicine, School of Medicine and Health, Munich, Germany

12 Department of Head and Neck Surgery and Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands

13 Department of Otorhinolaryngology and Head and Neck Surgery, Radboud University Medical Center, Nijmegen, the Netherlands

14 Cancer Core Europe, VILLEJUIF, France

15 Drug Development Department, Gustave Roussy, DITEP, F-94805, VILLEJUIF France

16 Division of Translational Medical Oncology, German Cancer Research Center (DKFZ), and NCT Heidelberg, Heidelberg, Germany

17 Vall d'Hebron Institute of Oncology (VHIO), Radiomics Group, Barcelona, Spain

18 Department of Medical Oncology, AI-ON-Lab, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy

19 Karolinska Institutet, Department of Oncology and Pathology, 171 77 Solna

20 Karolinska University Hospital, Medical Diagnostics Karolinska, 171 76 Solna

21 Cancer Research UK Cambridge Centre, Cambridge, United Kingdom  
22 Department of Medicine I, University Hospital Dresden, Dresden, Germany  
23 Google Cloud, Munich, Germany  
24 Center for Personalized Medicine (ZPM), University Hospital of the Technical University of Munich (TUM), Munich, Germany

## Correspondence to:

Dr. med. Sonja Mathes  
Biedersteiner Straße 29  
80802 München  
E-Mail: [sonja.mathes@mri.tum.de](mailto:sonja.mathes@mri.tum.de)

## Abstract

Precision oncology leverages real-world data, essential for identifying biomarkers and therapies. Large language models (LLMs) can aid at structuring unstructured data, overcoming current bottlenecks in precision oncology. We propose a framework for responsible LLM integration into precision oncology, co-developed by multidisciplinary experts and supported by Cancer Core Europe. Five thematic dimensions and ten principles for practice are outlined and illustrated through application to uterine carcinosarcoma in a thought experiment.

## Introduction

Large Language Models (LLMs) have transformed human perception of artificial intelligence (AI) primarily due to their ability to respond and reason in human language. The development of user-friendly interfaces has made LLMs accessible to non-technical users. In medicine, these advancements open opportunities, as LLMs can support clinicians by providing quick access to information, assist in clinical decision-making, and can take over simple, repetitive tasks such as documentation, thereby freeing up valuable time for healthcare professionals<sup>1,2</sup>. The capacities of LLMs extend beyond classical AI approaches that primarily analyze or classify structured data, as they combine content generation with the analysis and classification of unstructured data. More recently, multimodal models, which can process images and text simultaneously, have shown impressive results in simulated diagnostic scenarios<sup>3,4</sup>.

However, a recent systematic review found that only 5% of researchers utilize real-world patient data to train and evaluate LLMs, instead, most rely on synthetic datasets, which are easier to access and avoid data privacy concerns. Due to the limited real-world implementation of LLMs in clinical routines, little real-world evidence has been generated to confirm the theoretical potential of LLMs in real-world healthcare settings<sup>5,6</sup>. This current lack of evidence hinders the adoption of LLMs in routine clinical care. To fully unlock their potential for everyday healthcare, validation studies using real-world patient data are crucial to advance both the field and its clinical applications. Navigating the responsible use of real-world patient data for LLMs in complex health contexts benefits from practice-oriented guidelines to avoid uncertainties and navigate the responsible use of such data for LLMs in complex healthcare contexts.

In oncology, the challenges associated with using real-world patient data are particularly complex. Oncology clinical decision-making often requires integrating multimodal data into a comprehensive and individualized treatment plan. Currently, physicians in precision oncology carry out a range of labor-intensive tasks, such as creating and interpreting medical reports, annotating molecular mutations, manually curating scientific evidence, and interpreting biomarker profiles. These time-intensive processes, e.g., of translating genomic data into clinically actionable treatment plans<sup>7</sup>, delay the timely initiation of targeted therapies<sup>8</sup>, and thereby adversely impact patient outcomes.

Clinical Decision Support Systems (CDSS) can leverage LLMs to overcome bottle necks in precision oncology: LLMs can extract structured information from unstructured patient data, enabling clinicians to make faster, more informed decisions (see Figure 1). With this capability LLMs offer a range of applications across the diagnostic and therapeutic process in precision oncology, supporting tasks such as literature review, patient chart reviews, and biomarker interpretation, risk stratification, treatment recommendations and clinical trial matching (Figure 1). These capabilities streamline workflows, enhance efficiency, and improve patient-centered care by minimizing delays in decision-making.

[Figure 1 ]

As LLMs become more prevalent in this field, it is essential to address both the advantages and limitations of using real-world data – an issue increasingly acknowledged by clinicians, patients, and policymakers.

Responsible AI (RAI) frameworks consider the ethical, legal, scientific, and clinical aspects necessary for the ethically sound development and deployment of AI<sup>9,10</sup>. They are essential to ensure the trustworthy application of AI in clinical settings due to the sensitivity of identifying data and the vulnerability of patients. As AI cannot be held accountable for clinical decisions, human operators retain accountability and may be legally liable for acting on AI-generated recommendations, especially when those recommendations do not align with established standard procedures<sup>11</sup>. This is crucial in precision oncology, where off-label treatments are often the only option for rare and treatment resistant conditions lacking established guidelines.

To provide practice-oriented guidance for the responsible use of real-world data in LLMs for precision oncology, we designed a framework of five overarching dimensions to serve as a foundation for practical recommendations. Building on this, we developed ten principles to guide the practical implementation of LLM-based CDSS in precision oncology. We extend this with an example use case of a patient with uterine carcinosarcoma, a rare disease lacking established guidelines, where time-intensive manual research can be significantly accelerated using LLMs to improve patient outcomes. By providing practice-oriented guidance grounded in multidisciplinary experience and research, we aim to ensure that LLM-based CDSS leveraging real-world patient data not only achieve high standards of clinical efficacy but also adhere to ethical and legal requirements, ultimately bridging the gap between AI's theoretical potential and its practical application in healthcare.

## Methods

Our literature review identified the need for specific recommendations on using real-world patient data responsibly in generative AI applications for CDSS in precision oncology. We then conducted an explorative literature review to identify a framework of dimensions for responsible use of LLMs in CDSS in precision oncology. We subsequently held group discussions with an interdisciplinary team of ten experts in clinical oncology, AI, digital health, policy, and ethics, which led to the refinement of the five dimensions. Next, we identified ten practice-oriented principles based on our practical experience and insights from the literature, which should guide practitioners to implement all five dimensions.

To develop the principles, the group allocated each of the principles to a team of two authors according to their expertise. The principles were developed between March 2024 and August 2024. Regular consensus meetings allowed the group to discuss and agree on the content of the text sections. In July 2024, author groups were reassembled to review and edit the manuscript in new teams, providing a second perspective. We contacted Cancer Core Europe for external validation of the manuscript through renowned experts from their network working at the intersection of AI and precision oncology. Cancer Core Europe is a consortium of seven leading cancer centers across Europe, aiming to advance cancer research and improve therapeutic outcomes through collaborative, cross-institutional efforts. After meeting consensus within the team of ten, an additional seven authors from Cancer Core Europe (CCE)<sup>12</sup> reviewed the manuscript between November 2024 and January 2025, revising each section until consensus was reached within the extended international team. To evaluate consensus on the relevance of the thematic dimensions for each principle, eleven available experts from our extended author group participated in an online survey (Google Forms) conducted after the formulation of the ten user-oriented principles. Using a 5-point Likert scale, they rated the extent to which each principle reflected aspects of the five thematic dimensions. The importance of each dimension within the respective principles was evaluated and measured quantitatively and is illustrated as a spider-plot in figure 3. After final review of the finalized text, figures, and tables by all contributing authors,

CCE was asked to endorse the manuscript. We provide a detailed overview of the writing process in Figure 2.

[Figure 2]

## **Five dimensions of responsible AI for using real-world data with LLM for Precision Oncology**

Our five dimensions represent large thematic complexes concerning the responsible use of LLMs with real-world patient data in precision oncology. For each dimension, we briefly summarize relevant concepts and considerations regarding the responsible implementation of generative AI with real-world patient data. We outline the relevance and scope of each dimension and highlight important contextual factors, including applicable regulations, standards, and guidelines.

