CURTAIN

D3.3 Guidelines on LLMs and cancer literacy

2026

Co-funded by the European Union

Co-funded by the European Union's EU4Health programme under grant agreement No 101219448. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

Project number: 101219448
Project name: Beating Cancer Inequalities through Literacy in Europe
Project acronym: CURTAIN
Funding programme: EU4Health, EU4H-2024-PJ-02

Version number	
Status	Final
Dissemination level	Public
Due date of deliverable	31/08/2026
Actual submission date	31/08/2026
Work package	WP 3 Design of the cancer literacy tools and interventions
Task name	T3.4: Developing Guidelines for the Appropriate Use of Large Language Models to Enhance Cancer Literacy
Lead partner	INOMED
Partner(s) contributing	BJCN, INSA, UA, DSMU, EUROC, FPG
Authors	
Main author name	Organization
Andra Păun	INOMED
Contributing authors	Organization acronym
Marius Geantă	INOMED
Ruxandra Schitea	INOMED
Bogdan Vidrașcu	INOMED
Reviewers	
Reviewer name	Organization acronym
Ivaylo Petrov	BJCN

Statement of originality *This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.*

Version Tracker

Date	Version number*	Author(s)	Change
13/07/2026	0.1	Andra Păun Bogdan Vidrașcu	First draft Methodology section
11/08/2026	0.2	Ivaylo Petrov	Review, content refinement, reference addition
15/08/2026	0.3	Andra Păun Ruxandra Schitea	Content refinement, Annex addition
21/08/2026	0.4	Ivaylo Petrov	Second thorough review
26/08/2026	0.5	Andra Păun	Final version
31/08/2026	1	Marius Geantă	Final review

Table of Contents

Table of Contents 3
List of Figures 5
List of Tables 6
List of Abbreviations 7
Executive Summary 8
Introduction 11
Objective and scope 11
Relation to other WPs and deliverables 13
Content of the deliverable 13
Methodology 14
Integrative review 15
Community-based qualitative validation 15
Expert consultation and co-creation of the guidelines 15
Multi-stakeholder validation 16
1. Cancer burden and AI fundamentals 16
1.1. Global cancer burden 16
1.2. Artificial Intelligence and Large Language Models concepts 17
1.3. Prompting engineering as a core competency in medical AI literacy 19
1.4. Role of LLMs in cancer literacy 20
1.5. Medical Digital Twins and personalized AI 21
2. Cancer health literacy as predictor of outcomes and European Initiatives 22
3. Applications of LLMs in health and cancer literacy 22
3.1. Readability and quality assessments 23
3.2. Performance in culturally specific contexts 24
3.3. Performance in clinical decision-support 24
4. AI-Based Chatbots in cancer education and care 25
4.1. Literature evidence from recent studies 25
4.2. Evidence in specific cancers 26
4.3. Empathic communication 27

4.4. Regulatory standards, ethical considerations, and compliance challenges for AI-based chatbots (EU context) 27
5. Existing EU and Global Regulatory Frameworks 29
5.1. European Union frameworks 29
5.2. Global and International frameworks 31
5.3. Sector-specific and professional guidance 34
5.4. Common principles and implementation framework 36
6. Current clinical guidelines and evidence use of AI in oncology 37
6.1. Existing guidelines and professional perspectives 37
6.2. Examples of approved AI tools for different cancer types 38
7. A bottom-up framework for developing guidelines on the use of LLMs in cancer literacy 40
7.1. Rationale, guiding principles, and governance considerations for this type of framework 40
7.2. Alignment with the ESMO ELCAP framework 42
7.3. Community-based validation: The Lerești Workshop (Romania) 43
8. AI-generated deepfakes and trustworthy cancer literacy tools 44
9. Co-creation and validation of the proposed recommendations 46
10. Recommendations for the Appropriate Use of Large Language Models to Enhance Cancer Literacy 48
10.1. Recommendations for citizens and patients 50
10.2. Recommendations for healthcare professionals 51
10.3. Recommendations for organizations 52
References 55
Appendix 67
Annex 1. Technical Workshop. Trust and usage of digital health solutions (including LLMs) 2025 67
Annex 2. Expert Focus Group 69
Annex 3. Recommendations Factsheet used for validation 72
Annex 4. Validation Results of the Recommendations Across Participant Groups 78
Frequently used AI terms (plain-language glossary) 81

List of Figures

Figure 1. Methodology used in creating the final list of recommendations.....14

List of Tables

Table 1 Classification of healthcare chatbots according to MDR risk classes 28
Table 2. Trustworthiness principles for cancer-focused LLMs 36
Table 3. Examples of approved AI tools across cancer type 38
Table 4. Classification of ELCAP applications according to user group and functional role 42
Table 5. Benefits and risks of traditional vs. AI-assisted health information access 45
Table 6. Final recommendations per target audience 49

List of Abbreviations

- AI: Artificial Intelligence
- ANI: Artificial Narrow Intelligence
- CHART: Chatbot Assessment Reporting Tool
- CRC: Colorectal Cancer
- DMCCA: Digital Markets, Competition, and Consumers Act 2024
- EADV: European Academy of Dermatology and Venereology
- ECII: European Cancer Imaging Initiative
- ECO: European Code of Cancer Practice
- EHDS: European Health Data Space
- ELF: European Lung Foundation
- ESMO: European Society for Medical Oncology
- EU: European Union
- FDA: Food and Drug Administration
- GPAI: General-Purpose AI
- GMAI: Generalist Medical AI
- EBAI: ESMO Basic Requirements for AI-based Biomarkers in Oncology
- GDPR: General Data Protection Regulation
- GQS: Global Quality Score
- IVDR: In Vitro Diagnostic Regulation
- JRC: Joint Research Centre
- LLM: Large Language Model
- MDR: Medical Device Regulation
- MDSW: Medical Device Software
- MedDTs: Medical Digital Twins
- MPNE: Melanoma Patient Network Europe
- NLP: Natural Language Processing
- OECD: Organisation for Economic Co-operation and Development
- RAG: Retrieval-Augmented Generation
- TAI: Trustworthy AI
- WHO: World Health Organization

Executive Summary

This report presents a systematic, evidence-based framework for the ethical, safe, and effective use of Large Language Models (LLMs) in cancer literacy. People are increasingly using (Artificial Intelligence) AI tools, particularly LLMs, as a first step when seeking medical advice. This growing reliance creates a clear need for guidance to prevent exposure to inaccurate information, overly complex explanations, or content that increases anxiety. It outlines the context and necessity for developing a comprehensive set of guidelines tailored to citizens, patients, healthcare professionals, and organizations to support the safe and effective improvement of cancer literacy. Grounded in expert input, real-world data, and established best practices, these guidelines were co-designed with multidisciplinary experts, then validated, to ensure they are scientifically robust, ethically sound, and applicable in real-world contexts.

Key findings:

1. **Ongoing gap in health literacy and AI readability:** Health and cancer literacy remain low across Europe, with around 46% of the EU population having limited health literacy. At the same time, publicly available/general-purpose LLMs outputs are written at a college reading level, far above the recommended 6th-grade level at high-school to college reading levels. This mismatch reduces the usefulness of AI tools for the people who may need them most.
2. **Real but uneven benefits in healthcare settings:** LLMs show potential for patient education, symptom explanation, and selected clinical workflows. However, performance remains variable and generally well below 80–99% for publicly available models, differing substantially by model, task, cancer type, language, and context. Therefore, LLM-generated information should not be interpreted as consistently reliable clinical guidance and requires appropriate verification, human oversight, and adaptation to the linguistic, cultural, and health-literacy context of the user. Human verification and clinical oversight therefore remain essential. However, performance varies across cancer types, languages, and cultural contexts.
3. **A developing but still incomplete regulatory framework in the EU:** Key regulations such as the EU AI Act, MDR/IVDR, GDPR, and EHDS are moving toward a combined compliance system for high-risk medical AI. However, in line with Trustworthy AI principles, important uncertainties remain for patient-facing AI tools, especially regarding legal responsibility, automation bias, and whether certain chatbots count as medical devices.
4. **Citizens want support, not surveillance:** The Lerești Living Lab workshop found citizens open to digital health tools but highly sensitive to anxiety triggered by

frequent or blunt AI feedback, wary of "robotization: and constant monitoring, and reliant on a trusted human (usually a doctor) to interpret AI outputs and act as a "trust anchor".

5. **People critical-evaluation skills are limited:** Expert consensus estimates that only a small minority of citizens and patients currently have the combined health and digital literacy needed to reliably distinguish trustworthy AI output from hallucinations or deepfakes, meaning guidelines must assume limited public literacy rather than compensate for it after the fact.

Taken together, these findings indicate that no single, generic disclaimer or training session is sufficient: safe adoption of LLMs in cancer literacy requires differentiated, role-specific guidance, transparent and trust-by-design AI tools, and continuous human oversight embedded at every level of use. Some recommendations include:

1. **Differentiate guidance by user group:** Guidance should not be one-size-fits-all. Citizens and patients primarily use LLMs to access, understand, and navigate cancer information; healthcare professionals may use them for evidence synthesis, patient communication, and selected clinical or administrative support; and organizations are responsible for safe implementation, governance, data protection, and oversight. Recommendations should therefore be tailored to each group, rather than presented as a single generic disclaimer.
2. **Keep humans at the center of every use case:** In all contexts, AI should remain a trustworthy, transparent, and human-supervised support tool, not an autonomous source of medical or practical advice. Human judgment, accountability, and ethical responsibility must remain central at every stage of use. Clinicians must retain final responsibility for clinical decisions, while patients should be consistently guided toward appropriate professional care whenever diagnosis, treatment, or other health-related decisions are involved.
3. **Strengthen public literacy through trusted intermediaries:** Safe and responsible use of AI is best supported through trusted sources such as clinicians, patient organizations, community leaders and other credible local networks. This approach should complement and not be replaced byechnical or digital literacy training, and more effective than relying solely on technical or digital literacy training.
4. **Keep the guidelines living and regularly updated:** Given the rapid evolution of AI technologies and regulation, these guidelines should be periodically reviewed, validated, and updated rather than treated as a fixed document.

This approach ensures that LLMs can improve access to reliable cancer information and support healthcare efficiency, while remaining a supervised and transparent tool that complements, not replaces, clinical judgment, patient autonomy, and evidence-based care.

Following approval, the Guideline will be made publicly available on the CURTAIN project's website to facilitate broad access and support its dissemination.

Introduction

Cancer remains a major and complex public health challenge in Europe, with well-documented and persistent differences in incidence, mortality, prevention, early detection, and treatment outcomes both between and within countries [1-3]. In parallel, people are increasingly turning to publicly available and easily accessible artificial intelligence (AI) tools, particularly large language models (LLMs), as an initial source of health information. While these systems are widely accessible and often easy to use, this growing reliance raises important concerns. There is a urgent need to protect, ensure that users are not exposed to inaccurate information, overly complex explanations, or emotionally distressing content. Current evidence shows a clear mismatch between the technical performance of LLMs and their demonstrated practical clinical value: although they can produce fluent and seemingly reliable responses, they do not yet demonstrate consistent, measurable improvements in patient outcomes, while hallucinations may remain difficult to identify and assess [13, 18, 19].

A key limitation is that AI-generated health information is often written at a reading level that is too complex for the general population, reducing its real-world usefulness [27]. In addition, these systems can generate plausible but incorrect statements, including content that resembles AI-generated "deepfake" misinformation [107-109]. In the multidisciplinary complexity of clinical oncology, inaccurate or incomplete LLM outputs can have serious consequences, such as delaying or bypassing appropriate medical assessment and consultation, influencing unsafe or context-inappropriate decisions, generating incorrect recommendations or increasing patient anxiety. In more complex clinical contexts, including decision support, LLMs may also omit relevant clinical details, leading to incomplete or misleading outputs. For these reasons, their use in healthcare requires careful framing, evidence verification and continuous human oversight.

Objective and scope

Rather than addressing these challenges and barriers through a single disclaimer or uniform approach, this document aims to adopt a structured, user-centered ethical framework – recognizes that citizens, healthcare professionals, and organizations engage with LLMs in fundamentally different ways and therefore require distinct types of guidance. Citizens need clear, reliable, and understandable health information, always accompanied by encouragement to seek professional medical advice. Healthcare professionals may use AI as a supportive tool for information synthesis, evidence retrieval, documentation, and selected decision-support tasks, but AI outputs must be independently verified and clinicians remain fully responsible for all clinical decisions.

Organizations, in turn, are responsible for ensuring that AI systems are implemented safely, transparently, and in accordance with ethical and regulatory requirements, with clearly documented limitations, disclaimers, accountability, and human oversight, while avoiding opaque or purely efficiency-driven applications in clinical settings.

Existing frameworks, such as the ESMO ELCAP [106] model, together with regulatory pathways for AI-enabled medical devices, support this approach by classifying AI tools according to their intended users and required level of human oversight. At the regulatory level, the European Union has established an evolving governance framework for AI in healthcare through the AI Act (2024) [48] and the 2026 update [134], the Medical Device Regulation [46], GDPR [51], and the European Health Data Space [50]. These instruments classifing healthcare AI applications for used for diagnostic and/or treatment decisions as high-risk and require safeguards such as human oversight, transparency, data protection, bias mitigation, and safety validation. However, important gaps remain in the application of these frameworks to generative, patient-facing AI in oncology, particularly regarding regulatory classification and medical-device status, allocation of legal and clinical responsibility, transparency of outputs and data sources, automation bias, protection of sensitive health data, performance validation across different populations and contexts, and continuous monitoring after deployment. These uncertainties are especially important when AI moves from general cancer information toward personalized interpretation, symptom assessment, or clinical decision support.

This deliverable (D3.3), part of Task 3.4 within Work Package 3 (Design of the cancer literacy tools and interventions) of the project, has four main objectives:

1. Establish the context by examining the global cancer burden, the current cancer literacy gap and inequalities, and the increasing role of Large Language Models (LLMs) as a source of cancer and health information.
2. Provide an evidence-based overview of LLMs in clinical oncology, including their underlying concepts, current applications, demonstrated benefits, limitations, and the ethical and regulatory requirements governing their use according to intended purpose, user group, level of clinical impact, and associated risk.
3. Identify key risks and future considerations, including inaccurate or unreferenced outputs, AI-generated misinformation, medical deepfakes, automation bias, privacy and data-governance risks, and the need Trustworthy AI principles, evidence verification, continuous human oversight, and regularly updated guidelines.
4. Develop and validate role- and purpose-specific recommendations for citizens and patients, healthcare professionals, and healthcare organizations through literature review, community engagement, and multidisciplinary expert consultation,

supporting safe, transparent, equitable, and human-supervised use of LLMs in cancer literacy.

Relation to other WPs and deliverables

Several sections of content structured in Deliverable D3.3 were directly informed by previous work done in Work Package 2, particularly Deliverables D2.1 and D2.2 that mapped the current European status of cancer literacy initiatives, barriers, facilitators and cancer-related disinformation narratives.

The present document is directly linked with the ongoing work done in Work Package 4, Task 4.4, Co-creation of real-life use framework for LLMs in cancer literacy, with citizens and patients. The guidelines here developed will serve as the backbone for developing the real-life use framework for LLMs in cancer literacy, ensuring that application is safe, trustworthy, context-sensitive, and adapted to different user groups and levels of risk.

At the same time, Deliverable 3.3 is closely linked to Work Package 6, particularly with the development of the Virtual Cancer Literacy Library, an ongoing process performed in Task 6.1. Based on the present guidelines, several informative sheets and pieces of content will be created to populate the Virtual Cancer Literacy Library, with the aim of helping citizens, patients, healthcare professionals and organizations to safely use LLMs to navigate cancer-related information.