### **Dimension 1: Adherence to medical standards**

Medical standards are striving to uphold a “standard of care”, as the established treatment that is beneficial for patients and subject to ongoing scientific evaluation<sup>13</sup>. Consensus on medical standards is, in part, summarized in medical guidelines. The constant evolution of medical standards is mirrored in the growing body of literature and the updating of medical guidelines. Generative AI in the form of LLMs holds the potential to contribute to the evolution of these standards by structuring previously inaccessible unstructured information, such as doctors' letters, pathology reports and clinical images for structured evaluation<sup>14</sup> and facilitating the review of recent literature. There can be a bidirectional interaction between evolving medical standards and AI: On the one hand, principles RAI can help ensure the implementation of medical standards in LLMs. On the other hand, LLMs can help advance medical standards, especially, where they are undefined (e.g., in precision oncology). LLM systems must undergo critical evaluation of their adherence to medical standards, ensuring safety, efficacy, and relevance in real-world settings. This includes medical device certification, which ensures that LLMs in precision oncology comply with medical standards and safeguard patient safety.

### **Dimension 2: Balancing technological benefits and risks**

LLMs are able to quickly and efficiently extract information from unstructured text, image, audio and video files, and create new, human-like content. This makes them valuable for managing the exponentially growing volume of unstructured and unlabeled information generated in Electronic Health Records (EHR) each day<sup>15</sup>. This is especially relevant for precision oncology, where analysis of extensive data sources becomes essential for tailoring treatments when standard guidelines fail or are nonexistent.

Technical designs need to carefully balance benefits and risks of new technologies. For LLMs, this includes weighing their ability to analyze unstructured textual data quickly and reliably against shortcomings such as 'hallucinations', that are factually incorrect statements, that are hard to

detect as such<sup>16</sup>. When technology leads to risks, these risks must be acknowledged and addressed<sup>17</sup>, also involving structured assessments, such as context-specific technology impact evaluations<sup>18</sup>.

### Dimension 3: Compliance with regulatory frameworks

Laws and regulations mandate medical standards and provide guidance on handling technology risk. Examples are the Medical Device Regulation (MDR) and the In Vitro Diagnostic Regulation (IVDR) in the European Union, and Federal Drug Administration (FDA) and Health Insurance Portability and Accountability Act (HIPAA) in the United States<sup>19,20</sup>. Additionally, the newly published EU AI Act mandates implementation of controls for AI-products, based on the risk level of their application contexts<sup>21</sup>. Many LLM-based use cases in medicine fall under the "high risk" category as they support clinicians and patients to make healthcare decisions. Any modification or in-house operation of a general purpose LLM can lead to clinics becoming AI system providers<sup>22</sup>. While the AI Act gradually comes into full effect until 2027 business and IT decision makers in healthcare institutions were required to become familiar with the obligations of AI system providers and implement relevant governance frameworks and controls already by August 2026<sup>23</sup>.

### Dimension 4: Respecting ethical considerations

Ethical considerations form another thematic complex, inherently relevant to responsible use of LLMs in precision oncology. Ethical standards and guidelines in medicine give the highest priority to promoting patient well-being<sup>24-28</sup>. Historical examples underscore the importance of critical ethical guidance in research and development at this<sup>29</sup>. In AI for precision oncology, AI-specific novel ethical frameworks are increasingly established<sup>30-33</sup>, while a broad range of ethical principles has already been articulated in the existing literature<sup>31,34</sup>. New ethical considerations and principles add perspectives, while traditional perspectives remain valuable, e.g., reflecting the four principles of biomedical ethics by Beauchamp and Childress<sup>27,32</sup>. Novel near-term and long-term ethical concerns have been named, and including bias, fairness as important considerations, calling for responsible AI<sup>9</sup>. As a new concept developing ethical standards for the creation of responsible AI-systems, RAI has emerged from related scientific fields<sup>35</sup>. To ensure that algorithms comply with ethical requirements, embedding ethicists in development teams has been suggested<sup>36</sup>. In the context of precision oncology, the rapid technological evolution of LLMs continues to raise new ethical questions, especially as the race for advances in AI accelerates. Our team of authors proposes incorporating established ethical views alongside emerging ethical considerations when developing LLMs that utilize real-world patient data.

## Dimension 5: Providing scientific and clinical benefit

The increase of precision in medicine is data-driven, continuously integrating insights from biomedical research and health information. By structuring, synthesizing, validating and analyzing this abundance of information across fragmented data sources, AI can substantially contribute and thereby advance precision medicine across healthcare sectors<sup>37-40</sup> and accelerate scientific innovation<sup>4,41</sup>. Aims to prevent risks of AI in research and clinical applications, like dual-use potentials and data-safety concerns, are mirrored in extensive governance frameworks, such as the GDPR<sup>42</sup>, EU AI Act<sup>21</sup>, or DSA/DMA, partially hindering technological advancement<sup>43</sup>. Additional considerations for clinical usability include social and financial factors, such as fostering public acceptance through explainability and education, and calculating the increasing costs of in-depth testing and refinement for adequate monitoring<sup>43</sup>. These factors are particularly relevant in the healthcare sector, where a substantial portion of funding comes from social systems. While the successful implementation of LLMs in science and clinical routine might eventually yield scientific and clinical benefits, including returns on investment, increased accessibility of precision oncology for patients<sup>44,45</sup> and potentially incentivizing further LLM-centered research, these factors must be carefully weighed.

## Ten principles as practical guidance for responsible use of LLMs in precision oncology

Drawing on the content identified as relevant within the thematic dimensions, we proceed to curate ten user-oriented principles for practical guidance. Each principle presents a concise collection of relevant aspects from the dimensions to aid at the actual application of generative AI with real world patient data in precision oncology. The contribution of the overarching dimensions to each respective principle, as assessed by the experts in our author team, is mapped in Figure 3. To provide relevant practice-oriented guidance we contextualize collections of relevant aspects within current challenges related to the practical implementation of LLMs in precision oncology use cases involving real-world patient data. We intentionally permit individual aspects from the thematic dimensions to be addressed across different principles, as the thematic dimensions may offer different facets, depending on the principle's narrower thematic focus, all of which can contribute to directive recommendations.

[Figure 3]

## Principle 1: Data Provenance and Quality

Data provenance is defined as "information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness"<sup>46</sup>. Accordingly, LLMs, like other forms of ML systems, leveraging real-world patient data for precision oncology require rigorous standards for reporting on data sources, data composition, training data, data collection, data quality metrics as well as bias and variability, including data sources and demographic information on patients (see Table 1 for detailed recommendations).



Collecting provenance and quality information of all data used as input for inference, e.g., from information systems or prompts, their pre-processing steps, like de-identification, normalization or coding/annotation, and corresponding results, are minimal necessary requirements for inference audit trails.

[Table 1]

## Principle 2: Bias and Fairness

Fairness in connection with avoidance of bias are core values of medical ethics<sup>24,27</sup> and therefore critical concepts to consider for responsible use of LLMs in precision oncology. Both aim to achieve increased transparency and equitable healthcare outcomes across diverse populations. At this, fairness also emphasizes health equity, implying fair access to health resources from the outset. In computational health, diverse definitions of fairness are used<sup>47</sup>, with many criteria for fairness being observational, such as examining the distribution of factors to decide about access to precision oncology<sup>47</sup>. Concepts of fairness are often conflicting, e.g., conflicts between individual and group fairness<sup>47</sup> necessitating balancing principles of group fairness, like demographic parity (distribution of resources based on a singular criterion, e.g., gender), with fairness from the individual's point of view. However, this approach of deliberate differentiation between groups can result in disadvantages for singular individuals, which can be addressed by mechanisms like equality of odds<sup>48</sup>. Apart from known group differentiators, subconscious discrimination in training data for models can lead to algorithmic bias, necessitating critical interpretation, evaluation, and monitoring, including limiting use to suitable application contexts and implementing regular, sample-based human reviews of AI suggestions<sup>49</sup>.

## Principle 3: Explainability

Maximizing the explainability of generative AI enhances transparency and fosters trust among medical professionals. Many AI models, particularly deep learning architectures, are often criticized for their “black box” nature<sup>22</sup>, which is critical in the sensitive context of precision oncology, when the rationale behind their predictions cannot be tracked by the responsible clinicians.

To address this challenge at least partially, researchers have developed explainable AI (XAI) techniques helping users understand why an AI model came to a certain conclusion<sup>50,51</sup>.

A notable example is a study by Kuenzi *et al.*, which demonstrated the potential of XAI in precision oncology by developing an interpretable deep learning model for predicting drug responses in cancer patients. This model achieved high accuracy and provided explanations traceable and consistent with existing biological knowledge, thereby enhancing its clinical utility<sup>52</sup>.

However, trade-offs between explainability and model performance should be acknowledged<sup>53,54</sup>. Inherently explainable models can be limited by their susceptibility to confounders, while post-hoc explainability techniques often fail to transparently reveal the reasoning behind a model's

decisions, potentially misleading users. These methods lack performance guarantees, frequently rely on approximations, and may even reduce model accuracy. Crucially, while XAI methods provide *descriptive* insights into model behavior, they do not offer *normative* evaluations of whether that behavior is justified or appropriate. Attempting to bridge this gap through intuition risks introducing additional biases. Moreover, explainability can paradoxically foster over-reliance on AI systems, potentially reducing user vigilance and critical evaluation. To ensure the safety, efficacy, and fairness of AI systems, thorough validation across diverse populations remains the most reliable approach. XAI should primarily serve as a tool for developers and auditors to probe and refine models, rather than as a justification for individual decisions<sup>55</sup>. Therefore, caution is warranted when interpreting explainability outputs as indicators of clinical or operational correctness. For generative AI models, XAI can involve citing the sources of their recommendations and enabling users to validate outcomes. Techniques like Retrieval Augmented Generation (RAG) technically provide relevant documents as lookup data for the model and plays a crucial role in advancing the explainability, reliability, and auditability of AI assisted clinical decision-making<sup>1</sup>.

## Principle 4: Auditability

Auditability is crucial for ensuring the accountability of AI systems in precision oncology. It is also a key requirement in much of the prevailing regulation governing the use of AI in medicine (see Regulatory Frameworks). This involves the ability to track and review the decision-making process of AI models, identify potential sources of error, and correct them promptly. To achieve auditability, AI systems should incorporate logging and monitoring mechanisms that systematically capture relevant data throughout the decision-making process, such as source document references, model outputs, and precise timestamps. This data can then be analyzed to assess the performance of the model, detect anomalies, and pinpoint areas for improvement<sup>56</sup>.

## Principle 5: Patient Consent, De-Identification, and Privacy

Since the latter half of the 20th century, patient consent and privacy considerations have been considered essential for respecting patient autonomy and physician beneficence<sup>57</sup>.

The clinical and scientific use of genetic data is especially sensitive due to its highly personal nature. Meaningful consent to use genetic data requires patients to understand associated privacy concerns and must enable them to make informed decisions<sup>58</sup>. To mitigate patient privacy concerns reasonably, anonymization and pseudonymization are common data protection methods from the General Data Protection Regulation (GDPR)<sup>59</sup>. Pseudonymization, relying on cryptographic algorithms or lookup tables, renders patients theoretically still identifiable<sup>59</sup>. Anonymization, involving removing all personally identifiable information (PII), is preferred with respect to data protection for patients, as it theoretically renders re-identification impossible<sup>60</sup>. However, true anonymization is challenging, especially for mutational information, since indirect identifiers like age, location, and family and medical history can still enable reidentification<sup>61</sup>.

Simply omitting these data points may lead to increased information loss<sup>62,63</sup>. Thorough selection of anonymization strategies may contribute to acceptable information loss only minimally affecting data utility<sup>62,63</sup>. Anonymization methods should be evaluated across diverse data streams and demographic groups, with detailed documentation of the applied techniques.

Local operation of AI models and protected data storage may enhance data protection, even if full anonymization is not feasible<sup>54</sup>. Technological evolution may necessitate clarifications and adaptations of data protection regimes where appropriate, especially if they only add complexity without enhancing data security.

## **Principle 6: Infrastructure Robustness and Cyber Security**

For protection of patient data and patient well-being robust infrastructure is a key principle of applying generative AI to real-world patient data, necessitating cyber security safeguards and regular internal and external audits. These requirements are enforced by current regulation<sup>21</sup>. In accordance with data minimization, only data that is strictly necessary for the given use case should be stored<sup>59</sup>. Practitioners need to ensure that they use tested and validated software and infrastructure and implement robust governance, risk, and compliance (GRC) frameworks. Operators of an AI system need control over stored data and local data processing (without internet access) is therefore preferable. For good research practice, adherence to FAIR principles (Findable, Accessible, Interoperable, Reusable) is beneficial<sup>64,65</sup>.

## **Principle 7: Continuous Evaluation and Monitoring**

In the SARS-CoV2 pandemic, quickly emerging epidemiological changes in health data were seen, leading to dataset shifts. AI may struggle to recognize such deviations, requiring human supervision to ensure AIs work in situations that differ from their training situations<sup>66</sup>. Due to the highly sensitive applications of generative AI in precision oncology, continuous, multifaceted evaluation and monitoring are required, considering specifics of the respective application context<sup>67</sup>.

Molecular Tumor Boards (MTB) are a key element of precision oncology<sup>68</sup> and see benefits from integration of AI in the field, e.g., for literature search and synthesis<sup>69</sup>. Legal requirements such as the GDPR<sup>42</sup> and EU AI Act<sup>21</sup>, regulate generative AI applications in precision oncology, mandating mechanisms for evaluation and monitoring for high-risk applications. These expectations, next to evaluating real-world performance, can be met by implementing technical measures for monitoring the performance of (generative) AI applications, leveraging established metrics such as Ragas<sup>70</sup> or BERTScore<sup>71</sup>. Additionally, AI system recommendations should be reviewed by clinicians (human-in-the-loop), with comprehensive sample-based reviews<sup>34,72</sup>.

## Principle 8: Validation, Governance, Regulatory Compliance

AI has increasingly been focused on as a topic in jurisdiction, with growing regulations acknowledging this trend<sup>22</sup>. The EU AI Act is a landmark in international regulation, becoming mandatory in August 2026<sup>21,73</sup>. Complexity of legal requirements, particularly in data protection<sup>74</sup>, complicate processes for researchers and may incentivize the development of lower-risk medical AI-devices rather than complex solutions<sup>22</sup>.

Thorough validation is essential before integrating LLMs into clinical workflows, as incorrect AI outputs may lead to considerable risks for individuals<sup>75</sup>. This has led to calls for transparency and accountability in AI prosecution<sup>76</sup>. Current literature on liability discusses factors influencing physicians' accountability, such as increased responsibility when following more unconventional AI recommendations<sup>11</sup>. Edge cases like this may arise more naturally in precision oncology, where standard treatments are either exhausted or unavailable.<sup>77</sup> Researchers suggest resolving these issues by allowing for moral customization of AI settings by patients, returning autonomy to the individual<sup>78</sup>. Quick responses of legislators to trends and innovations may be crucial to protect patients, their families, clinicians, and researchers from uncertain legal situations.

## Principle 9: Human AI Collaboration

Humans and AI are projected to increasingly collaborate on clinical tasks<sup>79,80</sup>. As AI solutions become more advanced, it is crucial to ensure that these technologies respect human uniqueness and dignity, including awareness on influences of AI on users<sup>31</sup>. This implies effective and secure integration of AI applications in clinical routine, enhancing, rather than replacing, the expertise of physicians and providing explainable and responsible AI applications for patients<sup>30-33</sup>. Currently established principles lack clarity at guiding Human AI Collaboration<sup>30</sup>. To achieve an appropriate balance between uncritical acceptance of AI proposals and overt distrust of beneficial AI applications<sup>81</sup>, AI outputs require human validation during model development and in every specific case of use.

Currently, AI validation and training relies heavily on human-annotated data and feedback, especially for LLMs as they generate non-reproducible, non-deterministic output<sup>82,83</sup>.

Risks can be mitigated with "human-in-the-loop" refinement<sup>69</sup>. In precision oncology, this implies ensuring alignment with clinical requirements and by integrating physician specialist domain knowledge. This requires fine-tuning, additional prompt engineering or incorporating trusted (online-) data sources via RAG<sup>84</sup> or In-Context Learning<sup>85</sup>.

## Principle 10: Education and Awareness

Even though easy and comfortable to use<sup>86,87</sup>, AI application in vulnerable patient groups like in precision oncology necessitates specific education for all stakeholders, rather than leaving their responsible use primarily to interested early adopters in the field, creating awareness for potential

risks as well as opportunities. Education can leverage AI for teaching directly, enabling hands-on-learning concepts, together with immersive learning methods, which have been shown to outperform established teaching concepts<sup>88</sup>. Education of patients, physicians and researchers on AI is crucial for valuable and responsible integration of LLM in clinical applications and physicians' training should meet the standards of medical education<sup>89</sup>. This comprises learning on and through AI-tools in three primary ways: directly in the teaching process, as a supplementary tool for educators, and by empowering physicians in training<sup>90</sup>. Since interactions with AI may involve real-world data and personal information of all participants, effective measures for data and privacy protection are essential to ensure learners' legal integrity<sup>91</sup>.

## Thought experiment: Blueprint for LLMs in Clinical Decision Support Systems for Precision Oncology

To validate the usability of the ten principles for the application of generative AI for precision oncology comprehensively, we have applied them to a Rare Gynecological Tumor (RGT) use case.