The responsible uptake of LLMs also has implications for Europe's capacity to remain competitive in digital health innovation. While the EU is developing one of the most comprehensive regulatory environments for AI and health data, fragmented implementation, regulatory uncertainty, and uneven institutional readiness may slow the translation of trustworthy AI from research into real-world healthcare practice. CURTAIN can therefore contribute not only to safer AI use, but also to demonstrating how European values, through human oversight, transparency, evidence-based practice, data protection, and equity, can be translated into practical implementation models that support both public trust and responsible innovation.

Content of the deliverable

This deliverable is structured to move from background and evidence to a validated set of practical recommendations. After this introduction and the methodology used in outlining the guidelines, it first presents the global burden of cancer and key concepts related to AI and large language models (LLMs). It then reviews scientific evidence on the performance of LLMs and chatbots in cancer education and clinical decision support.

Next, it outlines the relevant European and international regulations and clinical guidelines for the use of AI in oncology. The deliverable then addresses the risks of medical deepfakes and AI-driven misinformation, as well as the role of trustworthy AI tools in helping to reduce these risks. This is followed by the development of the guideline framework, which was created and validated through a citizen workshop in Lerești and a multidisciplinary expert consultation, covering citizens, healthcare professionals, and organizations.

The deliverable concludes with the final set of co-developed recommendations, a glossary of key terms, references, and detailed annexes.

Methodology

The methodological approach combines four components: an integrative literature and policy review, community-based qualitative validation, and structured expert consultation and a final multi-stakeholder validation. This approach was adopted because the intersection of LLMs and cancer health literacy involves highly complex clinical, regulatory, and socio-cultural dimensions, requiring the concurrent integration of empirical clinical evidence, evolving legal frameworks, theoretical governance models, and primary qualitative user data.

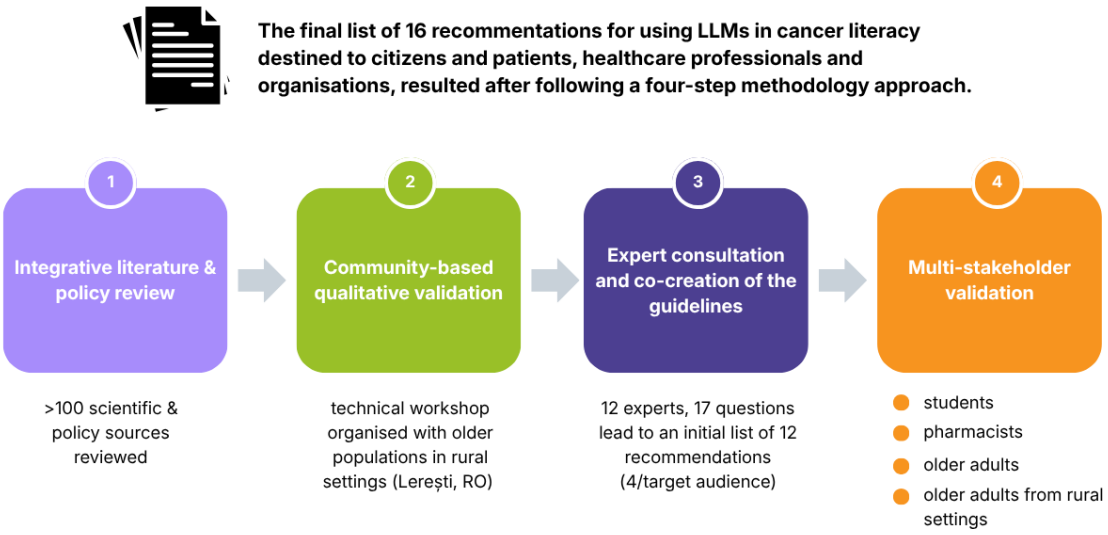

Figure 1. Methodology used in creating the final list of recommendations

Integrative review

The first phase identified a critical gap in guidance for the safe and effective use of LLMs in cancer literacy, in the context of increasing AI adoption. It was informed by WP2 findings, particularly Deliverable D2.2 ("Regional and platform-specific patterns of cancer-related disinformation"), which provides the evidence base on the dissemination of cancer misinformation across food supplement, natural remedy, and institutional distrust narratives, as well as across platforms such as TikTok, YouTube, Telegram, and Facebook.

These findings highlighted significant disparities in cancer literacy across countries, particularly linked to age, education, socioeconomic status, and healthcare experience, with key difficulties in understanding risk factors, screening, and healthcare navigation. A targeted literature search was then conducted, focusing on studies relevant to evaluating the outcomes, risks, and appropriate use of LLMs in healthcare, rather than providing an exhaustive review of all available AI literature. The evaluation stage synthesized three main evidence domains: empirical studies on LLM performance in healthcare (including readability, accuracy, cultural adaptation, and clinical support functions); regulatory and legal frameworks within the EU (including the AI Act, MDR, GDPR, and EHDS); and international ethical and implementation guidelines from organizations such as WHO and OECD.

Community-based qualitative validation

A technical workshop was held in November 2025 in Lereşti Living Lab, Argeş County, Romania. Thematic analysis of the workshop findings informed key refinements to the guidelines, highlighting the community's heightened sensitivity to "anxiety loops" associated with frequent health monitoring and concerns regarding "robotization" of care. A central finding was the importance of healthcare professionals serving as trusted intermediaries between LLMs and citizens.

Expert consultation and co-creation of the guidelines

A focus group involving 12 experts from all sectors of the penta-helix model was conducted to co-develop the framework required for the safe integration of AI into cancer care, with a focus on enhancing patient cancer literacy while mitigating risks related to misinformation and deepfakes. The consultation was guided by 17 structured questions, developed based on scientific literature, real-world observations of AI use, and insights from the Lereşti Living Lab. Questions were organized across three implementation levels: Citizens and Patients, Healthcare Professionals, and Organizations. Responses were collected through interactive polling and open discussion and analyzed thematically.

Multi-stakeholder validation

The recommendations developed through the previous steps were finally validated through a survey targeting four stakeholder groups: (i) older adults, (ii) older adults living in rural areas, (iii) pharmacists, and (iv) university students. These groups were selected to include participants with diverse backgrounds, experiences, and information needs in order to cover possible blind spots.

The recommendations were translated into Romanian to make them easier for participants to read and evaluate. A modular questionnaire was then designed to assess the quality and usefulness of the recommendations. The questionnaire asked participants to evaluate several aspects, including accessibility, readability, comprehensibility, relevance, and applicability with a mixed, quantitative and qualitative approach.

This approach made it possible to identify both the strengths of the recommendations and areas that could be improved. Each stakeholder group received a validation package that included the questionnaire, a factsheet with the relevant recommendations, and a short summary of the guidelines. For the two community groups (older adults and older adults living in rural areas), only the recommendations intended for citizens and patients were provided, together with the corresponding questionnaire. Pharmacists and university students received the full validation package. The feedback collected from all participants was included in a report that was later used to refine the recommendations and improve the final version of the guidelines before publication.

1. Cancer burden and AI fundamentals

1.1. Global cancer burden

Cancer remains one of the most pressing public-health challenges of the twenty-first century. New cancer cases are projected to reach 30.5 million by 2050, with an estimated 18.6 million deaths, driven by population aging, growth, and persistent exposure to preventable risk factors. The burden is expected to fall disproportionately in low- and middle-income countries, where limited resources hinder early detection and equitable access to treatment [1].

Despite the known role of modifiable risk factors, a major “prevention gap” persists between scientific knowledge and the population’s ability to effectively reduce risk. This gap is particularly evident in the European Union (EU), where non-communicable diseases remain the leading cause of preventable death. In 2022, over 725,000 deaths

were considered avoidable through stronger public health measures, with cancer as the leading contributor. Lung cancer alone accounts for 18% of preventable mortality, highlighting the limited reach of traditional prevention strategies among high-risk groups [2].

The economic and social burden is substantial, with 2.4 million potential productive life years lost in 2022, further exacerbating pressure on the EU workforce amid demographic decline. Recent Joint Research Centre (JRC) data report 2.7 million new cancer cases in 2024 and confirm persistent geographic inequalities across Europe. Higher incidence in Northern and Western Europe contrasts with significantly higher mortality in Central and Eastern Europe, driven by systemic differences in screening access, care continuity, and health system capacity, alongside socio-economic disparities. Colorectal cancer (CRC) exemplifies these challenges, with increasing incidence and mortality projections, including a rise in early-onset cases. Together with lung, breast, and prostate cancers, these trends have prompted expanded EU screening initiatives aimed at improving early detection and reducing inequities in outcomes across Member States [3].

1.2. Artificial Intelligence and Large Language Models concepts

Artificial intelligence (AI) refers to computer systems designed to perform tasks that typically require human intelligence, including understanding language, recognizing patterns, solving problems, and learning from experience. Over the past decade, AI has evolved from rule-based automation systems to adaptive, data-driven systems capable of learning from large and complex datasets [4].

- **Functional classification:**

AI can be functionally classified into two main categories relevant to healthcare and research applications:

1. **Artificial Narrow Intelligence (ANI)** [5] refers to systems designed for specific, well-defined tasks, such as identifying features in medical images, translating languages, or summarizing documents. Although limited in scope, ANI systems are highly effective within their specific functions and are generally easier to validate and regulate.
2. **General-Purpose AI (GPAI)** [6,8] refers to advanced and flexible AI systems that can process, understand, and connect different types of information and perform a wide range of tasks. These systems are trained on very large amounts of data, allowing them to adapt to different tasks without needing to be specifically retrained for each one. Under the EU regulatory framework, an AI system is

considered GPAI when it is trained using substantial computational resources and is capable of performing a broad range of tasks. In contrast, AI systems designed for a single or narrowly defined purpose are not generally considered GPAI. In healthcare, GPAI can be particularly valuable because it can bring together complex medical information from different sources and support healthcare professionals in clinical decision-making.

3. In healthcare, an additional emerging concept is **Generalist Medical AI (GMAI)** [7]: foundation-model-based systems designed to perform multiple medical tasks and potentially integrate different clinical data modalities. GMAI remains an evolving paradigm and should not be considered equivalent to validated clinical AI. Its safety, regulatory status, and level of required human oversight depend on the intended purpose and the specific healthcare context in which it is deployed.

4. Many contemporary LLMs are examples of generative general-purpose AI models; however, their general capabilities do not imply clinical validity. When incorporated into healthcare applications, regulatory and safety requirements depend on the intended purpose, user group, data processed, clinical impact, and level of autonomy.

- **How Large Language Models work:**

Large Language Models (LLMs) [9,10] are a core component of generative AI systems, trained to process and generate human language by learning statistical relationships between tokens and predicting contextually probable sequences of tokens. Trained on vast corpora of text, including books, scientific articles, and web content; they can perform tasks such as summarization, translation, question answering, and reasoning. LLM development typically involves pre-training followed by optional post-training or fine-tuning, while prompting occurs during model use (inference) to guide the generated output:

- **Pre-training:** The model is trained on very large datasets to learn statistical patterns, linguistic structures, contextual relationships, and associations between tokens. This provides broad language capabilities before adaptation to specific tasks or domains.

- **Fine-tuning:** The model can be further trained on specific datasets for a certain domain, such law or medical, following pre-training. This enables it to provide more precise and pertinent answers in that area.

- **Prompting** is the process by which users instruct the model by providing it with directives, or prompts. There are various approaches to prompting:
  - *Zero-shot:* When facing a task, the model attempts to respond without any examples. It is dependent on the knowledge it acquired prior to instruction.
  - *Contextual/few-shot learning*: The prompt provides the model with a few instances to help it grasp the specific type of response that is desired.
  - *Chain-of-thought:* The prompt asks the model to structure a complex problem into intermediate steps, which can improve performance for some multi-step tasks. However, generated explanations should not be assumed to represent the model's actual internal reasoning or to guarantee the correctness of the final answer.

- **Transformers:** The majority of LLMs are based on the transformer architecture. Transformers give the model the ability to comprehend the connections between every word in a sentence, regardless of how far apart they are. This facilitates the model's effective understanding of text's dependencies, meaning, and context.

Retrieval-Augmented Generation (RAG) [11] combines LLM generation with retrieval of information from external knowledge sources. Relevant information is retrieved and provided to the model as context, which can improve factual grounding and access to more current or domain-specific knowledge. In healthcare, its reliability depends strongly on the quality, relevance, currency, and traceability of the retrieved sources. RAG can reduce, but does not eliminate, hallucinations, retrieval errors, or misleading synthesis; therefore, medical outputs still require source verification and appropriate human oversight.

1.3. Prompting engineering as a core competency in medical AI literacy

Prompt engineering [12] refers to the practice of designing clear and precise instructions to guide LLM outputs. Effective prompting is an important component of AI literacy for healthcare professionals, educators, researchers, and institutions, supporting tasks such as information synthesis, evidence retrieval, patient communication, research, education, administrative activities, and selected decision-support applications.

However, LLMs have important limitations. LLMs may lack access to current, complete, or patient-specific information unless they are appropriately connected to validated external sources or authorized clinical systems. They may also generate inaccurate, incomplete, outdated, or fabricated information. Access to additional data does not itself guarantee clinical accuracy or safety. Outputs used for clinical interpretation, patient-

specific recommendations, or decision support therefore require verification against authoritative sources and appropriate review by qualified healthcare professionals.

Effective prompt engineering includes strategies such as being specific, providing context, clear objectives and boundaries, requesting structured outputs and supporting sources where appropriate, and instructing the model to acknowledge uncertainty or missing information [13]. Techniques such as few-shot prompting and iterative refinement can further improve output quality. Few-shot prompting and iterative refinement may improve output relevance and structure, but they do not guarantee factual or clinical accuracy. Sensitive or identifiable patient information should not be entered into publicly available or non-authorized AI systems.

As AI systems evolve, effective prompting is becoming part of a broader competency in human–AI interaction that includes appropriate tool selection, contextualization, source verification, privacy awareness, recognition of uncertainty, and continuous human oversight [14].

1.4. Role of LLMs in cancer literacy

LLMs have the potential to support healthcare delivery in oncology across three main domains: clinical decision-making, patient education and engagement, and access to oncology research [13].

In clinical decision-making, LLMs have the potential to support cancer literacy and oncology information use across several interconnected domains: citizen and patient education and navigation, healthcare professional communication and evidence synthesis, access to oncology knowledge and research, and selected clinician-supervised decision-support applications.

Despite their potential, LLMs present limitations including variable performance across cancer types, bias risks, misinformation, and accountability concerns. Their safe use requires governance frameworks, patient consent, and integration into clinician-supervised workflows.

A European Commission report on AI in healthcare [14] highlights that approximately 97% of health data remains underutilized. AI systems show promise in oncology, rare disease diagnostics, radiology, and healthcare optimization. Applications include multimodal data integration for precision oncology, tumor detection in imaging, and improved allocation of healthcare resources to underserved populations.

Importantly, the potential value of LLMs for cancer literacy should be assessed according to who is using the system, for what purpose, with which data, and with what potential consequences. A low-risk use such as explaining general cancer terminology is fundamentally different from patient-specific interpretation, symptom assessment, or treatment decision support. The required safeguards, evidence, validation, and level of human oversight should increase accordingly.

1.5. Medical Digital Twins and personalized AI

Medical Digital Twins (MedDTs) [15] represent an emerging paradigm in precision medicine. An MDT is a dynamic, continuously updated virtual representation of a patient that integrates multimodal data to simulate health and disease processes. The MedDT framework consists of five components: the patient as the source of multimodal data (electronic health records, pathology, molecular/genomic data, imaging, laboratory data, patient-reported outcomes, and potentially wearable data), a data integration layer, a computational patient model, a clinician interface, and a synchronization mechanism that continuously updates the virtual model.

Although promising, MedDTs remain predominantly an emerging research and translational concept, with important challenges related to clinical validation, interoperability, data quality, uncertainty quantification, governance, privacy, and regulatory implementation.