RGTs present a major challenge in oncology. In gynecology, they account for over 50% of all malignancies and include more than 30 distinct histologic subtypes for this specialty alone, with an estimated 80,000 new cases annually in Europe<sup>92</sup>. The rarity of RGTs, with an incidence of less than 6 per 100,000 women, due to limited patient numbers for each subtype, hinders the development of robust clinical practice guidelines. Consequently, treatment decisions often rely on less standardized approaches, such as retrospective studies, case reports, expert opinions, or extrapolating data from analogous tumors in other anatomical sites<sup>93</sup>.

Uterine carcinosarcoma (UCS) exemplifies the challenges in managing rare tumor entities. Despite established endometrial cancer treatment protocols, UCS remains a distinct, aggressive malignancy with limited therapeutic options. This RGT is frequently excluded from clinical trials, contributing to a poor prognosis characterized by low survival rates and high recurrence. Advanced UCS management is further complicated by treatment resistance, lack of standardized second-line therapies, and poor outcomes with subsequent treatment lines<sup>94</sup>. While the molecular characterization of UCS offers potential therapeutic targets, translating genomic data into effective clinical strategies remains challenging. Here we want to introduce LLM-enabled CDSS applications with a potential to improve disease management in a UCS patient who has progressed on first-line therapy.

We provide a case-based workflow especially underscoring the critical role of real-world patient data integration in the development and refinement of LLMs for precision oncology.

## Thought experiment and Workflow for responsible use of LLMs in Precision Oncology

Upon admission to an oncological care unit of an UCS-patient with progressive disease after first-line treatment, patient data would be collected from existing records and through a detailed anamnesis, ensuring adequate data provenance. Meaningful Informed Consent should be obtained from the UCS patient, ensuring understanding of relevant risks and benefits of Molecular Profiling and LLM use.

Annotation of real-life data and thorough literature research, together with (AI-supported) diagnostic assessment, is necessary at this step. This process is time-consuming, resource-intensive, and prone to suboptimal patient care due to the absence of standardized treatment guidelines and the challenges associated with manual literature review. At this, a trained professional can leverage an LLM-enabled CDSS - respecting medical standards, data protection requirements, as well as ethical requirements to alleviate these tasks. By instantly analyzing vast amounts of real-world patient data, streamlining literature reviews, and aiding in the generation of treatment plans, a CDSS offers a potential solution to overcome the existing bottlenecks at this point and improve patient care.

The developed treatment suggestions from the annotation-process facilitated through this approach form the basis for Shared Decision-Making (SDM) on identified treatment options. At this step, patients must be informed and educated about the use of CDSS and on the therapeutic options. LLMs can support patient-trial matching, if trials are available, or, otherwise, can assist in reviewing current evidence to suggest treatments, including potential off-label options for rare diseases. The paucity of UCS-specific clinical trials usually necessitates off-label treatment decisions based on the available evidence. Therefore, inclusion into clinical trials/a Cost Coverage Request is issued, again by the help of an LLM-enabled CDSS, and finally, an actionable treatment plan is developed. The treatment plan is executed, and the documented results are analyzed with support of LLMs that strictly adhere to legal and ethical requirements, avoid bias to the extent possible and enable clinicians and researchers to interact in a responsible, efficient way with the AI. Clinical outcomes are documented, using LLMs that strictly follow data protection and ethical requirements. Resulting multimodal data of a variety of patients is then analyzed and used to enhance LLM performance in supporting effective UCS treatment. This blueprint use case of a gynecological precision oncology patient with uterine carcinosarcoma serves at highlighting a framework for the application of similar methodologies to refractory diseases for other fields of precision oncology. The exemplary workflow is illustrated in the flowchart in Figure 4.

[Figure 4]

With this thought experiment, we demonstrate how rigorously applying our design principles for RAI in CDSS ensures that an LLM will address all five dimensions of RAI in precision oncology with real world data. The ensuing system adheres to medical standards, balances technology benefits and risks, complies to regulatory frameworks, respects ethical considerations, and ensures that the system will provide benefits to patients, clinicians, and the scientific community.

A detailed checklist for the application RAI in comparable Precision Oncology settings is provided in the Supplement (Supplement 1).

In our thought experiment, we propose a workflow that operationalizes our framework of five dimensions and ten practice-oriented principles for real-world applications. The UCS patient case can be readily replaced by anonymized real-world cases, such as those from The Cancer Genome Atlas (TCGA) UCS cohort established in the United States<sup>95,96</sup>.

While we highlighted the principal applicability of our framework in precision oncology contexts, we note the limitations of our framework. The thematic dimensions may be broadly transferable to other domains of precision medicine, such as neurology, the principles reflect context-specific guidance tailored to precision oncology. Further refinement and adaptation will be required to apply the principles to other medical fields. Also, given the continuous evolution of the field, critical engagement with ongoing developments within the context of this framework remains essential.

## Conclusion

Generative AI in oncology holds potential for improving patient, scientific and medical outcomes. Responsibly integrating real-world patient data is required to fully realize this potential. Recent advancements have made AI more accessible and user-friendly, but its responsible use in precision oncology requires comprehensive education for all stakeholders. This review explores five dimensions for AI implementation: adherence to medical standards, promotion of scientific and medical benefits, consideration of technological advantages and risks, regulatory frameworks, ethical considerations; as well as ten guiding principles for the application of generative AI in precision oncology: Data Provenance and Quality, Bias and Fairness, Explainability, Auditability, Patient Consent, De-identification and Privacy, Infrastructure Robustness and Cybersecurity, Continuous Evaluation and Monitoring, Validation, Governance and Regulatory Compliance, Human-AI Collaboration, Education and Awareness. These principles will assist practitioners in integrating generative AI within LLMs for precision oncology, using real-world patient data in alignment with RAI principles.

## Acknowledgements

We would like to thank the Board of Directors of CCE for critically reading, revising, and approving this paper. This study received no funding.

1084

## 1085 **Author Contributions**

1086 SM has made substantial contributions the conception, drafting, writing, revising, and finalizing  
 1087 the manuscript and is the main author of this work. DF has made substantial contributions at  
 1088 drafting, writing of the manuscript, and has critically revised this work. TD has made substantial  
 1089 contributions at conceptualizing the manuscript and has substantially revised this work. KB has  
 1090 made substantial contributions at conceptualizing the manuscript and has substantially revised  
 1091 this work. LM has made substantial contributions at conceptualizing the manuscript and has  
 1092 substantially revised this work. TW has made substantial contributions at conceptualizing the  
 1093 manuscript and has substantially revised this work. RD has substantially revised this work. JV  
 1094 has substantially revised this work. SK has substantially revised this work.  
 1095 RPL has made substantial contributions at designing this work and substantially revised the  
 1096 manuscript. AP has substantially revised this work. FS has substantially revised this work.  
 1097 RDB has made substantial contributions at designing this work and substantially revised the  
 1098 manuscript. MB has made substantial contributions at conceptualizing the manuscript and has  
 1099 substantially revised this work. JNK has made substantial contributions at conceptualizing the  
 1100 manuscript and has substantially revised this work. MT has contributed substantially to the  
 1101 conception and drafting of the manuscript and at the analysis of data and has critically revised  
 1102 this work. JL has made substantial contributions at the conception, drafting, writing, revising  
 1103 and finalizing the manuscript and has supervised this work. All authors have approved the final  
 1104 version of the manuscript. All authors agreed both to be personally accountable for the  
 1105 author's own contributions and to ensure that questions related to the accuracy or integrity of  
 1106 any part of the work, even ones in which the author was not personally involved, are  
 1107 appropriately investigated, resolved, and the resolution documented in the literature.