This approach enables personalized simulation of treatment strategies. For example, in oncology, MedDTs are being explored as a means of simulating disease trajectories and comparing potential responses to different treatment strategies, with the longer-term goal of supporting more personalized treatment planning.

LLMs may potentially serve as an interface layer within these systems, helping clinicians and patients interrogate complex multimodal information and translate model outputs into understandable language. However, an LLM interface does not validate the underlying model or its predictions, and generated explanations still require appropriate verification and human oversight. From a cancer-literacy perspective, future MedDT systems may help make complex personalized information more understandable, provided that uncertainty, data provenance, and the distinction between explanatory information and validated clinical recommendations remain transparent.

2. Cancer health literacy as predictor of outcomes and European Initiatives

Before examining what LLMs can offer, it is worth grounding the discussion on why cancer literacy matters so much in the first place, both for individual outcomes and for the sustainability of health systems.

Persisting gaps in cancer literacy limit patients' ability to make informed decisions despite advances in oncology, undermining shared decision-making, and making it hard for patients to interpret complex information on diagnosis, prevention, and treatment. Low general health literacy affects 46% of Europeans [16].

Personal cancer literacy is an individual's capacity to access, understand, evaluate and use information related to cancer prevention, diagnosis, treatment and care to make informed decisions; organizational health literacy refers to how effectively healthcare organizations and systems make information, services and navigation accessible, understandable and usable for people with different literacy needs [17]. Higher cancer literacy empowers patients to engage actively with clinicians asking informed questions and weighing options, rather than passively following instructions. Literacy gaps often mirror social inequalities: patients with lower education, lower income, minority status, or rural residence tend to have less ability to navigate cancer information [18, 19].

European organizations are embedding literacy into care: the European Code of Cancer Practice (ECO) defines patient rights on information and participation, [20] while ESMO Patient Guides [21] and My Cancer Navigator [22] further support access to understandable, evidence-based cancer information across Europe.

3. Applications of LLMs in health and cancer literacy

The digital era creates tension between rapid misinformation spread and the opportunity to disseminate accurate knowledge [23]. While AI integration brings specialist-level expertise to clinical practice, especially imaging and diagnostics, concerns are growing about dehumanized medicine, weakened patient-doctor trust, and safety risks from the swift adoption of LLMs.

The GRADE approach provides an established methodological framework for translating clinical questions and research evidence into transparent evidence assessments and recommendations [24,25]. The official GRADE Book, developed by the GRADE Working Group, structures this process around answerable clinical questions, assessment of certainty of evidence, evidence synthesis and Evidence-to-Decision frameworks. These principles are increasingly relevant to the development of AI-supported clinical decision-support systems, where transparency of the evidence underlying generated outputs is particularly important.

An emerging example in oncology is the A.P.C.I. (Ambient Patient-Clinical Intelligence) system within the PrOPA® framework, which applies standalone LLM processing within a modified GRADE-based and PICO-structured approach. Patient-specific clinical information is transformed into structured clinical questions, while relevant evidence and guideline information are retrieved and processed to support evidence-informed diagnostic and therapeutic assessment. Rather than treating the LLM output itself as evidence, the approach aims to preserve traceability between the patient-specific question, retrieved scientific evidence, assessment of evidence certainty, and the resulting decision-support output. The system is currently undergoing retrospective clinical validation against multidisciplinary expert decisions [26]; therefore, its clinical performance should be considered under evaluation rather than established.

Importantly, applying a modified GRADE framework to an LLM does not itself validate the AI-generated recommendation; GRADE assesses the certainty of the underlying body of evidence and supports transparent decision-making, while the performance, safety, reproducibility and clinical validity of the AI system require separate evaluation.

3.1. Readability and quality assessments

A 2025 study evaluated ChatGPT's responses to 57 common cancer-patient questions (from a 508-patient survey on colorectal, breast, lung, and prostate cancers) using Flesch Reading Ease, grade-level scales, DISCERN, and Global Quality Score (GQS). Responses required college-level reading (mean Flesch Reading Ease 48.18, mean grade-level 13.21), well above the recommended 6th grade, though overall quality was good (DISCERN 51.98, GQS 3.91 near-good). Breast cancer responses were slightly easier to read and higher quality than other cancers; responses often lacked references, and only a few included disclaimers advising consultation with a healthcare provider. The study concluded that ChatGPT can serve as a useful informational companion but must be used cautiously given its reading-level demands and variable completeness [27].

3.2. Performance in culturally specific contexts

Recent research in Romania has explored the potential of large language models (LLMs) to improve cancer literacy and patient education, focusing first on prostate cancer. The initial study evaluated ChatGPT 3.5 by having eight Romanian oncology experts assess its performance in answering 25 validated questions based on the official Romanian Patient Guide for prostate cancer. The evaluation criteria included accuracy, timeliness, comprehensiveness, and readability. The study found that ChatGPT 3.5 consistently provided more precise, detailed, and understandable information than both other AI models and the official patient guide. This demonstrated that, even in a non-English, culturally specific context, advanced LLMs can enhance the accessibility of complex medical information for patients and non-specialist healthcare providers. Notably, the study highlighted that ChatGPT 3.5 was particularly strong in explaining medical terms in plain language, structuring responses logically, and providing comprehensive answers that often exceeded the scope of the official guide [28].

Building on these findings, a subsequent study compared multiple LLMs (including ChatGPT 3.5, CoPilot, and Gemini) to evaluate their relative performance in delivering cancer-related information in Romanian. The results confirmed ChatGPT 3.5 as the most accurate and informative model, outperforming CoPilot, which showed slightly lower coverage and clarity, and Gemini, which tended to produce neutral or incomplete answers. The study emphasized that the success of LLMs depends not only on their underlying architecture but also on their adaptation to local languages, cultural context, and healthcare guidelines. Additionally, the research discussed the importance of ongoing oversight and regulation of AI in healthcare, referencing European frameworks such as the EU Artificial Intelligence Act, which mandates transparency, bias mitigation, and human oversight. Overall, these studies indicate that well-trained and culturally adapted LLMs can serve as reliable tools for enhancing patient understanding, supporting shared decision-making, and complementing traditional educational resources, while highlighting the need for careful implementation under expert supervision [29].

3.3. Performance in clinical decision-support

An evaluation of ChatGPT-3.5 in breast cancer tumor boards found its treatment recommendations matched expert physician decisions in 70% of cases (7 of 10 patients), performing especially well at summarizing clinical cases and explaining medical reasoning, as confirmed by high scores from independent radiologists. However, in one case the model omitted key diagnostic data (HER2 FISH results) that could have led to inappropriate treatment advice, and it systematically excluded radiologist input from its recommendations, and did not include radiologist involvement in its multidisciplinary recommendations, which the authors suggested might reflect limitations or biases in the

information represented in its training data. This small proof-of-concept study suggests that LLMs may support case summarization and explanation but does not establish their reliability for treatment decision-making [30].

Emerging research is also exploring multi-agent or agentic AI architectures for complex clinical tasks. One preprint evaluating the Doctronic [31], multi-agent system across 500 virtual urgent-care encounters reported 81% diagnostic concordance with clinicians and 99.2% treatment-plan alignment, with no clinical hallucinations identified according to the study's predefined assessment [32]. These findings should be interpreted cautiously, as the study has not undergone peer review, was conducted outside oncology, and all authors reported equity ownership in the company developing the system. It therefore illustrates an emerging technical approach rather than established evidence for oncology decision support.

These readability, cultural-adaptation, and decision-support findings set the stage for Chapter 4's closer look at AI-based chatbots specifically, where similar patterns of promise and risk recur in patient-facing form.

4. AI-Based Chatbots in cancer education and care

Chatbots occupy a distinct place within the broader LLM landscape: ather than supporting clinicians directly, they interact with patients and the public in real time, making their design, oversight, and regulatory classification especially consequential for cancer literacy.

4.1. Literature evidence from recent studies

Chatbots simulate human conversation to support patient education, screening, treatment support, symptom management, and survivorship [33]. A 2023 review of 21 studies found most patients reported positive experiences with usability, convenience, and personalized communication; chatbots improved knowledge, adherence, symptom monitoring, and decision-making, sometimes outperforming standard care across breast, ovarian, cervical, colorectal, prostate, lung, and head-and-neck cancers, delivered via text, web, or voice platforms. Benefits include improved access for underserved populations and reduced clinician workload; limitations include early-stage inaccuracies, training-data bias, limited voice interfaces, privacy concerns, and regulatory gaps [34].

A broader systematic review of 137 generative-AI health-chatbot studies found major inconsistencies in design and reporting: most focused on closed-source models like ChatGPT without documenting version, technical settings, or prompts, making replication difficult, and fewer than a third addressed ethical, safety, or regulatory considerations despite the high-stakes healthcare context. The review calls for standardized reporting frameworks and proposes a Chatbot Assessment Reporting Tool (CHART) to support consistent evaluation and responsible integration [35]. A 2025 article similarly frames chatbots as promising for oncology education, delivering personalized, real-time information to patients, caregivers, and clinicians, but flags risks around convincing human mimicry, misinformation, undisclosed AI-nature, privacy breaches, and reliance on proprietary datasets, recommending robust oversight, transparent algorithmic development, regularly updated data, and safety guardrails [36].

4.2. Evidence in specific cancers

- **Prostate cancer:** PROSCA, an evidence-based NLP chatbot developed to complement standard patient information, initially demonstrated high usability and perceived knowledge gains in a small proof-of-concept study [37]. A subsequent randomized controlled trial involving 112 men found a significantly greater reduction in information needs among patients offered PROSCA alongside standard information compared with standard information alone, while also reporting high usability and willingness to reuse the tool. These findings support its potential as an adjunct to, rather than a replacement for, clinician–patient communication, although the evidence remains based on a specific single-center prostate cancer setting [133].
- **HPV and cervical cancer**: VaxBot-HPV is a chatbot developed to improve health literacy and promote HPV vaccination uptake. It was fine-tuned on a knowledge base of 451 documents, which were distilled into 202 question-answer pairs and their supporting context, along with 39 questions generated by GPT-4 used for training and testing. In the study's evaluation, VaxBot-HPV reported higher answer relevancy and faithfulness than the GPT-3.5 and GPT-4 comparators. The findings demonstrate that fine-tuning mitigates hallucinations and misinformation while allowing the model to provide explanations and reasoning for its answers, ultimately enhancing transparency and user trust in public health communication [38]. A 2025 trial of an AI chatbot promoting HPV vaccination among parents of 12–15-year-old girls in China (2,671 participants randomized to chatbot or usual-care control) found vaccine scheduling rose to 7.1% in the chatbot group versus 1.8% in controls over two weeks, alongside improved parental knowledge and misinformation-detection ability, with consistent effects across urban and rural

population, though overall vaccine confidence did not significantly shift, warranting further research on long-term scalability [39].

- **Breast cancer:** A 2025 randomized controlled trial involving 122 women with breast cancer evaluated structured AI chatbot-based education as an adjunct to standard care and reported significantly improved knowledge, perceived empowerment, and attitudes toward AI compared with standard care alone. These findings support the potential of chatbots as supplementary educational tools, although they do not establish clinical effectiveness or safety for diagnostic or treatment decision-making [40].

4.3. Empathic communication

A meta-analysis of 15 studies (2023–2024) examined whether GPT-3.5/GPT-4 chatbots could match or surpass human clinicians in empathic communication across oncology, neurology, mental health, rheumatology, and dermatology, with empathy assessed by proxy raters (clinicians, students, laypeople) rather than patients themselves. AI chatbots were frequently rated more empathetic than medical experts (GPT-4 consistently outperforming the more inconsistent GPT-3.5) likely reflecting AI responses' tendency to be longer, more structured, and encouraging in tone. Dermatology was the only specialty where humans occasionally outperformed AI. Limitations include unvalidated single-item empathy scales, observer- rather than patient-rated assessments, exclusion of non-verbal cues in text-only interactions, and no joint evaluation of empathy alongside accuracy; future research should incorporate multimodal interactions and direct patient perceptions, especially of whether awareness of "talking to AI" changes perceived empathy [41].

4.4. Regulatory standards, ethical considerations, and compliance challenges for AI-based chatbots (EU context)

AI chatbots are increasingly used to support interactions between patients and healthcare systems, raising ethical concerns related to decontextualized care, overlooked emotional and social factors, and an "illusion of certainty" surrounding AI-generated responses. These challenges may affect patient autonomy, shared decision-making, and quality of care. Accountability for chatbot-related harm remains unclear within the EU, particularly regarding the responsibilities of developers, providers, and healthcare institutions. Although the AI Act requires human oversight for high-risk systems, automation bias remains a concern. While the GDPR establishes requirements for lawful processing, transparency, data minimization, security, and protection of health data, it does not by itself address the full range of clinical safety, medical-device qualification, performance, and accountability issues arising from conversational AI [42].

Whether an AI chatbot qualifies as Medical Device Software (MDSW) under Regulation (EU) 2017/745 depends primarily on its intended purpose. Software intended to provide information used for diagnostic or therapeutic decisions may qualify as MDSW and is classified according to MDR Rule 11, with the risk class determined by the potential clinical consequences of the decisions it informs [43–45]. Software whose intended purpose is limited to general information, communication, or administrative support may fall outside the MDR." In contrast, chatbots providing only general educational information without analysing individual patient data are not considered medical devices, such as the Geneva University Hospitals' *conflAnce* chatbot. Chatbots analysing laboratory or diagnostic data may instead fall under the In Vitro Diagnostic Regulation (IVDR 2017/746) [44, 46].

Table 1 Classification of healthcare chatbots according to MDR risk classes

Class	Description	Example	Certification requirement
Outside MDR	No intended medical purpose; general information, communication or administrative functionality	General cancer education, appointment information, administrative navigation	MDR medical-device conformity assessment not applicable
I	MDSW not falling within the higher-risk provisions of Rule 11	Certain medical-purpose software with functions not used for diagnostic/therapeutic decisions or physiological monitoring	Generally manufacturer self-declaration of conformity*
IIa	Provides information used for diagnostic or therapeutic decisions; or monitors physiological processes	Clinical decision-support information where an incorrect decision would not normally cause serious deterioration, surgical intervention, death or irreversible deterioration	Notified Body conformity assessment
IIb	Diagnostic/therapeutic decisions may cause serious deterioration or require surgical intervention; or monitoring of vital parameters where	Higher-impact treatment or monitoring decision support	Notified Body conformity assessment

	variation could create immediate danger		
III	Diagnostic/therapeutic decisions may result in death or irreversible deterioration of health	Decision-support software informing potentially life-critical diagnostic or therapeutic decisions	Notified Body conformity assessment with the highest level of MDR scrutiny

Medical AI chatbots require clinical evaluation, technical documentation, CE marking, and ongoing post-market surveillance. Manufacturers must maintain an appropriate quality-management system under the MDR/IVDR; harmonized standards such as EN ISO 13485 may be used to support demonstration of conformity. The EU AI Act complements the MDR and IVDR through a risk-based framework. Under Article 6(1), AI-enabled medical devices, including chatbots used for diagnosis, monitoring, or treatment-qualify as high-risk AI systems when they are subject to third-party conformity assessment. Together with the MDCG 2025-6 guidance, this framework requires high-quality training data, human oversight, cybersecurity, traceability, continuous monitoring, and appropriate user training throughout the AI lifecycle [47, 48]. Following the 2026 Digital Omnibus update, general transparency obligations for these chatbots apply from 2 August 2026, while the broader high-risk compliance requirements are deferred to 2 August 2028 [134].

For cancer literacy, regulatory status depends on the chatbot's intended purpose. Chatbots providing only general educational information on cancer prevention, screening, or treatment are not considered medical devices. However, once they analyse individual symptoms, health data, laboratory results, or imaging findings to generate patient-specific recommendations, they become Medical Device Software and are subject to both MDR and AI Act requirements. Having established how AI chatbots are regulated, the next chapter broadens the discussion to the wider European and international governance framework for AI in cancer literacy.

5. Existing EU and Global Regulatory Frameworks

5.1. European Union frameworks

1. **EU Artificial Intelligence Act (2024).** The world's first comprehensive AI-specific legal framework mandates bias management, transparency, documentation, and human oversight through a four-tier risk system: Unacceptable Risk (prohibited social scoring, manipulative AI, workplace emotion recognition, biometric scraping),

High-Risk (health, education, employment, justice), Limited Risk (transparency duties, e.g. chatbots, deepfakes), and Minimal Risk (largely unregulated, e.g. video games, spam filters). Obligations fall primarily on developers and providers, extending to non-EU entities whose systems are used within the Union, requiring continuous lifecycle risk management, unbiased and representative training data, comprehensive technical documentation and performance transparency, and clear human oversight mechanisms ensuring clinicians retain the ability to interpret or override AI decisions. Crucially, a formal third-party conformity assessment is mandatory before any AI system can be marketed or deployed in the EU, demonstrating proof of accuracy, robustness, and cybersecurity [48].

2. **EU Ethics Guidelines for Trustworthy AI (2019).** Built on four foundational principles: respect for human autonomy (AI should support, not replace, human decision-making), prevention of harm, fairness, and explicability, with seven operational requirements spanning human oversight to technical robustness and safety. The guidelines pay special attention to vulnerable populations (children, persons with disabilities, disadvantaged groups) and imbalanced power relationships, recommending continuous testing and active stakeholder involvement throughout development. [49]

3. **European Health Data Space (EHDS).** In force since March 2025 with phased implementation through 2031, EHDS pursues better healthcare (patient-controlled cross-border data sharing with providers) and better research (secure, governed access to anonymized health data for innovation). It establishes a harmonized single market for Electronic Health Record systems with standardized technical specifications; implementation phases run March 2027 (implementing rules), 2029 (Patient Summaries/ePrescriptions), 2031 (images/lab results), and 2034 (HealthData@EU cross-border access). Data protection builds on GDPR with health-specific safeguards, explicitly prohibiting detrimental decisions based on health data analysis and requiring mandatory EHR certification with full transparency on data access. [50]

4. **Patient perspectives on data governance**. A 2024 survey of 220 participants across 28 EU countries found 90% GDPR awareness but low practical engagement; most had never exercised rights like access or erasure. Patients preferred tools enabling direct, individual control (granular consent, data portability) over collective mechanisms delegating decisions to trusted organizations, though willingness to share data altruistically rose substantially when patient representatives participated in governing organizations' decision-making. Researchers caution that pure individual control can become burdensome and risky, recommending better-

balanced EHDS provisions, stronger transparency mechanisms, and more meaningful patient involvement in data governance structures [51].

5. **EU Living Guidelines on Generative AI in Research (2026)**. Published by the European Commission in May 2026, this non-binding ERA Forum framework, anchored in the European Code of Conduct for Research Integrity, is designed to adapt continuously to technological change and emphasizes four core principles: reliability in verifying AI outputs, honesty through transparent disclosure, respect for privacy and intellectual property, and ultimate human accountability for all research outputs. Researchers hold ultimate responsibility for scientific outputs: AI cannot be credited as author or co-author, and must critically verify against training-data bias, sycophantic prompt bias, invented citations, hallucinations, and plagiarism, while accounting for AI's stochastic nature, which can produce different outputs from identical prompts and challenge reproducibility. They must transparently disclose AI use, share prompts and outputs following FAIR principles where possible, avoid uploading sensitive data to external systems without consent, and refrain from using generative AI for sensitive tasks like peer review. Research organizations must foster an environment where disclosure carries no penalty, provide training on verifying outputs, and where possible host their own secure, locally governed AI tools; funding bodies must mandate applicant transparency, prohibit AI from evaluating scientific content, and guard against deceptive practices such as hidden prompts embedded in application files. All stakeholders: researchers, institutions, and funders, are also required to actively minimize the environmental footprint associated with generative AI use [52].

5.2. Global and International frameworks

1. **World Health Organization (WHO).** WHO's ethical guidance rests on six principles: protecting autonomy and promoting wellbeing, safety and the public interest, transparency and explainability, responsibility and accountability, inclusiveness and equity, and responsiveness and sustainability. [53] It identifies applications spanning clinical diagnosis and care delivery to patient-guided self-management, administrative automation, medical education, and research, while flagging risks including overreliance from overestimated model reliability, embedded training-data bias, inaccurate medical information, privacy concerns, and cybersecurity threats. For health literacy specifically, WHO recommends a structured three-step approach: Development (training LLMs on high-quality, representative healthcare data), Provision (careful adaptation and rigorous validation of outputs for target audiences, checking for accuracy and clarity), and Deployment (comprehensive user guidance on safe usage, system limitations, and

verifying information authenticity), underpinned throughout by clear liability frameworks and robust oversight mechanisms to manage potential errors or misinformation.

A WHO European-region readiness report found 66% of Member States have a general AI strategy but only 8% a health-specific one, with oversight generally managed by existing bodies rather than new agencies; 78% of national strategies name health a priority, though many lack the funding to move from aspiration to implementation. AI adoption is driven by patient-care improvement (98% of respondents) and workforce relief (92%), concentrated in diagnostics (64%) and patient chatbots (50%). Yet only 24% of countries offer AI training for professionals and just 20% integrate it into pre-service medical education; stakeholder engagement (72%) often excludes patients and the public, risking overlooked user needs. Only 8% of countries have clear liability standards and 6% specific rules for generative AI, while legal uncertainty (86%) and cost (78%) remain the top adoption barriers, though 92% agree clear liability rules are crucial and 90% stress transparency guidance for building trust. [54] A companion country-profiles report covering 50 countries confirms uneven progress: strong government/provider engagement (81% /75%) but weak patient involvement (22%); France, Finland, and Estonia lead on data-hub governance and diagnostics/chatbot deployment, Denmark and Spain already integrate chatbots into patient care, while Romania, North Macedonia, Poland, and Czechia cite financial, infrastructure, or early-stage pilot constraints [55].

2. **OECD.** The OECD AI Principles (*updated May 2024*) promote inclusive growth and wellbeing, safeguarding human rights and democratic values, transparency and explainability, and robust, secure, accountable AI systems. A December 2025 OECD report notes cardiovascular disease's combined health/informal-care costs in the EU are sometimes double cancer's, benchmarking that shows costs for major conditions extend far beyond direct hospital treatment, framing AI adoption in oncology as a necessary strategy to offset a comparably massive economic and social burden, not merely a technical upgrade [56].

3. **FDA, Health Canada, and UK MHRA.** Their joint Good Machine Learning Practice principles cover the AI/ML device lifecycle: engaging multidisciplinary experts throughout development, robust software engineering and security protocols, training data representative of the intended patient population, strict separation of training and testing datasets to prevent overfitting, credible reference-standard validation, human-AI interface design focused on collaboration, real-world testing beyond theoretical performance, clear communication of capabilities and limitations to users, and continuous post-deployment monitoring with careful management of model updates [57].

4. **American Hospital Association.** Its "AI Action Plan" framework rests on three pillars: people (strong leadership, change management, and collaboration among clinical, administrative, technical, and operational teams), process (structured channels for idea submission, transparent vetting criteria, and centralized capabilities to scale successful pilots), and technology (data stewardship, governance, modern IT infrastructure, and cybersecurity), illustrated by case studies from Atrium, Cleveland Clinic, Corewell, HCA Healthcare, Mayo Clinic, Providence, and Sutter Health, showing that successful AI adoption always begins with these three pillars [58].

5. **Singapore Artificial Intelligence in Healthcare Guidelines (AIHGle 2.0).** Developed by Singapore's Ministry of Health and Health Sciences Authority, AIHGle provides a healthcare-specific governance framework for the responsible development, deployment and use of AI. It differentiates responsibilities among developers, healthcare organizations as deployers, and healthcare professionals as users, emphasizing patient-centred implementation, transparency, risk assessment, human accountability and continuous monitoring. As a living framework, AIHGle provides a useful international example of translating high-level Trustworthy AI principles into operational healthcare governance [135].

6. **IEEE Standards for Trustworthy and Medical AI.** The IEEE Standards Association provides a complementary family of technical and ethical standards relevant to healthcare AI. IEEE 7000 [136] establishes processes for integrating ethical considerations throughout system design; IEEE 7001 [137] addresses measurable transparency; IEEE 7002 [138] establishes privacy-engineering processes; and IEEE 7003 [139] provides methods for identifying and mitigating algorithmic bias and defining appropriate application boundaries. For medical AI specifically, IEEE 2801 [140] addresses lifecycle quality management of medical-AI datasets, while IEEE 2802 [141] establishes terminology for performance and safety evaluation. IEEE 2894 [142] further provides an architectural framework for explainable AI. Together, these standards operationalize principles such as transparency, privacy, fairness, traceability, data quality and human-centred governance that are directly relevant to safe LLM implementation in cancer literacy and clinical contexts. IEEE CertifAIEd™ [70] complements these standards through conformity-assessment and certification activities.

7. **International consensus frameworks for trustworthy medical AI:**
   - **The FUTURE-AI framework** offers 30 consensus recommendations across six core principles: explainability (clinically meaningful, end-user-validated justifications), fairness (identifying and mitigating demographic and clinical biases such as sex, age, ethnicity, or skin tone), robustness and universality (consistent performance across data sources and clinical environments via representative

datasets and domain adaptation), usability (intuitive human-AI collaboration and end-user training), traceability (ongoing auditing, model cards, and AI passports), and regulatory compliance (alignment with the AI Act, FDA SaMD guidance, and GDPR). It promotes a full lifecycle approach: design, development, evaluation, and deployment, with continuous updates, real-time monitoring, and periodic auditing as clinical practice evolves [59].

- **CHEERS-AI** complements this with a 38-item checklist (28 items from the established CHEERS 2022 standards, 8 AI-specific elaborations, and 10 new extensions) for economic evaluation of AI in medicine, addressing a critical problem: many current cost-effectiveness studies on AI are poorly reported, making them hard for decision-makers to trust and compare. New requirements include disclosing an AI system's degree of autonomy, whether it can learn over time, and how it was developed and validated. Together with FUTURE-AI's ethical blueprint, CHEERS-AI gives health technology assessment bodies and hospital administrators the standardized, high-quality evidence needed for confident funding and implementation decisions [60-62].

5.3. Sector-specific and professional guidance

1. **ESMO.** ESMO's guidance envisions LLMs as strictly assistive, human-supervised instruments that enhance, but never replace, clinical reasoning, communication, and decision-making in cancer care. Its ELCAP framework defines three contexts: patient-facing chatbots (supervised, with clear escalation pathways and strict privacy controls), clinician-facing decision-support tools (formally validated, explainable, with explicit human oversight of all outputs), and institutional background systems (continuously governed, bias-monitored, periodically re-validated), stressing that the clinical value of LLMs depends critically on who uses them and how they are integrated into workflows, and that generative applications should be treated as regulated medical devices requiring clinical testing and outcome validation before deployment. Education and digital literacy are identified as central components, requiring oncologists to understand AI principles, recognize system limitations, and appreciate the risks of overreliance [63, 64].

2. **NHS England.** The NHS AI/ML framework emphasizes clinician–engineer–data-scientist collaboration, robust information governance, and rigorous validation of proof-of-concept models before implementation. It stresses that healthcare professionals need confidence in AI technologies, addressed through transparent communication, education, and involvement in development, and that clinicians should actively shape technology use rather than passively accept implemented systems. Practical tools include model cards informing users of algorithm capabilities and limitations, proactive equality assessments to prevent discrimination, and post-market incident reporting to

manufacturers and the MHRA; the framework also anticipates system-wide impacts on clinical workloads and payment structures. The UK's Digital Markets, Competition, and Consumers Act 2024 (DMCCA) further empowers the Competition and Markets Authority to curb anti-competitive practices, aiming to keep the AI healthcare market competitive, innovative, and equitable for all populations [65, 66].

3. **European Cancer Imaging Initiative (ECII).** Part of Europe's Beating Cancer Plan, ECII connects fragmented cancer imaging data repositories across member states through the federated EUCAIM (European Federation for Cancer Images) infrastructure, funded under the Digital Europe Programme. Launching with 21 clinical sites across 12 countries and expanding toward 30+ sites and 100,000 anonymized cases, it will host AI tools and clinical prediction models under strict ethical and data-protection standards, aiming for a fully operational European imaging repository by 2026 [67].

4. **European Lung Foundation (ELF)**, with 66 partner organizations, has issued a joint statement calling on the European Commission, Parliament, and national governments for clear AI regulation, reliability testing, patient safety safeguards, increased funding for collaborative research, improved data sharing through the EHDS, and seamless integration of AI into hospital workflows [68].

5. **Melanoma Patient Network Europe (MPNE)**, via its 2024 Berlin Consensus developed through extensive stakeholder engagement, promotes risk-proportionate data protection avoiding both under- and over-regulation, treating health data as a common good rather than a resource for exclusive commercial profit, zero-trust security distributing control across multiple stakeholders to minimize insider risk, and robust, legally enforceable, future-proof governance with meaningful patient participation [69].

6. **IEEE CertifAIEd™** offers certification against transparency, accountability, algorithmic bias prevention, and privacy protection, a verifiable, recognized quality mark that helps healthcare organizations build patient trust beyond baseline legal compliance [70].

7. **Clinical AI benchmarking.**
   - **MedAgentBench**, a FHIR-based virtual EHR testbed, tests AI agents against hundreds of physician-curated tasks from simple data retrieval to complex, multi-step clinical workflows; its finding that top models achieve only a 69.67% success rate provides objective evidence that current systems are not ready for autonomous clinical use, validating the cautious, human-supervised approach advocated by ESMO's ELCAP [71, 72].

- **OpenAI's HealthBench,** built with input from 260+ doctors, uses 5,000 realistic multi-turn conversations in multiple languages and 48,000 physician-written criteria across five domains (Accuracy, Completeness, Context Awareness, Communication Quality, Instruction Following) to test whether AI knows when to escalate to emergency care, respond cautiously when unsure, or ask for more information. Even advanced models struggle on the tougher "HealthBench Hard" subset; a 34-item "HealthBench Consensus" list, where doctors nearly universally agree on the ideal response, serves as a gold-standard benchmark, graded at scale via an automated GPT-4.1 grader that closely matches expert human judgment [73].The 2026 launch of ChatGPT Health operationalizes these safety principles in a secured, isolated environment integrating EHR and wearable data (from sources like Apple Health) that is never used to train underlying models, translating HealthBench's guardrails into practice, knowing when to escalate to human care, while explicitly avoiding diagnosis and improving health literacy and visit preparation instead [74, 75].

5.4. Common principles and implementation framework

Table 2 consolidates ethical imperatives drawn from the AI Act, WHO, OECD, and ESMO frameworks.

Table 2. Trustworthiness principles for cancer-focused LLMs

Principle	Description	Implications for cancer literacy LLMs
Human Oversight	Critical outputs must be reviewable and overridable by humans	Oncologists/trained staff supervise educational outputs and verify accuracy
Accountability	Developers and institutions own performance and error management	Clear roles for validation, updates, and documented interventions
Privacy & Data Protection	GDPR compliance; anonymized, secure processing	De-identified examples; EHDS/GDPR-aligned tools

Equity & Inclusivity	Prevent bias; ensure accessibility	Consider language, literacy level, and cultural context
Continuous Validation & Safety	Ongoing performance and risk monitoring	Periodic audits and patient/clinician feedback loops
Ethical, Human-Centered Use	Respect autonomy; prevent harm	Support, not replace, clinicians and informed decision-making
Stakeholder Engagement	Involve patients, clinicians, and regulators in governance	Patients and advocacy groups shape relevance and trust

Taken together, these principles show that no single regulation fully addresses patient-facing generative AI in oncology alone: the AI Act supplies the risk-based governance backbone, MDR/IVDR determine when a tool counts as a medical device, GDPR and the EHDS govern the underlying data, and WHO/OECD/ESMO guidance fills in ethical and clinical nuance. For cancer-literacy LLMs specifically, this means transparency, human oversight, and accountability must be designed in from the start rather than added retroactively.