## 1108 **Competing Interest Statement**

1109 SM has received speaker honoraria and travel reimbursements from ALK-Abelló (Copenhagen,  
 1110 Denmark) and Bencard Allergie GmbH (Munich, Germany). DF has received an OpenAI research  
 1111 grant and holds shares and is an employee at Synagen GmbH, Germany. TD has no COI. KJB  
 1112 declares speaker honoraria from Accuray inc. (Sunnyvale, California, United States), travel  
 1113 reimbursement from Zeiss group (Oberkochen, Baden-Württemberg, Germany) and consulting  
 1114 services for NEED Inc., Santa Monica, CA 90405, United States. LM has no COI. TW has no COI.  
 1115 FS has received speaker fees from Pfizer, Sweden, speaker fees from Lunit, South Korea, and  
 1116 is a shareholder of ClearScanAI AB, Sweden; MB declares no COI. A.P. has a consulting/advisory  
 1117 role for BMS, AstraZeneca, Novartis, MSD, Lilly, Amgen, Pfizer, and Johnson & Johnson; travel,  
 1118 accommodations, or other expenses paid or reimbursed by Roche and Johnson & Johnson; and  
 1119 serves as the principal investigator for Spectrum Pharmaceuticals, BMS, Bayer, MSD, and Lilly  
 1120 outside the submitted work. RDB acknowledges support from: Cancer Research UK, Cambridge  
 1121 Experimental Cancer Medicine Centre and Cambridge Biomedical Research Centre (BRC-1215-  
 1122 20014). RDB declares: consulting services / honoraria (to institution) from AstraZeneca,  
 1123 Genentech/Roche, Molecular Partners, Novartis, Shionogi; and research grants from



1124 AstraZeneca, Boehringer-Ingelheim, Genentech/Roche. JNK declares consulting services for  
1125 Bioprimus, France; Owkin, France; DoMore Diagnostics, Norway; Panakeia, UK; AstraZeneca,  
1126 UK; Scaillyte, Switzerland; Mindpeak, Germany; and MultiplexDx, Slovakia. Furthermore, he  
1127 holds shares in StratifAI GmbH, Germany, Synagen GmbH, Germany, and has received a  
1128 research grant by GSK, and has received honoraria by AstraZeneca, Bayer, Daiichi Sankyo,  
1129 Eisai, Janssen, MSD, BMS, Roche, Pfizer and Fresenius. MT has no COI. JL received speaker  
1130 honoraria by the Forum for Continuing Medical Education (Germany), AstraZeneca (Germany)  
1131 and Novartis (Germany).

1132

ARTICLE IN PRESS

## References

- 1 Ferber, D. *et al.* GPT-4 for information retrieval and comparison of medical oncology guidelines. *Nejm Ai* **1**, Alcs2300235 (2024).
- 2 Zakka, C. *et al.* Almanac—retrieval-augmented language models for clinical medicine. *Nejm ai* **1**, Aloa2300068 (2024).
- 3 Buckley, T., Diao, J. A., Rodman, A. & Manrai, A. K. Accuracy of a vision-language model on challenging medical cases. *arXiv preprint arXiv:2311.05591* (2023).
- 4 Eriksen, A. V., Möller, S. & Ryg, J. Vol. 1 Alp2300031 (Massachusetts Medical Society, 2024).
- 5 Bedi, S. *et al.* A systematic review of testing and evaluation of healthcare applications of large language models (LLMs). *medRxiv*, 2024.2004. 2015.24305869 (2024).
- 6 Bedi, S. *et al.* Testing and Evaluation of Health Care Applications of Large Language Models: A Systematic Review. *Jama* **333**, 319-328, doi:10.1001/jama.2024.21700 (2025).
- 7 Good, B. M., Ainscough, B. J., McMichael, J. F., Su, A. I. & Griffith, O. L. Organizing knowledge to enable personalization of medicine in cancer. *Genome biology* **15**, 1-9 (2014).
- 8 Oncology, T. L. Vol. 25 945 (2024).
- 9 Stahl, B. C. Embedding responsibility in intelligent systems: from AI ethics to responsible AI ecosystems. *Scientific Reports* **13**, 7586 (2023).
- 10 Fleisher, L. A. & Economou-Zavlanos, N. J. in *JAMA Health Forum*. e241369-e241369 (American Medical Association).
- 11 Tobia, K., Nielsen, A. & Stremitzer, A. When does physician use of AI increase liability? *Journal of Nuclear Medicine* **62**, 17-21 (2021).
- 12 Carmona, J. *et al.* Cancer Core Europe: Leveraging institutional synergies to advance oncology research and care globally. *Cancer discovery* **14**, 1147-1153 (2024).
- 13 Moffett, P. & Moore, G. The standard of care: legal history and definitions: the bad and good news. *Western Journal of Emergency Medicine* **12**, 109 (2011).
- 14 Meineke, F., Modersohn, L., Loeffler, M. & Boeker, M. in *Caring is Sharing—Exploiting the Value in Data for Health and Innovation* 835-836 (IOS Press, 2023).
- 15 Tayefi, M. *et al.* Challenges and opportunities beyond structured data in analysis of electronic health records. *WIREs Computational Statistics* **13**, e1549, doi:https://doi.org/10.1002/wics.1549 (2021).
- 16 Martino, A., Iannelli, M. & Truong, C. in *European Semantic Web Conference*. 182-185 (Springer).
- 17 Balagurunathan, Y., Mitchell, R. & El Naqa, I. Requirements and reliability of AI in the medical context. *Physica Medica* **83**, 72-78 (2021).
- 18 Longhurst, C. A., Singh, K., Chopra, A., Atreja, A. & Brownstein, J. S. Vol. 1 Alp2400223 (Massachusetts Medical Society, 2024).
- 19 .
- 20 Regulation - 2017/746 - EN - Medical Device Regulation - EUR-Lex. (2017).
- 21 The Act Texts.
- 22 Gottlieb, S. in *JAMA Health Forum*. e242691-e242691 (American Medical Association).

- 1177 23 Busch, F. *et al.* Navigating the european union artificial intelligence act for healthcare.  
1178 *npj Digital Medicine* **7**, 210 (2024).
- 1179 24 World Medical Association Declaration of Geneva. *Afr Health Sci* **17**, 1203,  
1180 doi:10.4314/ahs.v17i4.30 (2017).
- 1181 25 ABIM, A.-A. & Charter, E. A. P. Project of the ABIM Foundation. *ACP-ASIM Foundation,*  
1182 *and European Federation of Internal Medicine: ABIM Foundation, ACP-ASIM*  
1183 *Foundation, and European Federation of Internal Medicine* (2002).
- 1184 26 Pappworth, M. Modern Hippocratic Oath. *British Medical Journal* **1**, 668 (1971).
- 1185 27 Beauchamp, T. & Childress, J. Vol. 19 9-12 (Taylor & Francis, 2019).
- 1186 28 Beauchamp, T. L. & Childress, J. F. *Principles of biomedical ethics*. (Edicoes Loyola,  
1187 1994).
- 1188 29 U, W. The Declaration of Helsinki—Its History and Its Future. (2014).
- 1189 30 Mittelstadt, B. Principles alone cannot guarantee ethical AI. *Nature machine intelligence*  
1190 **1**, 501-507 (2019).
- 1191 31 Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. *Nature*  
1192 *machine intelligence* **1**, 389-399 (2019).
- 1193 32 Floridi, L. *et al.* AI4People—an ethical framework for a good AI society: opportunities,  
1194 risks, principles, and recommendations. *Minds and machines* **28**, 689-707 (2018).
- 1195 33 Rubeis, G. *Ethics of medical AI*. (Springer, 2024).
- 1196 34 Buckley, R. P., Zetzsche, D. A., Arner, D. W. & Tang, B. W. Regulating artificial  
1197 intelligence in finance: putting the human in the loop. *Sydney Law Review, The* **43**, 43-  
1198 81 (2021).
- 1199 35 Peters, D., Vold, K., Robinson, D. & Calvo, R. A. Responsible AI—two frameworks for  
1200 ethical design practice. *IEEE Transactions on Technology and Society* **1**, 34-47 (2020).
- 1201 36 McLennan, S. *et al.* An embedded ethics approach for AI development. *Nature Machine*  
1202 *Intelligence* **2**, 488-490 (2020).
- 1203 37 Thirunavukarasu, A. J. *et al.* Large language models in medicine. *Nature medicine* **29**,  
1204 1930-1940 (2023).
- 1205 38 Clusmann, J. *et al.* The future landscape of large language models in medicine.  
1206 *Communications medicine* **3**, 141 (2023).
- 1207 39 Uprety, D., Zhu, D. & West, H. ChatGPT—A promising generative AI tool and its  
1208 implications for cancer care. *Cancer* **129**, 2284-2289 (2023).
- 1209 40 Chandak, P., Huang, K. & Zitnik, M. Building a knowledge graph to enable precision  
1210 medicine. *Scientific Data* **10**, 67 (2023).
- 1211 41 From Einstein to AI: how 100 years have shaped science. *Nature* **624**, 474,  
1212 doi:10.1038/d41586-023-04021-2 (2023).
- 1213 42 (2016 ).
- 1214 43 Yoshizawa, G. *et al.* Limiting open science? Three approaches to bottom-up governance  
1215 of dual-use research of concern. *Pathogens and Global Health* **118**, 285-294 (2024).
- 1216 44 Khanna, N. N. *et al.* in *Healthcare*. 2493 (MDPI).
- 1217 45 Mehta, M. C., Katz, I. T. & Jha, A. K. Transforming global health with AI. *New England*  
1218 *Journal of Medicine* **382**, 791-793 (2020).
- 1219 46 PROV-Overview.
- 1220 47 Dwork, C., Hardt, M., Pitassi, T., Reingold, O. & Zemel, R. in *Proceedings of the 3rd*  
1221 *innovations in theoretical computer science conference*. 214-226.