6. Current clinical guidelines and evidence use of AI in oncology

6.1. Existing guidelines and professional perspectives

ESMO Guidance on the Use of Large Language Models in Clinical Practice (ELCAP). In 2025, the European Society for Medical Oncology published the first oncology-specific framework for the structured use of LLMs in clinical practice. ELCAP distinguishes three contexts of use: patient-facing systems, systems used directly by healthcare professionals, and institutional or background systems. The framework emphasizes that risks and safeguards differ according to the intended user and purpose, requiring appropriate human oversight, protection of patient data, systematic validation, transparency and continuous monitoring. Patient-facing systems should complement

rather than replace professional advice, while clinically consequential applications require stronger validation and professional oversight [63].

ESMO Basic Requirements for AI-based Biomarkers in Oncology (EBAI). Complementing ELCAP, the EBAI framework establishes minimum requirements for the validation and clinical adoption of AI-based biomarkers in oncology. It emphasizes appropriate ground truth, performance evaluation, calibration, generalisability across intended populations and settings, fairness, and progressively stronger clinical evidence according to the impact of the AI output on prognosis or treatment selection. EBAI reinforces an important principle for clinical AI: performance demonstrated for one population, cancer type, modality or intended purpose should not be generalized beyond the evidence supporting that specific use [143].

The 2024 European melanoma guideline [76] confirms that in controlled experimental settings, AI algorithms trained on vast image libraries can match or exceed dermatologists' diagnostic accuracy for melanoma, but this performance does not reliably translate to real-world clinical settings. A particular concern involves direct-to-consumer AI skin cancer apps such as **SkinVision** [77] and **DermaSensor** [78], where no proven safety profile exists, making their use potentially dangerous for patients.

The European Academy of Dermatology and Venereology (EADV) AI Task Force [79] identifies eight principal concerns with AI-powered dermatology apps: inaccurate results paired with poor user education, a potential decline in clinicians' diagnostic skills, significant data security risks, commercial influence, problematic cost structures, and inconsistent regulatory approval. It recommends mandatory developer transparency on accuracy, data use, privacy, and cost; a consistent user experience for all individuals; and stricter, more consistent European regulatory rules guaranteeing app safety and accuracy.

6.2. Examples of approved AI tools for different cancer types

A wide range of clinically validated tools now span imaging, pathology, screening, and treatment workflows across major cancer types, illustrating the breadth of AI applications already integrated into oncology practice and the rapid expansion of certified, real-world deployments beyond the research stage [80]. Table 3 summarizes representative examples by cancer type.

Table 3. Examples of approved AI tools across cancer type

Cancer type	Tool (Developer)	Function
Breast cancer	Transpara (ScreenPoint Medical) [81]	Mammography reading aid identifying suspicious regions
Breast cancer	MIA (Kheiron Medical) [82]	AI detection support for screening mammograms
Breast cancer	ProFound AI (iCAD Inc.) [83]	2D and digital breast tomosynthesis (DBT) computer-assisted detection
Breast cancer	ER/PR APP (Visiopharm) [84]	Automated pathology scoring for hormone receptor expression
Lung cancer	InferRead CT Lung (Infervision) [85]	AI-supported CT lung cancer screening
Lung cancer	Auto Lung Nodule Detection (Samsung) [86]	AI for pulmonary nodule detection
Lung cancer	Veye Chest (Aidence) [87]	AI triage for lung nodules
Prostate cancer	SmartTarget (SmartTarget Ltd) [88]	MRI-based image-guided interventional planning
Prostate cancer	JPC-01K [89]	Prostate MRI detection assistance
Prostate cancer	Paige Prostate [90]	AI-supported pathology for biopsy evaluation
Colorectal cancer	GI Genius (Medtronic/Cosmo AI) [91]	AI-assisted polyp detection during colonoscopy
Colorectal cancer	ColonFlag (EarlySign) [92]	Risk stratification for colorectal cancer
Brain tumors	MRCAT Brain (Philips) [93]	Automated MRI segmentation for radiotherapy planning
Brain tumors	QyScore (Qynapse) [94]	Quantitative brain lesion volumetrics
Skin cancer	DERM (Skin Analytics) [95]	AI decision-support for skin lesion evaluation
Gynecologic cancers	AVEC (MobileODT) [96]	Cervical cancer screening support
Gynecologic cancers	Cintec Plus Cytology (Ventana) [97]	Dual p16/Ki-67 immunocytochemical assay
Multi-cancer/general	Arterys Oncology DL [98]	Cloud-based CT/MRI lesion tracking
Multi-cancer/general	SubtlePET (Subtle Medical) [99]	Noise reduction for faster PET acquisition

Multi-cancer/general	RayCare/RayStation (RaySearch) [100]	Workflow management and radiotherapy treatment planning

7. A bottom-up framework for developing guidelines on the use of LLMs in cancer literacy

As our primary objective is to develop guidelines for the appropriate use of LLMs to enhance cancer literacy, it is essential to adopt an innovative, bottom-up approach that encourages thinking beyond conventional frameworks. Therefore, we must consider the development of guidelines addressing three distinct categories: (i) patients and citizens, (ii) healthcare professionals, and (iii) clinical institutions. The starting point is the recognition that people are already using LLMs as sources of information; therefore, there is a need for education and guidance in this regard. This structure aligns with emerging oncology-specific governance models, notably the ESMO ELCAP framework.

7.1. Rationale, guiding principles, and governance considerations for this type of framework

LLM accuracy, transparency, and reliability vary widely, risking misinformation and anxiety-inducing content that may influence health-seeking behavior without structured guidance; professionals face similar risks around bias, contextual accuracy, and responsibility, while institutions must ensure system-level oversight, governance, and quality control.

A bottom-up, co-creation approach supports the development of guidance tailored to the different roles, knowledge levels, needs, and responsibilities of patients and citizens, healthcare professionals, and organizations across the cancer care continuum.

A UK NHS case study shows why bottom-up works: the first NLP system built inside the National Health Service succeeded because clinical experts were closely involved from the beginning. As Au Yeung et al. (2024) explain, when experts label and explain clinical text directly at its source, their knowledge becomes embedded in the model, helping it understand local context and medical language; a step-by-step cycle of clinician annotation and correction let the model improve gradually. Their finding, that the best AI projects are those where clinicians clearly identify the problem to be solved, confirms that

a user-focused, bottom-up approach is essential for building cancer literacy LLMs that are both effective and trustworthy [101]. A complementary Trustworthy AI (TAI) design framework formalizes this stakeholder-centric methodology by mapping ethical principles: human agency, transparency, fairness, into concrete, actionable specifications explicitly tied to the different needs of patients, clinicians, and institutions, supplying the architectural blueprint for consistently safe, reliable systems [102].

Within this framework, guidelines are structured according to three stakeholder groups:

- *Patients and citizens*: LLMs should support access to understandable cancer information grounded, where possible, in authoritative, current, and traceable evidence sources. Outputs must be adapted to users' literacy levels and language, clearly disclose their AI nature, comply with GDPR requirements, and consistently direct users back to professional care rather than substituting clinical advice.
- *Healthcare professionals:* LLMs should function as clinical and educational support tools, assisting with evidence synthesis, literature summarization, innovation tracking, and interdisciplinary decision-making. Outputs should provide traceable sources where possible and explicitly identify uncertainty, missing information, and limitations. Numerical confidence estimates should only be presented where they have been appropriately validated and calibrated for the intended use. Continuous AI literacy training is required, and human oversight remains mandatory, with healthcare professionals retaining full responsibility for all clinical decisions.
- *Organizations:* Institutions should establish governance frameworks defining accountability, quality control, and ethical oversight across the AI lifecycle. Governance should define approved purposes and boundaries of use and include pre-deployment validation, continuous performance and safety monitoring, incident reporting, change management, and re-evaluation following material model or system updates. They should ensure secure integration with clinical systems and registries, promote equitable access to AI literacy resources for both patients and staff, and regularly evaluate AI outputs for accuracy, bias, and cultural sensitivity. Organizational responsibilities also include maintaining ethical oversight committees and ensuring that LLMs support, rather than replace, cross-department communication and standardized educational initiatives aligned with national and EU-level cancer prevention strategies.

In addition, emerging evidence highlights systemic risks associated with the increasing use of AI in administrative healthcare functions. A 2026 analysis describes a potential "AI arms race" in decisions such as insurance authorization and treatment approval, where reliance on automated systems risks reducing meaningful human review and increasing opacity in decision-making processes. In oncology, even small delays in approval can

significantly affect patient outcomes, while biased historical data may reinforce inequities unless robust safeguards are implemented. Therefore, AI systems should be designed to support administrative efficiency while ensuring that any decision restricting access to care remains under transparent human responsibility [103]. Furthermore, AI ethics audits currently face significant practical limitations, including fragmented standards, resource constraints, and a tendency to focus narrowly on technical metrics while overlooking broader socio-technical risks, creating a gap between regulatory intent and real-world evaluation [104].

Across all three stakeholder groups, the level of evidence, validation, transparency, and human oversight should be proportionate to the intended purpose, user, data involved, degree of system autonomy, and potential consequences of an incorrect output.

7.2. Alignment with the ESMO ELCAP framework

Developed by a 20-expert ESMO panel (November 2024–February 2025) comprising oncologists, AI specialists, biostatisticians, digital health experts, and patient advocates, and released at the ESMO Congress 2025, the ELCAP framework provides an oncology-specific governance model grounded in clinical accountability, transparency, and data protection. It translates high-level ethical principles into 23 consensus "Do's and Don'ts" across three categories of AI applications, with mandatory human oversight across all use cases. Patients must be informed of AI involvement, and healthcare professionals retain final responsibility for clinical decisions. ELCAP currently focuses on assistive LLMs operating under human supervision, while acknowledging that more autonomous agentic systems will require further regulatory development [105, 106]. This classification, presented in Table 4, provides a practical foundation for implementation, whereby increasing system autonomy or proximity to clinical decision-making requires progressively stricter validation, monitoring, and governance.

Table 4. Classification of ELCAP applications according to user group and functional role

ELCAP Type	Primary User	Main Purpose	Requirements
Type 1 – Patient-Facing	Patients, caregivers	Education and symptom support	Supervised pathways, escalation mechanisms, data protection
Type 2 – Healthcare Professional-Facing	Healthcare professionals	Decision support, documentation, literature synthesis	Formal validation, transparency, human accountability

Type 3 – Background Systems	Institutional systems	Data extraction, trial matching, administrative support	Pre-deployment testing, bias monitoring, governance frameworks

7.3. Community-based validation: The Lerești Workshop (Romania)

To ensure these guidelines are relevant, the framework incorporates findings from the technical workshop conducted on November 20, 2025, in Leresti, Romania (see Annex 1). This case study serves as a "bottom-up" validation, highlighting how real-life input must shape the deployment of LLMs and the usage of digital tools for health.

The evidence from the workshop indicates that while citizens value digital tools for health, they are highly sensitive to "anxiety loops" caused by frequent monitoring. To ensure psychological safety, LLMs must provide supportive and polite dialogue rather than blunt messages that could cause significant shock or distress. Beyond technical accuracy, ethical concerns focus on a fear of "robotization" and a feeling that AI might intrusively "guide" rather than just monitor human behavior. While sharing data is generally accepted if the results are useful, the presence of cameras on AI digital devices is a major constraint.

The success of AI is also tied to the doctor's role as a "trust anchor." Participants feel that AI has limited value at home without a professional to explain the results and support a diagnosis. Guidelines must ensure that technology supports, but never replaces, "foundational medical skills" and manual expertise. There is a clear expectation for clinicians to keep direct interaction and eye contact, ensuring the AI remains a helpful tool.

Trust in these systems depends more on "strong and firm regulation" than on technical complexity. In rural or traditional areas like Lerești, institutions should involve trusted figures like local priests, the mayor, to build confidence in new tools. Finally, institutions should prioritize placing AI devices in clinical or workplace settings where experts can ensure the technology is properly coordinated and the results are robust and reliable.

8. AI-generated deepfakes and trustworthy cancer literacy tools

The digital health information landscape is increasingly exposed to AI-generated deepfakes, which represent a non-theoretical and growing risk in oncology. Such content may deliberately target vulnerable moments in a patient's care journey by promoting false "miracle cures," discrediting effective therapies, or fabricating side effects of standard treatments. These manipulations can undermine trust in clinicians, weaken confidence in evidence-based medicine, and lead to delayed or unsafe treatment decisions. Real-world cases, such as the misuse of a medical professional's likeness to promote unverified products, demonstrate how easily authoritative identities can be exploited for misinformation, a risk that is particularly harmful in oncology contexts where patients are emotionally and medically vulnerable [107-109].

Addressing this risk requires a layered response combining source and content verification, provenance mechanisms, public AI and health literacy, trusted intermediaries, platform-level safeguards, and appropriately governed AI tools capable of directing users toward authoritative evidence. Unlike static health information portals, which require high levels of health literacy and self-interpretation, trustworthy AI systems enable interactive, conversational access to validated medical knowledge. Through "trust-by-design" principles, these systems explicitly communicate limitations, explain reasoning, and operate under human oversight, shifting from passive information delivery to personalized interpretation of medical content (e.g., "what does this mean for me?") [110-111].

A key enabling approach is Retrieval-Augmented Generation (RAG), which can ground LLM outputs in curated, evidence-based medical sources by retrieving relevant information from approved repositories and supplying it as context for generation. This can improve traceability and reduce hallucination risk, particularly when users are provided with access to the underlying sources. However, RAG does not guarantee factual consistency or eliminate hallucinations, retrieval errors, outdated evidence, or inappropriate synthesis. Trustworthy implementation therefore requires source governance, provenance, validation, uncertainty handling, and appropriate human oversight [112–116]. In Table 5, we present the main differences between traditional and AI-assisted health information access for the large public. This comparison illustrates why a trustworthy, RAG-based AI assistant is not simply a faster search engine: by mediating access to the same verified medical library, it lowers the literacy barrier to understanding

complex information while, only if properly designed, avoiding the fabricated content that makes deepfakes and generic chatbots risky in the first place.

Table 5. Benefits and risks of traditional vs. AI-assisted health information access

Feature	Static health portal	AI tool
How users interact	Manual search and self-interpretation	Plain-language questions, AI summarizes and answers
Source of information	Full articles, self-extracted	Verified library, directly extracted, no unverified content
Effort required	High: needs strong health literacy	Low: AI explains in real time
Accuracy & Compliance	Risk of misinterpretation	Avoids hallucination; flags missing data
Emotional Nuance	User interprets tone themselves	Empathetic delivery, though limited vs. human bedside sensitivity

This matters more given the current misinformation landscape. CURTAIN's Deliverable D2.2 (from WP 2) shows cancer misinformation forms an interconnected ecosystem across three domains: curative food supplements (highly amplified, simplified cure narratives), natural remedies framed against medical treatment (the most persistent narrative, casting cancer as a lifestyle imbalance), and distrust in health authorities, pharmaceutical companies, and vaccines (an enabling layer legitimizing alternative explanations), with risk unevenly distributed: Romania shows high convergence of all three domains at scale, especially on TikTok, Bulgaria centers on natural remedies, while Portugal, Moldova, and Montenegro show more limited or selective exposure [117]. TikTok drives rapid amplification even at moderate publication volumes; YouTube and Telegram sustain cross-platform persistence over time; Facebook provides a high-participation structural backbone embedding narratives within broader health discussions. Narratives adapt culturally, pseudo-scientific framing in Southern Europe (Italy, Portugal), tradition and folklore ties in Eastern Europe (Romania, Bulgaria), survival and political-polarization framing in crisis contexts (Ukraine, Moldova), and low content volume does not indicate low risk, since platform amplification can create outsized exposure from few posts, often delivered in neutral, non-alarmist tones that evade easy detection. As AI tools become more widely available, existing misleading narratives risk being reinforced through AI-generated content, including realistic deepfake videos, images, and voices, making trustworthy AI assistants increasingly important for helping patients and families know which information to trust.