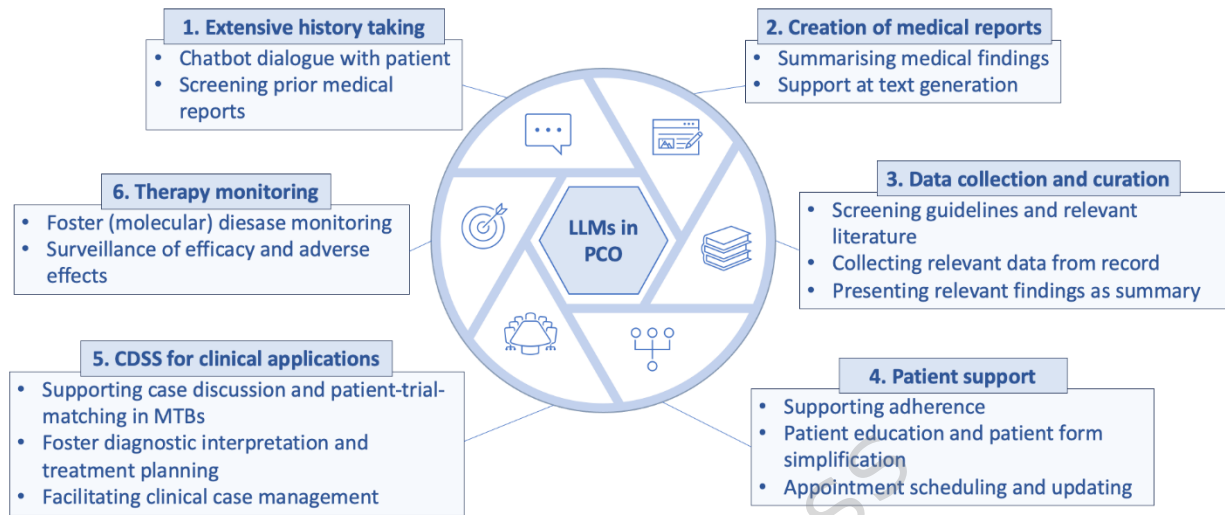
- 1222 48 Yang, Y., Zhang, H., Gichoya, J. W., Katabi, D. & Ghassemi, M. The limits of fair medical  
1223 imaging AI in real-world generalization. *Nature Medicine* **30**, 2838-2848 (2024).
- 1224 49 Franklin, G. *et al.* The sociodemographic biases in machine learning algorithms: a  
1225 biomedical informatics perspective. *Life* **14**, 652 (2024).
- 1226 50 Zafar, M. R. & Khan, N. M. DLIME: A deterministic local interpretable model-agnostic  
1227 explanations approach for computer-aided diagnosis systems. *arXiv preprint*  
1228 *arXiv:1906.10263* (2019).
- 1229 51 Nohara, Y., Matsumoto, K., Soejima, H. & Nakashima, N. Explanation of machine  
1230 learning models using shapley additive explanation and application for real data in  
1231 hospital. *Computer Methods and Programs in Biomedicine* **214**, 106584 (2022).
- 1232 52 Kuenzi, B. M. *et al.* Predicting drug response and synergy using a deep learning model  
1233 of human cancer cells. *Cancer cell* **38**, 672-684. e676 (2020).
- 1234 53 Herm, L.-V., Heinrich, K., Wanner, J. & Janiesch, C. Stop ordering machine learning  
1235 algorithms by their explainability! A user-centered investigation of performance and  
1236 explainability. *International Journal of Information Management* **69**, 102538 (2023).
- 1237 54 Wiest, I. C. *et al.* Anonymizing medical documents with local, privacy preserving large  
1238 language models: The LLM-Anonymizer. *medRxiv*, 2024.2006. 2011.24308355 (2024).
- 1239 55 Ghassemi, M., Oakden-Rayner, L. & Beam, A. L. The false hope of current approaches  
1240 to explainable artificial intelligence in health care. *The Lancet Digital Health* **3**, e745-  
1241 e750 (2021).
- 1242 56 Brundage, M. *et al.* Toward trustworthy AI development: mechanisms for supporting  
1243 verifiable claims. *arXiv preprint arXiv:2004.07213* (2020).
- 1244 57 Gostin, L. Health care information and the protection of personal privacy: ethical and  
1245 legal considerations. *Annals of Internal Medicine* **127**, 683-690 (1997).
- 1246 58 Burgess, M. M. Beyond consent: ethical and social issues in genetic testing. *Nature*  
1247 *Reviews Genetics* **2**, 147-151 (2001).
- 1248 59 European, P. & Council. Regulation (EU) 2016/679 of the European Parliament and of  
1249 the Council of 27 April 2016 on the protection of natural persons with regard to the  
1250 processing of personal data and on the free movement of such data, and repealing  
1251 Directive 95/46/EC (General Data Protection Regulation). 1-88 (2016).
- 1252 60 Malin, B. & Sweeney, L. in *Proceedings of the AMIA Symposium*. 423.
- 1253 61 El Emam, K. & Dankar, F. K. Protecting privacy using k-anonymity. *Journal of the*  
1254 *American Medical Informatics Association* **15**, 627-637 (2008).
- 1255 62 Neubauer, T. & Heurix, J. A methodology for the pseudonymization of medical data.  
1256 *International journal of medical informatics* **80**, 190-204 (2011).
- 1257 63 Mehtälä, J. *et al.* Utilization of anonymization techniques to create an external control  
1258 arm for clinical trial data. *BMC Medical Research Methodology* **23**, 258 (2023).
- 1259 64 Good Research Practice.
- 1260 65 Lamprecht, A.-L. *et al.* Towards FAIR principles for research software. *Data Science* **3**,  
1261 37-59 (2020).
- 1262 66 Finlayson, S. G. *et al.* The clinician and dataset shift in artificial intelligence. *New*  
1263 *England Journal of Medicine* **385**, 283-286 (2021).
- 1264 67 van de Sande, D. *et al.* To warrant clinical adoption AI models require a multi-faceted  
1265 implementation evaluation. *NPJ Digital Medicine* **7**, 58 (2024).

- 1266 68 Boos, L. & Wicki, A. The molecular tumor board—a key element of precision oncology.  
1267 *memo-Magazine of European Medical Oncology* **17**, 190-193 (2024).
- 1268 69 Lammert, J. *et al.* Expert-guided large language models for clinical decision support in  
1269 precision oncology. *JCO precision oncology* **8**, e2400478 (2024).
- 1270 70 Es, S., James, J., Anke, L. E. & Schockaert, S. in *Proceedings of the 18th Conference of*  
1271 *the European Chapter of the Association for Computational Linguistics: System*  
1272 *Demonstrations*. 150-158.
- 1273 71 Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q. & Artzi, Y. Bertscore: Evaluating text  
1274 generation with bert. *arXiv preprint arXiv:1904.09675* (2019).
- 1275 72 Cohen, I. G. *et al.* How AI can learn from the law: putting humans in the loop only on  
1276 appeal. *npj Digital Medicine* **6**, 160 (2023).
- 1277 73 Gilbert, S. The EU passes the AI Act and its implications for digital medicine are unclear.  
1278 *NPJ Digital Medicine* **7**, 135 (2024).
- 1279 74 Li, H., Yu, L. & He, W. Vol. 22 1-6 (Taylor & Francis, 2019).
- 1280 75 Computers make mistakes and AI will make things worse - the law must recognize that.  
1281 *Nature* **625**, 631, doi:10.1038/d41586-024-00168-8 (2024).
- 1282 76 Marshall, P. *et al.* Recommendations for the probity of computer evidence. *Digital*  
1283 *Evidence & Elec. Signature L. Rev.* **18**, 18 (2021).
- 1284 77 Rieke, D. T. *et al.* Feasibility and outcome of reproducible clinical interpretation of high-  
1285 dimensional molecular data: a comparison of two molecular tumor boards. *BMC*  
1286 *medicine* **20**, 367 (2022).
- 1287 78 Loh, E. Medicine and the rise of the robots: a qualitative review of recent advances of  
1288 artificial intelligence in health. *BMJ leader*, leader-2018-000071 (2018).
- 1289 79 Rajpurkar, P. & Lungren, M. P. The current and future state of AI interpretation of  
1290 medical images. *New England Journal of Medicine* **388**, 1981-1990 (2023).
- 1291 80 Scheetz, J. *et al.* A survey of clinicians on the use of artificial intelligence in  
1292 ophthalmology, dermatology, radiology and radiation oncology. *Sci Rep* **11**, 5193,  
1293 doi:10.1038/s41598-021-84698-5 (2021).
- 1294 81 Messeri, L. & Crockett, M. Artificial intelligence and illusions of understanding in scientific  
1295 research. *Nature* **627**, 49-58 (2024).
- 1296 82 Ouyang, L. *et al.* Training language models to follow instructions with human feedback.  
1297 *Advances in neural information processing systems* **35**, 27730-27744 (2022).
- 1298 83 Shah, N. H., Entwistle, D. & Pfeffer, M. A. Creation and adoption of large language  
1299 models in medicine. *Jama* **330**, 866-869 (2023).
- 1300 84 Lewis, P. *et al.* Retrieval-augmented generation for knowledge-intensive nlp tasks.  
1301 *Advances in neural information processing systems* **33**, 9459-9474 (2020).
- 1302 85 Brown, T. *et al.* Language models are few-shot learners. *Advances in neural information*  
1303 *processing systems* **33**, 1877-1901 (2020).
- 1304 86 Toros, E., Asiksoy, G. & Sürücü, L. Refreshment students' perceived usefulness and  
1305 attitudes towards using technology: a moderated mediation model. *Humanities and*  
1306 *Social Sciences Communications* **11**, 1-10 (2024).
- 1307 87 Mustafa, S., Zhang, W., Anwar, S., Jamil, K. & Rana, S. An integrated model of UTAUT2  
1308 to understand consumers' 5G technology acceptance using SEM-ANN approach.  
1309 *Scientific Reports* **12**, 20056 (2022).

- 88 Du Boulay, B. Artificial intelligence as an effective classroom assistant. *IEEE Intelligent Systems* **31**, 76-81 (2016).
- 89 Schwill, S. *et al.* The WFME global standards for quality improvement of postgraduate medical education: Which standards are also applicable in Germany? Recommendations for physicians with a license for postgraduate training and training agents. *GMS J Med Educ* **39**, Doc42, doi:10.3205/zma001563 (2022).
- 90 Narayanan, S., Ramakrishnan, R., Durairaj, E. & Das, A. Artificial intelligence revolutionizing the field of medical education. *Cureus* **15** (2023).
- 91 Franco D'Souza, R., Mathew, M., Mishra, V. & Surapaneni, K. M. Twelve tips for addressing ethical concerns in the implementation of artificial intelligence in medical education. *Medical Education Online* **29**, 2330250 (2024).
- 92 Gatta, G. *et al.* Rare cancers are not so rare: the rare cancer burden in Europe. *European journal of cancer* **47**, 2493-2511 (2011).
- 93 Lainé, A., Hanvic, B. & Ray-Coquard, I. Importance of guidelines and networking for the management of rare gynecological cancers. *Current Opinion in Oncology* **33**, 442-446 (2021).
- 94 Bogani, G. *et al.* Endometrial carcinosarcoma. *Int J Gynecol Cancer* **33**, 147-174, doi:10.1136/ijgc-2022-004073 (2023).
- 95 National Cancer Institute. Uterine carcinosarcoma study (TCGA). National Cancer Institute Center for Cancer Genomics, <https://www.cancer.gov/ccg/research/genome-sequencing/tcga/studied-cancers/uterine-carcinosarcoma-study> (accessed Sept 24, 2025).
- 96 Cherniack, A. D. *et al.* Integrated molecular characterization of uterine carcinosarcoma. *Cancer Cell* **31**, 411–423, doi:10.1016/j.ccell.2017.02.010 (2017).

## Figures

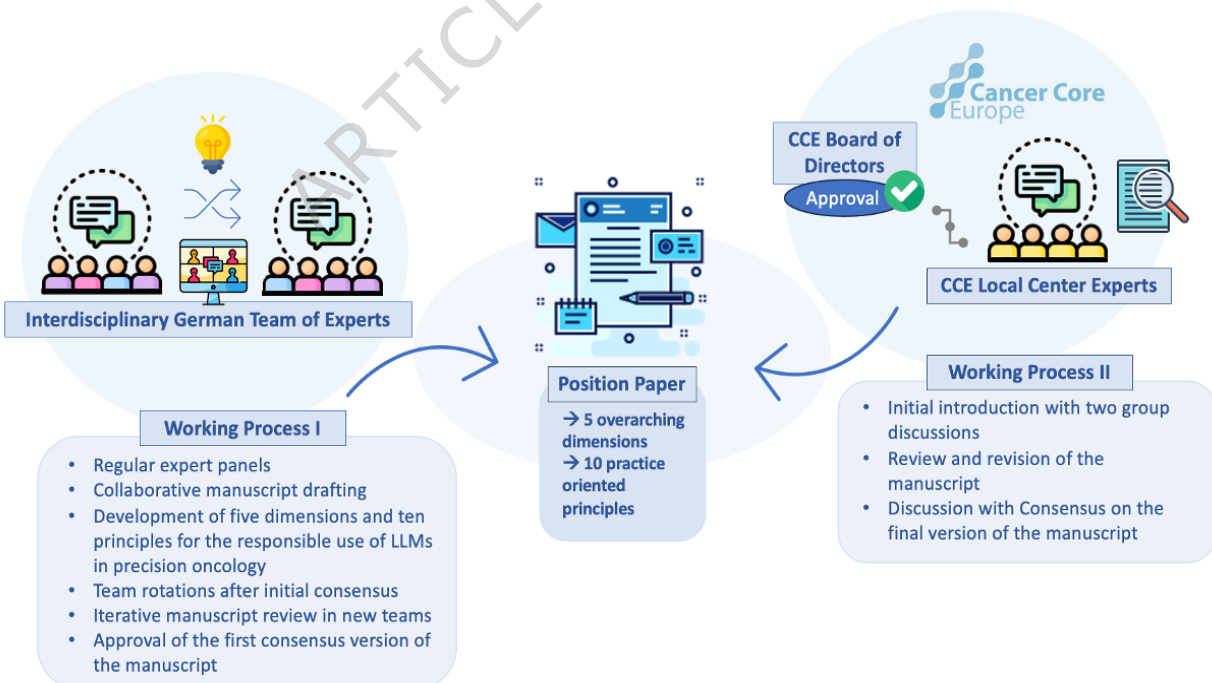
Figure 1: Applications of LLMs in Precision Oncology across the diagnostic and therapeutic process



LLM: Large Language Models, PCO: Precision Oncology, MTB: Molecular Tumor Board

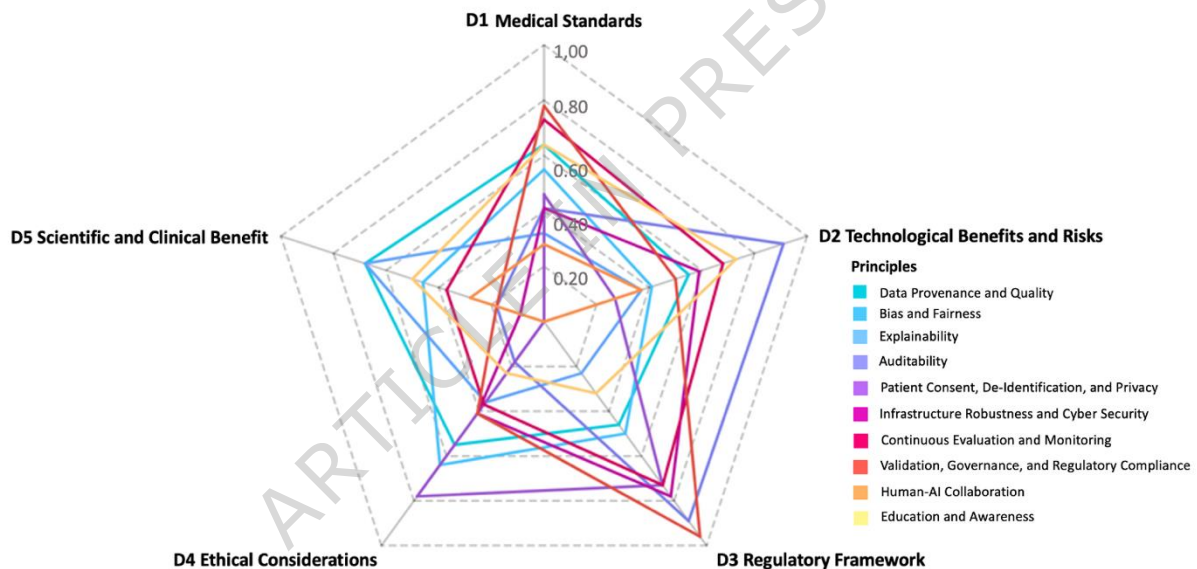
*Legend Figure 1: LLMs can assist clinicians and related professions in the field of precision oncology with various tasks to eventually develop actionable treatment plans for patients. These models can be tailored to support time-consuming tasks, thereby improving resource management, and enhancing the quality of care.*