Building on this risk landscape, the final chapter presents the co-developed and validated recommendations that translate these findings into concrete guidance for citizens, healthcare professionals, and organizations.

9. Co-creation and validation of the proposed recommendations

The recommendations presented in this deliverable were developed and validated through a multi-stage, stakeholder-centred process combining scientific and regulatory evidence, community-based qualitative input, multidisciplinary expert consultation, and targeted end-user assessment. This approach was selected because the safe use of LLMs in cancer literacy cannot be evaluated solely through technical performance: recommendations must also be understandable, relevant, applicable, ethically appropriate, and sensitive to differences in health literacy, digital literacy, vulnerability, professional responsibility, and organizational context.

A co-creation expert discussion led to a set of 12 initial recommendations, distributed in 4 directions per category (citizens & patients, healthcare professionals, organizations), as they are detailed in Annex 3. These initial recommendations underwent targeted stakeholder validation involving four groups with different information needs and levels of exposure to digital health technologies: older adults, older adults living in rural areas, pharmacists, and university students. The recommendations were translated into Romanian and assessed through a modular questionnaire using both quantitative and qualitative feedback. Participants evaluated key characteristics including accessibility, readability, comprehensibility, relevance, and practical applicability. Community participants assessed the recommendations intended for citizens and patients, while pharmacists and university students were provided with the broader validation package. Feedback was subsequently used to identify unclear wording, potential blind spots, and areas requiring refinement before finalization of the guidelines. Post-validation, the feedback received led to rephrasing several recommendations and establishing 4 additional recommendations: 2 targeting citizens, patients and healthcare professionals, both highlighting data confidentiality and 2 targeting organizations, emphasizing on AI governance and building role-specific AI literacy within employees.

The validation focused not only on whether individual recommendations were understandable, but also on whether the overall framework preserved several essential principles: differentiation according to user group and intended purpose; proportionality between potential risk and required safeguards; verification of clinically relevant AI

outputs against authoritative evidence; protection of personal and health data; transparency regarding AI capabilities and limitations; meaningful human oversight; professional accountability for clinical decisions; organizational governance across the AI lifecycle; and equitable access for people with different levels of cancer, health, and digital literacy.

Attention was given to avoiding two opposite risks: The first is under-protection, in which users may over-trust fluent AI-generated content, rely on fabricated or insufficiently referenced information, disclose sensitive health data, or use general-purpose AI for patient-specific decisions beyond its validated purpose. The second is over-restriction, in which overly broad warnings or prohibitions prevent appropriate use of AI for cancer education, information navigation, evidence synthesis, communication, and other legitimate supportive functions. The final recommendations therefore adopt a purpose- and risk-based approach: the level of evidence, validation, transparency, data governance, and human oversight should increase according to the intended use, user group, data involved, degree of system autonomy, and potential consequences of incorrect output.

The validation also confirmed the need to distinguish **cancer-literacy support from clinical decision support**. Helping a citizen understand medical terminology, prepare questions for a consultation, or navigate authoritative cancer information represents a fundamentally different use case from interpreting patient-specific symptoms, laboratory results, imaging, genomic data, or treatment options. Likewise, healthcare professionals may appropriately use AI for selected information-processing and evidence-support tasks, but general-purpose LLMs should not be treated as validated clinical decision-support systems. Organizations consequently have an important role in defining which systems are approved, for which purposes and users, under which data and security conditions, and with what mechanisms for monitoring, incident management, reassessment, and human oversight.

The term **validation** in this deliverable refers to the scientific, expert, stakeholder, usability, and contextual validation of the proposed recommendations. It should not be interpreted as clinical validation, regulatory conformity assessment, certification, or demonstration of safety and effectiveness of any particular LLM or AI system. Any AI application intended for clinically consequential use requires separate evaluation and, where applicable, validation and conformity assessment according to its intended purpose and the relevant regulatory framework.

The present validation also has limitations. The direct community and end-user assessment was conducted within selected stakeholder groups and cannot establish

universal applicability across all European populations, languages, healthcare systems, cancer types, or vulnerable communities. In addition, LLM capabilities, deployment models, clinical evidence, cybersecurity risks, and European regulatory requirements are evolving rapidly. The recommendations should therefore be considered a living framework rather than a fixed endpoint.

The next stage should move from stakeholder and content validation toward real-world implementation and iterative evaluation. Within CURTAIN, these guidelines will provide the evidence-based and governance backbone for Task 4.4 and the co-creation of a real-life use framework for LLMs in cancer literacy with citizens and patients. Experience from implementation across different linguistic, cultural, and healthcare contexts should be used to further assess usability, accessibility, unintended effects, equity, and the effectiveness of safeguards.

The recommendations should subsequently be reviewed and updated periodically, as well as whenever material changes occur in scientific evidence, LLM capabilities, regulatory requirements, identified safety risks, or patterns of real-world use. This continuous validation cycle is necessary to ensure that the framework remains scientifically credible, practically useful, technologically relevant, and aligned with the central principle of these guidelines: AI should enhance cancer literacy and access to trustworthy knowledge while remaining transparent, evidence-informed, proportionately governed, and under meaningful human oversight.

10. Recommendations for the Appropriate Use of Large Language Models to Enhance Cancer Literacy

To ensure that these guidelines are practical, safe, and aligned with current technological developments, a bottom-up co-creation approach was used. A purely top-down regulatory approach is not enough because it is important to understand how patients, healthcare professionals, and institutions use AI in real-life situations. In this context, an expert focus group within the CURTAIN project was organized to discuss and validate the proposed framework.

In line with the approach promoted by the European Commission, these recommendations were designed as "living guidelines." Because AI technologies are evolving rapidly, fixed rules can quickly become outdated. To keep the guidelines relevant, safe, and useful, the need for continuous evaluation, human oversight, and regular updates arises.

The insights gathered from the expert discussion, together with the validation process, helped shape the actionable recommendations presented below. These guidelines adapt broad European principles into practical measures while also introducing solutions specifically tailored to cancer literacy needs. For complete methodology on the co-creation expert discussion, see Annex 2.

Table 6 below provides an outline of the recommendations and their respective target audience:

Table 6. Final recommendations per target audience

Target audience	Recommendation
Citizens and patients	Look for clear AI warnings
	Learn to recognize warning signs ("red flags")
	Understand the limits of prompting and know what is safe to ask
	Protect personal and health information
	Seek guidance from trusted and trained community resources
Healthcare professionals	Verify everything with real evidence
	Understand the limits of AI in clinical workflows
	Protect patient confidentiality and use appropriately governed AI tools
	Maintain clinical responsibility and final decision-making
	Communicate with empathy and openness
Organizations	Ensure meaningful patient and citizen involvement throughout the AI lifecycle
	Establish a transparent process for evaluating and approving AI tools
	Establish strict data-governance and security boundaries
	Implement ethical workplace features
	Govern AI throughout its lifecycle
	Build role-specific AI literacy among staff

10.1. Recommendations for citizens and patients

These rules are designed to help the general public and cancer patients safely navigate AI tools without being misled, panicked, or urged to follow advice that is detrimental to their health:

1. Look for clear AI warnings.
AI tools should clearly disclose when users are interacting with an AI system. For health-related applications, users should also be provided with clear, accessible information about the system's intended purpose, limitations, appropriate uses, and circumstances in which professional medical advice should be sought.

2. Learn to recognize warning signs ("red flags").
Citizens and patients should be able to identify potentially unsafe or misleading AI content. Warning signs include:

- The AI does not encourage consultation with a qualified doctor (e.g. "Just follow these steps to manage your symptoms" omitting any recommendation to speak with the oncologist or your general practitioner);
- It promotes supplements or paid services (e.g. "Have you tried this immunity supplement?");
- It uses emotional, alarming, or urgent language (such as "Act now before it is too late");
- It promises "miracle cures" (e.g. "Follow this diet and the cancer won't spread in your body");
- It does not provide reliable medical sources (e.g. cannot give the proper link to scientific papers, real oncology societies or World Health Organization, or it may hallucinate scientific references);
- It shows unnatural images or voices (e.g. a social media video shows a medical practitioner presenting themselves as an expert in their field, actively giving health advice, but when searching their medical background online, there is no other information to be found apart from several other similar videos);
- It encourages long conversations mainly for commercial purposes (e.g. the chatbot keeps asking questions to prolong the chat like "Tell me a bit more about your daily habits so I can tailor your profile").

3. Understand the limits of prompting and know what is safe to ask.
General-purpose AI tools should not be assumed to have been clinically validated or equipped with safeguards appropriate for individual medical decision-making. Even well-

formulated questions may lead to misleading answers if important medical details are missing. Safe use does not depend only on how the user formulates or writes a question. For publicly available general-purpose AI, citizens and patients should primarily use these tools to support understanding of general health information(for example, explaining terminology or helping formulate questions for a healthcare professional), not as substitutes for diagnosis, treatment selection, or other patient-specific clinical decisions. They must never replace a medical professional or be used for decision-making, such as diagnosing symptoms, or choosing to stop or change a cancer treatment (e.g. "should I take this medication").

4. Protect personal and health information.
Do not enter identifiable health information, medical records, genomic data, test results, images, or other sensitive personal data into publicly available AI tools unless the system is explicitly authorized and governed for this purpose. Before sharing health information, users should understand how their data will be processed, stored, shared, and protected.

5. Seek guidance from trusted and trained community resources.
Citizens and patients are encouraged to discuss the use of digital health tools with trained community representatives, patient organizations, or trusted local support networks. These resources can help people better understand both the benefits and risks of using AI for health information.

10.2. Recommendations for healthcare professionals

These rules are designed for healthcare professionals to help them use AI efficiently in their daily work without compromising patient care or giving up their professional responsibility.

1. Verify everything with real evidence.
In line with the European Commission guidance, clinicians should carefully check any AI-generated summaries, references, or citations against trusted clinical guidelines and peer-reviewed medical literature. Professionals must remember that AI often invents highly realistic but completely fake medical references and authors. Clinicians should verify AI-generated clinical summaries, evidence claims, references, and citations against authoritative clinical guidelines, validated databases, and peer-reviewed literature before relying on them for patient care. LLMs may generate plausible but fabricated or inaccurate references, misrepresent evidence, or omit clinically important information

2. Understand the limits of AI in clinical workflows.

General-purpose AI tools may support tasks such as information synthesis, literature summarization, drafting, education, communication, and selected administrative activities. They should not be treated as validated clinical decision-support systems or relied upon independently for patient-specific diagnosis or treatment decisions. Clinically consequential AI applications require appropriate validation, governance and, where applicable, compliance with medical-device and AI regulatory requirements.

3. **Protect patient confidentiality and use appropriately governed AI tools.**
Healthcare professionals should not enter identifiable patient information, medical records, genomic data, images, or other sensitive health data into publicly available or non-authorized AI systems. Patient data should only be processed through systems approved for the intended healthcare purpose and operated within appropriate organizational, privacy, security, and data-governance safeguards.

4. **Maintain clinical responsibility and final decision-making.**
Following the EU principle that humans hold ultimate responsibility, any disagreement between AI output and clinical judgment should trigger further review and verification. AI recommendations should not be accepted automatically, but differences should also not be ignored. Clinical decisions must always remain under professional control.

5. **Communicate with empathy and openness.**
Healthcare professionals should be prepared for patients who arrive with AI-generated information or self-diagnoses. Instead of dismissing these inputs, clinicians should use them as an opportunity to build trust (for example, by saying "Let's look at this together"). Patients' efforts should be acknowledged, while clearly and calmly explaining the limits of AI in simple terms and guiding them toward reliable medical sources. When time is limited, patients should be advised to seek a second medical opinion where appropriate.

10.3. Recommendations for organizations

These rules are designed to help institutions build safe, transparent AI systems while protecting patient data and employee well-being.

1. **Ensure meaningful patient and citizen involvement throughout the AI lifecycle.**
Organizations should meaningfully involve patients, citizens, and relevant end users at appropriate stages of AI design, selection, implementation, evaluation, and monitoring. Engagement should be proportionate to the intended use and should inform accessibility, usability, acceptability, equity, and identification of real-world risks rather than being limited to one-time consultation.

2. Establish a transparent process for evaluating and approving AI tools.
Organizations should not endorse publicly available or unvalidated AI tools without appropriate assessment, nor use absolute labels such as "completely safe" or "error-free." Instead, they should establish transparent criteria for evaluating and approving AI systems according to intended purpose, evidence, performance, privacy, security, regulatory status, accessibility, and level of risk. Approved use should remain limited to the purposes and conditions for which the system has been assessed.

3. Establish strict data-governance and security boundaries.
Organizations should clearly define which AI systems are authorized to process which categories of information and for what purposes. Identifiable patient information, health records, genomic data, imaging, or other sensitive health data should not be entered into publicly available or non-authorized AI systems. Where AI systems are authorized to process health data, organizations should apply appropriate data minimization, access controls, security, retention policies, auditability, contractual safeguards, and applicable GDPR, cybersecurity, and medical-device requirements.

4. Implement ethical workplace features.
When using AI-based tools to support staff well-being (for example, reminders for breaks or workload management), organizations must ensure full transparency. These features should be optional and not enabled by default. To avoid any form of pressure or monitoring, institutions must clearly state that there are no consequences for ignoring these prompts, and that human resources or management cannot access any individual-level wellness data.

5. Govern AI throughout its lifecycle.
Organizations should maintain an inventory of AI systems in use and define approved purposes, accountable owners, and boundaries of use. Governance should include risk assessment and validation before deployment, monitoring of performance, safety, bias and incidents during use, procedures for reporting and escalation, and reassessment following significant model, data, software, or intended-use changes.

6. Build role-specific AI literacy among staff.
Organizations should provide AI literacy and competency development appropriate to employees' roles, experience, and the risks of the AI systems they use, including understanding of system limitations, verification, privacy, bias, appropriate escalation, and human oversight.

References

[1] The global, regional, and national burden of cancer, 1990–2023, with forecasts to 2050: a systematic analysis for the Global Burden of Disease Study 2023. The Lancet [Internet]. 2025 Sept 24;. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0140673625016356
[2] Synthesis report [Internet]. Public Health. 2024. Available from: https://health.ec.europa.eu/state-health-eu/synthesis-report_en
[3] The latest EU cancer data: what is new? [Internet]. The Joint Research Centre: EU Science Hub. 2025. Available from: https://joint-research-centre.ec.europa.eu/jrc-news-and-updates/latest-eu-cancer-data-what-new-2025-12-17_en
[4] Mohd Amin MR, Ismail I, Sivakumaran VM. Revolutionizing Education with Artificial Intelligence (AI)? Challenges, and Implications for Open and Distance Learning (ODL). Social Sciences & Humanities Open [Internet]. 2025 Jan 27;11:101308. Available from: https://www.sciencedirect.com/science/article/pii/S259029112500035X
[5] Staff C. What Is Artificial Narrow Intelligence? [Internet]. Coursera. 2025. Available from: https://www.coursera.org/articles/what-is-artificial-narrow-intelligence
[6] "Overview of Guidelines for GPAI Models | EU Artificial Intelligence Act." 2025. Artificialintelligenceact.Eu. 2025. https://artificialintelligenceact.eu/gpai-guidelines-overview/.
[7] Gilbert S, Jakob Nikolas K. Guardrails for the use of generalist AI in cancer care. Nature Reviews Cancer [Internet]. 2024 June [cited 2026 July 5];24(6):357–8. Available from: https://www.nature.com/articles/s41568-024-00685-8
[8] Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. Nature [Internet]. 2023 Apr 12;. Available from: https://www.nature.com/articles/s41586-023-05881-4
[9] Understanding LLMs: A Comprehensive Overview from Training to Inference [Internet]. arxiv.org. Available from: https://arxiv.org/html/2401.02038v2
[10] Jaiman A. Large Language Models [LLMs] [Internet]. Medium. 2024. Available from: https://ashishjaiman.medium.com/large-language-models-llms-260bf4f39007
[11] Gupta S. A Comprehensive Survey of Retrieval-Augmented Generation (RAG): Evolution, Current Landscape and Future Directions [Internet]. Available from: https://arxiv.org/pdf/2410.12837
[12] Meskó B. Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial. Journal of Medical Internet Research [Internet]. 2023 Oct 4;25(1):e50638. Available from: https://www.jmir.org/2023/1/e50638/
[13] Andrew A, Tizzard E. Large language models for improving cancer diagnosis and management in primary health care settings. Journal of Medicine, Surgery, and Public Health. 2024 Dec;4:100157.