Figure 2: Methodology of manuscript compilation with the development of 5 overarching dimensions and 10 distinct practice-oriented principles



*Legend Figure 2: Structured development of a position paper on the responsible use of LLMs in precision oncology based on expert discussions and iterative consensus-building. An interdisciplinary German team of experts drafted five dimensions, each representing a thematic complex of responsible LLM-use for precision oncology with real world data. From these more abstract complexes, ten practice-oriented, actionable principles were developed deductively through group discussions and repeated manuscript reviews. The manuscript underwent two working processes: initial drafting, team-based revisions, and consensus formation, followed by validation through discussions with local center experts within Cancer Core Europe (CCE). Final approval was granted by the CCE Board to ensure alignment with institutional and scientific standards.*

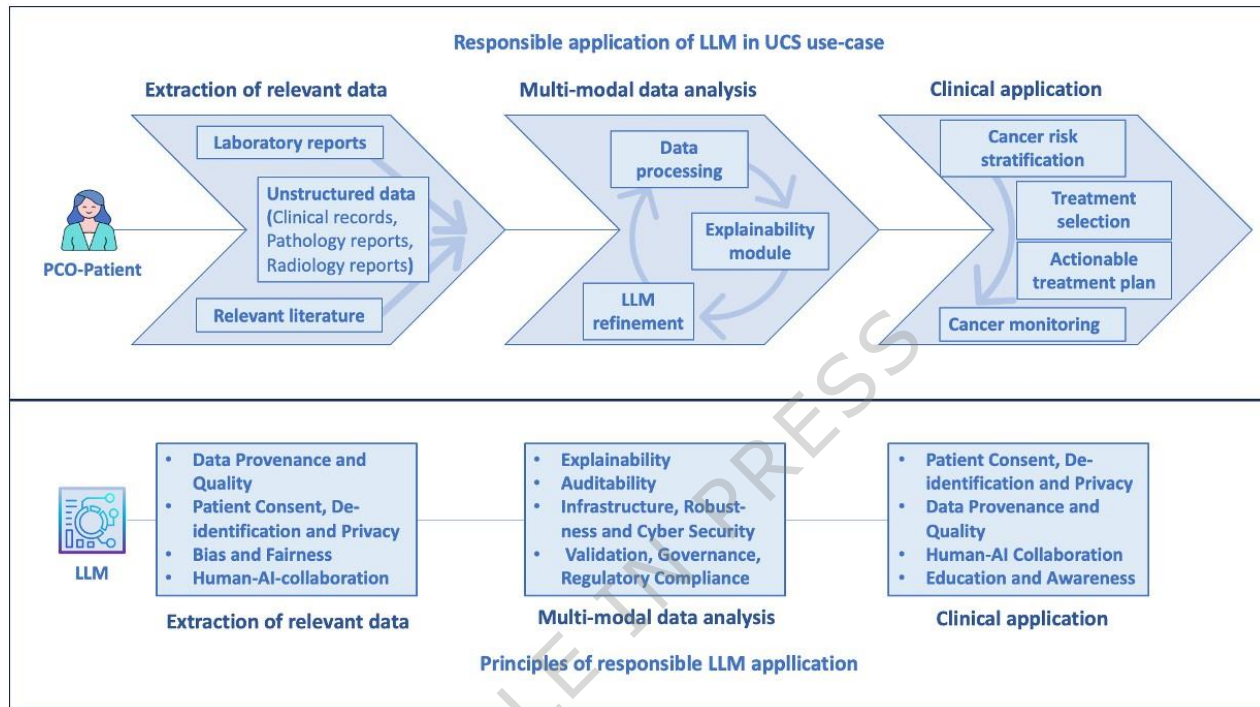
Figure 3: Mapping the influence of thematic dimensions on principle formulation based on a consensus-survey of 11 experts.



*Legend Figure 3: Evaluation of the ten user-oriented principles for applying LLMs in precision oncology based on their relevance across five thematic dimensions. Participating experts were asked to assess the contribution of each dimension to the respective principle: adherence to medical standards, technological benefits and risks, regulatory framework, ethical considerations, and scientific and clinical benefit, based on their professional experience. The mapping took place after the formulation of the principles was completed, to check for consensus on the relevance of the principle across thematic dimensions. For the mapping, the collection of aspects of each principle was rated dimension-wise for relevant overlap between principle and dimension by using a 5-point Likert scale.*



Figure 4: Flowchart for Integration of Responsible use of LLMs in Decision Support Systems (CDSS) with Real-world patient data for PCO-Patients, exemplified for Patients with Uterine Carcinosarcoma.



Legend Figure 4: Structured framework (blueprint) to the responsible integration of LLMs in precision oncology workflows. LLMs are applied across data extraction, multi-modal data analysis, and clinical decision-making to process laboratory reports, unstructured clinical data, and literature, support data analysis and explainability, and assist in cancer risk stratification, quality, patient privacy, explainability, auditability, and regulatory compliance, reflects their contextual relevance across workflow components to ensure responsible LLM deployment in clinical and research settings.

## Tables

Table 1: Suggestions to support adequate Data Provenance and Quality for applying generative AI to real-world patient data

<b>Data Sources</b>	<ul style="list-style-type: none"> <li>Specify origins and types of datasets used with inclusion of primary care data (e.g., electronic health records, clinical trials, and registries) as well as generic medical data (textbooks, PubMed, clinical coding systems).</li> </ul>
---------------------	---

<b><i>Dataset Composition</i></b>	<ul style="list-style-type: none"> <li>• List the total number of individuals, cases, or records, and duplicates included in the dataset.</li> <li>• Clearly describe inclusion and exclusion criteria and limitations of the dataset.</li> <li>• Consider distinct characteristics and differences between vendor-level LLM training requirements and hospital-specific application in emphasizing differing challenges in data provenance.</li> </ul>
<b><i>Data Quality Metrics</i></b>	<ul style="list-style-type: none"> <li>• Disclose detailed dataset statistics like distribution of data types (such as clinical notes, imaging, laboratory values) and descriptive statistics of data missingness.</li> <li>• Provide details of deidentification pipeline use.</li> <li>• Up-to-date indicating the collection period from start to end or details on coding systems and ontologies employed, such as ICD10 or SNOMED-CT.</li> </ul>
<b><i>Demographic Information</i></b>	<ul style="list-style-type: none"> <li>• Inform on geographic locations from which the data was collected.</li> <li>• Represent regional variations in disease prevalence, treatment approaches, and healthcare access in the data (health equity).</li> <li>• Ensure representation of ethnic minority and marginalized populations.</li> </ul>
<b><i>Bias and Data Variability</i></b>	<ul style="list-style-type: none"> <li>• Provide descriptions of all measures taken to detect and correct for biases in the data, addressing issues like over- or underrepresentation of specific groups.</li> </ul>
<b><i>Legal Requirements</i></b>	<ul style="list-style-type: none"> <li>• Main privacy requirements in compliance with HIPAA (US) or GDPR (EU), including consent for the incorporation of non-anonymized data into datasets.</li> <li>• Disclose details on data licensing and reuse terms for any published literature data used in model training.</li> </ul>
<b><i>Ethical aspects</i></b>	<ul style="list-style-type: none"> <li>• Detail any ethical approvals obtained for the use of data, especially at the use of sensitive data.</li> </ul>

Legend Table 1: This table outlines key recommendations for establishing robust data provenance and quality standards in the use of generative AI, like LLMs, for precision oncology. It addresses critical elements including the specification of data sources (e.g., electronic health records, clinical trials, registries, and general medical data), detailed documentation of dataset

1393 composition (such as case counts, inclusion/exclusion criteria, and application-specific  
1394 nuances), and reporting of data quality metrics (including data type distributions, temporal  
1395 coverage, and coding systems used). Additionally, it highlights the importance of systematically  
1396 identifying and mitigating biases and data variability to support fair and reliable AI applications.  
1397 These components form the minimal requirements for ensuring data provenance and quality for  
1398 responsible use of real-world patient data in AI-driven clinical decision support.  
1399

ARTICLE IN PRESS