[14] Study on the deployment of AI in healthcare - Publications Office of the EUrep [Internet]. Public Health. 2025. Available from: https://health.ec.europa.eu/publications/study-deployment-ai-healthcare-publications-office-eurep_en
[15] Christoph Sadée, Testa S, Barba T, Hartmann K, Schuessler M, Thieme A, et al. Medical digital twins: enabling precision medicine and medical artificial intelligence. The Lancet Digital Health [Internet]. 2025 June 1;100864–4. Available from: https://www.thelancet.com/journals/landig/article/PIIS2589-7500(25)00028-7/fulltext
[16] World Health Organization (2025). Low health literacy is costing health. World Health Organization. https://doi.org/10.2471/B09619. License: CC BY-NC-SA 3.0 IGO
[17] Cancer literacy – Informing patients and implementing shared decision making. Journal of Cancer Policy [Internet]. 2023 Mar [cited 2026 July 5];35:100375. Available from: https://www.sciencedirect.com/science/article/abs/pii/s2213538322000546
[18] Ryman C, Warnicke C, S. Hugosson, A.-B. Zakrisson, Dahlberg K. Health literacy in cancer care – A systematic review. European journal of oncology nursing. 2024 June 1;70:102582–2.
[19] Marron JM, Kyi K, Appelbaum PS, Magnuson A. Medical Decision-Making in Oncology for Patients Lacking Capacity. American Society of Clinical Oncology Educational Book. 2020 May;40(40):e186–96
[20] Agnese Konusevska. European Code of Cancer Practice [Internet]. European Cancer Organisation. 2020. Available from: https://www.europeancancer.org/content/european-code-of-cancer-practice.html
[21] ESMO - Patient Guides [Internet]. Esmo.org. 2026 [cited 2026 July 5]. Available from: https://patient-guides.esmo.org/#/
[22] My Cancer Navigator: information service for people with cancer [Internet]. Anticancer Fund. 2025 [cited 2026 July 5]. Available from: https://www.anticancerfund.org/en/we-help-patients
[23] For trustworthy AI, keep the human in the loop. Nature Medicine [Internet]. 2025 Oct [cited 2026 July 5];31(10):3207–7. Available from: https://www.nature.com/articles/s41591-025-04033-7
[24] GRADE Book [Internet]. Gradepro.org. 2024. Available from: https://book.gradepro.org/
[25] GRADE Working Group. How to Use the GRADE Book. [online]. 2026. [Accessed 26 August 2026]. Available from: https://book.gradepro.org/about
[26] Petrov I, Ivanov H, Kalev DN, Zamboglou C, Pathak A, Konsoulova-Kirova A. 363TiP Pseudo-anonymized non-interventional retrospective clinical validation of AI-ambient patient-clinical intelligence (A.P.C.I.) using the GRADE methodology. ESMO Real World Data and Digital Oncology [Internet]. 2025 Nov [cited 2026 July 5];10:100559. Available from: https://www.esmorwd.org/article/s2949-8201(25)00448-5/fulltext

[27] Erdat EC, Köksoy EB, Utkan G. Artificial Intelligence Chatbot as a Companion for Cancer Patients About Most Common Questions: Analysis of Readability and Quality. Acta Haematologica Oncologica Turcica [Internet]. 2025 Mar 7 [cited 2026 July 5];. Available from: https://actaoncologicaturcica.com/articles/artificial-intelligence-chatbot-as-a-companion-for-cancer-patients-about-most-common-questions-analysis-of-readability-and-quality/ahot.galenos.2024.2024-8-2
[28] Geantă M, Bădescu D, Chirca N, Nechita OC, Radu CG, Rascu S, et al. The Potential Impact of Large Language Models on Doctor–Patient Communication: A Case Study in Prostate Cancer. Healthcare. 2024;12(15):1548. Available from: https://www.mdpi.com/2227-9032/12/15/1548
[29] Geantă M, Bădescu D, Chirca N, Nechita OC, Radu CG, Rascu Ș, et al. The Emerging Role of Large Language Models in Improving Prostate Cancer Literacy. Bioengineering. 2024;11(7):654. Available from: https://www.mdpi.com/2306-5354/11/7/654
[30] Sorin V, Klang E, Miri Sklair-Levy, Cohen I, Zippel DB, Nora Balint Lahat, et al. Large language model (ChatGPT) as a support tool for breast tumor board. 2023 May 30;9(1).
[31] Doctronic. Doctronic, Your Trusted AI Doctor [Internet]. Doctronic. 2026 [cited 2026 July 5]. Available from: https://www.doctronic.ai/telehealth/
[32] Hayat H, Kudrautsau M, Makarov E, Melnichenko V, Tsykunou T, Varaksin P, Pavelle M, Oskowitz AZ. Toward the autonomous AI doctor: quantitative benchmarking of an autonomous agentic AI versus board-certified clinicians in a real world setting. medRxiv [online]. 2025. [Accessed 5 July 2026]. Available from: https://www.medrxiv.org/content/10.1101/2025.07.14.25331406v1
[33] Chow J, Xu L. Chatbot for Healthcare and Oncology Applications Using Artificial Intelligence and Machine Learning (Preprint). JMIR Cancer. 2021 Feb 9;7(4).
[34] Wang A, Qian Z, Briggs L, Cole AP, Reis LO, Trinh QD. The Use of Chatbots in Oncological Care: A Narrative Review. International Journal of General Medicine [Internet]. 2023 May [cited 2026 July 5];16:1591. Available from: https://pmc.ncbi.nlm.nih.gov/articles/pmc10162388/
[35] Huo B, Boyle A, Marfo N, Tangamornsuksan W, Steen JP, McKechnie T, et al. Large Language Models for Chatbot Health Advice Studies. JAMA Network Open [Internet]. 2025 Feb 4;8(2):e2457879. Available from: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2829839
[36] Lawson McLean A, Vagelis Hristidis. Evidence-Based Analysis of AI Chatbots in Oncology Patient Education: Implications for Trust, Perceived Realness, and Misinformation Management. Journal of Cancer Education [Internet]. 2025 Feb 18 [cited 2026 July 5];40(4):482. Available from: https://pmc.ncbi.nlm.nih.gov/articles/pmc12310775/

[37] Görtz M, Kilian Baumgärtner, Schmid T, Muschko M, Woessner P, Gerlach A, et al. An artificial intelligence-based chatbot for prostate cancer education: Design and patient evaluation study. Digital Health [Internet]. 2023 May 2 [cited 2026 July 5];9:20552076231173304. Available from: https://pmc.ncbi.nlm.nih.gov/articles/pmc10159259/
[38] Li Y, Li J, Li M, Yu E, Rhee D, Amith M, et al. VaxBot-HPV: a GPT-based chatbot for answering HPV vaccine-related questions. JAMIA Open [Internet]. 2025 Feb 19;8(1):ooaf005. Available from: https://doi.org/10.1093/jamiaopen/ooaf005
[39] Hou Z, Wu Z, Qu Z, Gong L, Peng H, Jit M, et al. A vaccine chatbot intervention for parents to improve HPV vaccination uptake among middle school girls: a cluster randomized trial. Nature Medicine [Internet]. 2025 Apr 7;. Available from: https://www.nature.com/articles/s41591-025-03618-6
[40] Shaban M, Osman YM, Mohamed NA, Shaban MM. Empowering breast cancer clients through AI chatbots: transforming knowledge and attitudes for enhanced nursing care. BMC Nursing. 2025 July 29;24(1).
[41] Howcroft A, Bennett-Weston A, Khan A, Griffiths J, Gay S, Howick J. AI chatbots versus human healthcare professionals: a systematic review and meta-analysis of empathy in patient care. British Medical Bulletin. 2025 Sept 22;156(1).
[42] Mennella C, Maniscalco U, Pietro G De, Esposito M. Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. Heliyon [Internet]. 2024 Feb 29;10(4):e26297. Available from: https://www.sciencedirect.com/science/article/pii/S2405844024023284
[43] European Union. EUR-Lex - 32017R0745 - EN - EUR-Lex [Internet]. eur-lex.europa.eu. 2017. Available from: https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng
[44] View of Chatbots in medicine: certification process and applied use case [Internet]. Www.smw.ch. 2026 [cited 2026 July 5]. Available from: https://www.smw.ch/index.php/smw/article/view/3954/6066
[45] Navarro A. EU MDR and IVDR: Classifying Medical Device Software (MDSW) [Internet]. NAMSA. 2024. Available from: https://namsa.com/resources/blog/eu-mdr-and-ivdr-classifying-medical-device-software-mdsw/
[46] Regulation - 2017/746 - EN - Medical Device Regulation - EUR-Lex [Internet]. eur-lex.europa.eu. Available from: https://eur-lex.europa.eu/eli/reg/2017/746/oj/eng
[47] Medical Devices [Internet]. [cited 2026 July 5]. Available from: https://health.ec.europa.eu/document/download/b78a17d7-e3cd-4943-851d-e02a2f22bbb4_en?filename=mdcg_2025-6_en.pdf
[48] EU AI Act. High-level Summary of the AI Act [Internet]. EU Artificial Intelligence Act. Future of Life Institute; 2024. Available from: https://artificialintelligenceact.eu/high-level-summary/
[49] European Commission. HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE SET UP BY THE EUROPEAN COMMISSION ETHICS GUIDELINES FOR

TRUSTWORTHY AI [Internet]. 2019 Apr. Available from: https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf
[50] European Commission. European Health Data Space Regulation (EHDS) [Internet]. Public Health. 2025. Available from: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_en
[51] Lalova-Spinks T, Robbe Saesen, Silva M, Geissler J, Iryna Shakhnenko, Jennifer Catherine Camaradou, et al. Patients' knowledge, preferences, and perspectives about data protection and data control: an exploratory survey. Frontiers in Pharmacology [Internet]. 2024 Feb 20;14. Available from: https://doi.org/10.3389%2Ffphar.2023.1280173
[52] ERA Forum Stakeholders" document RESPONSIBLE USE OF GENERATIVE AI IN RESEARCH Living guidelines on the [Internet]. 2024. Available from: https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf
[53] Ethics and governance of artificial intelligence for health Guidance on large multi-modal models [Internet]. Available from: https://iris.who.int/server/api/core/bitstreams/e9e62c65-6045-481e-bd04-20e206bc5039/content
[54] Data. Artificial intelligence is reshaping health systems: state of readiness across the WHO European Region [Internet]. Who.int. World Health Organization; 2025. Available from: https://www.who.int/europe/publications/i/item/WHO-EURO-2025-12707-52481-81028
[55] Data, Artificial Intelligence and Digital Health (DAD. Artificial intelligence is reshaping health systems: country profiles [Internet]. Who.int. World Health Organization; 2025 [cited 2026 July 5]. Available from: https://www.who.int/europe/publications/i/item/who-euro-2025-12707-52481-81470
[56] OECD. The OECD Artificial Intelligence (AI) Principles [Internet]. oecd.ai. OECD; 2019. Available from: https://oecd.ai/en/ai-principles
[57] Health CDR. Good Machine Learning Practice for Medical Device Development: Guiding Principles. FDA [Internet]. 2021 Oct 27;. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles
[58] How to Build and Implement Your AI Health Care Action Plan | AHA [Internet]. American Hospital Association. 2025. Available from: https://www.aha.org/aha-center-health-innovation-market-scan/2025-01-14-how-build-and-implement-your-ai-health-care-action-plan
[59] Lekadir K, Feragen A, Fofanah AJ, Frangi AF, Buyx A, Emelie A, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence

in healthcare. arXiv [online]. 2023. [Accessed 5 July 2026]. Available from: https://arxiv.org/pdf/2309.12325
[60] Introducing CHEERS-AI: improving health economic evaluation reporting for AI technologies [Internet]. NICE website: The National Institute for Health and Care Excellence. NICE; 2024. Available from: https://www.nice.org.uk/news/blogs/introducing-cheers-ai:-improving-health-economic-evaluation-reporting-for-ai-technologies
[61] Elvidge J, Hawksworth C, Avşar TS, Zemplenyi A, Chalkidou A, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards for Interventions That Use Artificial Intelligence (CHEERS-AI). Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research [Internet]. 2024 May 23;S1098–3015(24)02366-0. Available from: https://pubmed.ncbi.nlm.nih.gov/3879595
[62] Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) [Internet]. ISPOR | International Society For Pharmacoeconomics and Outcomes Research. Available from: https://www.ispor.org/heor-resources/good-practices/article/consolidated-health-economic-evaluation-reporting-standards-2022-cheers-2022-statement-updated-reporting-guidance-for-health-economic-evaluations
[63] E.Y.T. Wong, Verlingue L, Aldea M, M.A. Franzoi, Umeton R, Halabi S, et al. ESMO guidance on the use of Large Language Models in Clinical Practice (ELCAP). Annals of Oncology [Internet]. 2026 [cited 2026 July 5];36(12):1447–57. Available from: https://www.annalsofoncology.org/article/%20s0923-7534(25)04698-8/fulltext
[64] European Society for Medical Oncology (ESMO) [Internet]. Esmo.org. 2026 [cited 2026 July 5]. Available from: https://www.esmo.org/society-updates/esmo-publishes-first-guidance-on-the-safe-use-of-large-language-models-in-oncology-practice
[65] NHS England. NHS England » Artificial intelligence (AI) and machine learning [Internet]. www.england.nhs.uk. 2025. Available from: https://www.england.nhs.uk/long-read/artificial-intelligence-ai-and-machine-learning/
[66] Narayan S. Public health at the frontier of AI: aligning UK competition policy with national health and innovation goals. International Journal of Science and Research [online]. 2025;14(7):7–11. [Accessed 5 July 2026]. Available from: https://www.ijsr.net/archive/v14i7/SR25627220306.pdf
[67] European Cancer Imaging Initiative | Shaping Europe's digital future [Internet]. digital-strategy.ec.europa.eu. 2023. Available from: https://digital-strategy.ec.europa.eu/en/policies/cancer-imaging
[68] lucyrobinson. EU's vision for artificial intelligence to help transform cancer treatment across Europe - European Lung Foundation [Internet]. European Lung Foundation. 2025. Available from: https://europeanlung.org/en/news-and-blog/eu-vision-for-artificial-intelligence-to-help-transform-cancer-treatment-across-europe/

[69] Patient Consensus on Data and AI | MPNE [Internet]. MPNE. 2020 [cited 2026 July 5]. Available from: https://www.mpneurope.org/patient-consensensus-data-and-ai
[70] Hayes M. Why the Consideration of Ethical AI Matters: The Role of IEEE CertifAIEdTM in Building Trustworthy Technology [Internet]. IEEE Standards Association. 2025 [cited 2026 July 5]. Available from: https://standards.ieee.org/beyond-standards/why-the-consideration-of-ethical-ai-matters-the-role-of-ieee-certifaied-in-building-trustworthy-technology/
[71] Jiang Y, Black KC, Geng G, Park D, Zou J, Ng AY, et al. MedAgentBench: a virtual EHR environment to benchmark medical LLM agents. NEJM AI [online]. 2025;2(9):1-9. doi:10.1056/AIdbp2500144. Available from: https://ai.nejm.org/doi/10.1056/AIdbp2500144
[72] Nikitin M. Unlocking EHR Integration with SMART on FHIR: What Healthcare Leaders Should Know [Internet]. Itirra. 2025 [cited 2026 July 5]. Available from: https://itirra.com/blog/unlocking-ehr-integration-with-smart-on-fhir-what-healthcare-leaders-should-know/
[73] Introducing HealthBench [Internet]. Openai.com. 2025. Available from: https://openai.com/index/healthbench/
[74] OpenAI. Introducing ChatGPT Health. OpenAI [online]. 2026. [Accessed 5 July 2026]. Available from: https://openai.com/ro-RO/index/introducing-chatgpt-health
[75] Barnhart AJ, Comerci G, Prainsack B, Braun M. Solidarity or segregation? ChatGPT Health and US health care disparities. Journal of Medical Internet Research [online]. 2026;28:e94972. doi:10.2196/94972. [Accessed 5 July 2026]. Available from: https://www.jmir.org/2026/1/e94972
[76] Garbe C, Amaral T, Peris K, Hauschild A, Arenberger P, Basset-Seguin N, et al. European consensus-based interdisciplinary guideline for melanoma. Part 1: Diagnostics - Update 2024. European Journal of Cancer [Internet]. 2024 Nov 28;215:115152. Available from: https://www.sciencedirect.com/science/article/pii/S0959804924017593
[77] SkinVision. SkinVision | Skin Cancer Melanoma Detection App | Check Your Skin [Internet]. SkinVision. 2018. Available from: https://www.skinvision.com/
[78] AI Skin Cancer Detection & Melanoma Detection | DermaSensor [Internet]. DermaSensor. 2026 [cited 2026 July 5]. Available from: https://www.dermasensor.com/#ds_device
[79] Sangers TE, Kittler H, Blum A, Braun RP, Barata C, Cartocci A, et al. Position statement of the EADV Artificial Intelligence (AI) Task Force on AI-assisted smartphone apps and web-based services for skin disease. Journal of the European Academy of Dermatology and Venereology : JEADV [Internet]. 2024 Jan;38(1):22–30. Available from: https://pubmed.ncbi.nlm.nih.gov/37766502/

[80] Luchini C, Pea A, Scarpa A. Artificial intelligence in oncology: current applications and future perspectives. British Journal of Cancer [Internet]. 2021 Nov 26;126:1–6. Available from: https://www.nature.com/articles/s41416-021-01633-1
[81] Welcome to ScreenPoint Medical - Home [Internet]. Screen Point. Available from: https://screenpoint-medical.com/
[82] Mia (Acquired by DeepHealth) - Kheiron Medical Technologies [Internet]. Health AI Register. 2026 [cited 2026 July 5]. Available from: https://healthairegister.com/radiology/products/kheiron-mia
[83] AI Breast Cancer Detection | iCAD [Internet]. iCAD is now part of DeepHealth. 2023 [cited 2026 July 5]. Available from: https://www.icadmed.com/breast-health/ai-breast-cancer-detection/
[84] ER APP, Breast Cancer - Visiopharm · Pathology image analysis software [Internet]. Visiopharm · Pathology image analysis software. 2020 [cited 2026 July 5]. Available from: https://visiopharm.com/app-center/app/er-app-breast-cancer/
[85] InferRead CT Lung - AI for Lung CT Scans [Internet]. Infervision.com. 2025 [cited 2026 July 5]. Available from: https://global.infervision.com/products/inferread-ct-lung
[86] Boston Imaging. Samsung Receives FDA Clearance for AI Algorithms [Internet]. Samsunghealthcare.com. Samsung HME America; 2021 [cited 2026 July 5]. Available from: https://usa.samsunghealthcare.com/news-center/digital-radiography/fda-clearance-for-ai-algorithms
[87] Aidence - DeepHealth [Internet]. DeepHealth. 2025 [cited 2026 July 5]. Available from: https://www.aidence.com/
[88] SmartTarget Ltd. SmartTarget Ltd launches a device for the detection of prostate cancer. SmartTarget [online]. 2016. [Accessed 5 July 2026]. Available from: https://www.smarttarget.co.uk/2016/09/30/smarttarget-announces-product-launch/
[89] Super User. Prostate: JPC-01K [Internet]. Ismi.ie. 2019 [cited 2026 July 5]. Available from: https://www.ismi.ie/medical-solutions/it-and-ai-solutions/prostate-jpc-01k.html
[90] Introducing FDA-Approved Paige Prostate [Internet]. info.paige.ai. Available from: https://info.paige.ai/prostate
[91] GI GeniusTM Intelligent Endoscopy Module [Internet]. Medtronic.com. 2014. Available from: https://www.medtronic.com/en-us/healthcare-professionals/products/digestive-gastrointestinal/gastrointestinal-artificial-intelligence/gi-genius-intelligent-endoscopy-module.html
[92] Medial EarlySign. EarlySign. Medial EarlySign [online]. n.d. [Accessed 5 July 2026]. Available from: http://earlysign.com/
[93] Philips. MRCAT Brain MR-RT clinical application. Philips Healthcare [online]. n.d. [Accessed 5 July 2026]. Available from: https://www.usa.philips.com/healthcare/product/HCNMRF320/mrcat-brain-mr-rt-clinical-application

[94] QyScore® - Qynapse - The future of neuroimaging is here [Internet]. Qynapse - The future of neuroimaging is here. 2021 [cited 2026 July 5]. Available from: https://qynapse.com/qyscore/
[95] Diginomic-Riaz. DERM performance - Skin Analytics [Internet]. Skin Analytics. 2023 [cited 2026 July 5]. Available from: https://skin-analytics.com/ai-pathways/derm-performance/
[96] Liger Medical. MobileODT acquisition announcement. Liger Medical [online]. n.d. [Accessed 5 July 2026]. Available from: https://ligermedical.com/mobileodt-acquisition-announcement/
[97] CINtec® PLUS Cytology [Internet]. Diagnostics. 2024 [cited 2026 July 5]. Available from: https://diagnostics.roche.com/us/en/products/lab/cintec-plus-cytology-rtd001259.html
[98] Medical Imaging Cloud AI for Radiology | Arterys [Internet]. Arterys.com. 2026 [cited 2026 July 5]. Available from: https://www.arterys.com/
[99] Subtle Medical. SubtlePET. Subtle Medical [online]. n.d. [Accessed 5 July 2026]. Available from: https://subtlemedical.com/subtlepet/
[100] The next-generation oncology information system - RayCare | RaySearch Laboratories [Internet]. RaySearch Laboratories. 2026. Available from: https://www.raysearchlabs.com/raycare/
[101] Yeung JA, Shek A, Searle T, Zeljko Kraljevic, Dinu V, Mart Ratas, et al. Natural language processing data services for healthcare providers. BMC Medical Informatics and Decision Making. 2024 Nov 26;24(1).
[102] A design framework for operationalizing trustworthy artificial intelligence in healthcare: Requirements, tradeoffs and challenges for its clinical adoption. Information Fusion [Internet]. 2026 Mar [cited 2026 July 5];127:103812. Available from: https://www.sciencedirect.com/science/article/pii/s1566253525008747#
[103] Mello MM, Trotsyuk AA, Jalil A, Char D. The AI Arms Race In Health Insurance Utilization Review: Promises Of Efficiency And Risks Of Supercharged Flaws. Health Affairs. 2026 Jan 1;45(1):6–13.
[104] Schiff DS, Kelley S, Camacho Ibáñez J. The emergence of artificial intelligence ethics auditing. Big Data & Society. 2024 Dec;11(4
[105] Wong EYT, L. Verlingue, Aldea M, Franzoi MA, R. Umeton, Halabi S, et al. ESMO guidance on the use of Large Language Models in Clinical Practice (ELCAP). Annals of Oncology. 2025 Oct 18;36(12):1447–57.
[106] European Society for Medical Oncology (ESMO). ESMO publishes first guidance on the safe use of large language models in oncology practice. ESMO [online]. 2025. [Accessed 5 July 2026]. Available from: https://www.esmo.org/society-updates/esmo-publishes-first-guidance-on-the-safe-use-of-large-language-models-in-oncology-practice

[107] AI Impersonation in Medicine: Risks for Physicians and Patients | Docwire News [Internet]. Docwire News. 2025 [cited 2026 July 5]. Available from: https://www.docwirenews.com/post/ai-impersonation-in-medicine-risks-for-physicians-and-patients
[108] Study finds most cancer patients exposed to misinformation; UF researchers pilot “information prescription” - UF Health [Internet]. Ufhealth.org. 2025. Available from: https://ufhealth.org/news/2025/study-finds-most-cancer-patients-exposed-to-misinformation-uf-researchers-pilot-information-prescription
[109] Campbell D. AI deepfakes of real doctors spreading health misinformation on social media [Internet]. the Guardian. The Guardian; 2025. Available from: https://www.theguardian.com/society/2025/dec/05/ai-deepfakes-of-real-doctors-spreading-health-misinformation-on-social-media
[110] Vani MS, Sudhakar RV, A. Mahendar, Sukanya Ledalla, Marepalli Radha, M. Sunitha. Personalized health monitoring using explainable AI: bridging trust in predictive healthcare. Scientific Reports. 2025 Aug 29;15(1).
[111] Ahadian P, Xu W, Liu D, Guan Q. Ethics of trustworthy AI in healthcare: Challenges, principles, and practical pathways. Neurocomputing [Internet]. 2025 Oct 28;131942. Available from: https://www.sciencedirect.com/science/article/pii/S0925231225026141
[112] Lars M, Meuth SG, Pawlitzki M. Evaluating base and retrieval augmented LLMs with document or online support for evidence based neurology. npj Digital Medicine [Internet]. 2025 Mar 4 [cited 2026 July 5];8(1):137. Available from: https://www.nature.com/articles/s41746-025-01536-y
[113] Kelly A, Eoin Noctor, Ryan L, van. The effectiveness of a custom AI chatbot for T2DM health literacy: An evaluation study (Preprint). Journal of Medical Internet Research [Internet]. 2024 Dec 16;. Available from: https://www.jmir.org/2025/1/e70131
[114] Lars M, Meuth SG, Pawlitzki M. Evaluating base and retrieval augmented LLMs with document or online support for evidence based neurology. npj Digital Medicine [Internet]. 2025 Mar 4 [cited 2026 July 5];8(1):137. Available from: https://www.nature.com/articles/s41746-025-01536-y
[115] Neha F, Bhati D, Shukla DK. Retrieval-Augmented Generation (RAG) in Healthcare: A Comprehensive Review. AI. 2025 Sept 11;6(9):226.
[116] Wong L, Ali A, Xiong R, Shen SZ, Kim Y, Agrawal M. Retrieval-augmented systems can be dangerous medical communicators. arXiv [online]. 2025. [Accessed 5 July 2026]. Available from: https://arxiv.org/pdf/2502.14898
[117] CURTAIN Project. (2025). Vladu, L., Pîrvan, M., Enache, M., Petre, G., Nastasiu, C., & Paicu, S. D2.2 Regional and platform-specific patterns of cancer-related disinformation. Deliverable D2.2. https://curtainproject.eu/public-deliverables

[118] GeeksforGeeks. What is Artificial Intelligence(AI)? [Internet]. GeeksforGeeks. 2017. Available from: https://www.geeksforgeeks.org/artificial-intelligence/what-is-artificial-intelligence-ai/
[119] Prompt engineering | OpenAI API [Internet]. Openai.com. 2025. Available from: https://developers.openai.com/api/docs/guides/prompt-engineering
[120] Coursera Staff. What Is Artificial Narrow Intelligence? [Internet]. Coursera. 2025 [cited 2026 July 6]. Available from: https://www.coursera.org/articles/what-is-artificial-narrow-intelligence
[121] Petrenko A. Generative Artificial Intelligence (GAI) for Research and Creative Activities. Studies in Systems, Decision and Control. 2025 Nov 12;251–65.
[122] GeeksforGeeks. Architecture and Working of Transformers in Deep Learning [Internet]. GeeksforGeeks. 2024. Available from: https://www.geeksforgeeks.org/deep-learning/architecture-and-working-of-transformers-in-deep-learning/
[123] GeeksforGeeks. What is Pre Training and its Objective [Internet]. GeeksforGeeks. 2025 [cited 2026 July 6]. Available from: https://www.geeksforgeeks.org/artificial-intelligence/what-is-pre-training-and-its-objective/
[124] Bergmann D. Fine Tuning [Internet]. Ibm.com. 2024. Available from: https://www.ibm.com/think/topics/fine-tuning
[125] Thakker Y. Zero-Shot vs Few-Shot vs Chain-of-Thought Prompting: Complete Guide 2026 [Internet]. Explainx.ai. explainx.ai; 2026 [cited 2026 July 6]. Available from: https://www.explainx.ai/blog/zero-shot-few-shot-chain-of-thought-prompting-guide-2026
[126] GeeksforGeeks. What is a Large Language Model (LLM) [Internet]. GeeksforGeeks. 2023. Available from: https://www.geeksforgeeks.org/artificial-intelligence/large-language-model-llm/
[127] GeeksforGeeks. What is RetrievalAugmented Generation (RAG) ? [Internet]. GeeksforGeeks. 2024. Available from: https://www.geeksforgeeks.org/nlp/what-is-retrieval-augmented-generation-rag/
[128] Stryker C. Agentic AI [Internet]. Ibm.com. 2025. Available from: https://www.ibm.com/think/topics/agentic-ai
[129] Coursera Staff. What Is a Chatbot? Definition, Types, and Examples [Internet]. Coursera. 2023 [cited 2026 July 6]. Available from: https://www.coursera.org/articles/what-is-a-chatbot
[130] IBM. What Are AI Hallucinations? [Internet]. IBM. 2023. Available from: https://www.ibm.com/think/topics/ai-hallucinations
[131] Katsoulakis E, Wang Q, Wu H, Shahriyari L, Fletcher R, Liu J, et al. Digital twins for health: a scoping review. npj Digital Medicine. 2024 Mar 22;7(1):1–11.

[132] Dhanya Lakshmi, D Jude Hemanth. An Overview of Deepfake Methods in Medical Image Processing for Health Care Applications. Frontiers in artificial intelligence and applications [Internet]. 2024 Feb 7;. Available from: https://ebooks.iospress.nl/doi/10.3233/FAIA231448
[133] Görtz M, Baumgärtner K, Schmid T, Muschko M, Woessner P, Gerlach A, et al. An artificial intelligence-based chatbot for prostate cancer education: design and patient evaluation study. Digital Health [online]. 2023;9:20552076231173304. [Accessed 26 August 2026]. Available from:https://doi.org/10.1177/20552076231173304
[134] Technology.org. EU AI Act: What actually applies on 2 August 2026 [Internet]. Technology.org; 2026. Available from: https://www.technology.org/2026/07/17/eu-ai-act-what-actually-applies-on-2-august-2026/
[135] "MOH MHC Cir 0024_2026 Updated MOH Artificial Intelligence in Healthcare Guidelines (AIHGle 2.0)." n.d. HCSA. https://www.hcsa.gov.sg/licensable-healthcare-services/moh-mhc-cir-0024-2026-updated-moh-artificial-intelligence-in-healthcare-guidelines--aihgle-2-0-/moh-mhc-cir-0024-2026-updated-moh-artificial-intelligence-in-healthcare-guidelines--aihgle-2-0-/.
[136] I "IEEE SA - IEEE 7000-2021." n.d. SA Main Site. Accessed August 26, 2026. https://standards.ieee.org/ieee/7000/6781/.
[137] "IEEE Standards Association." n.d. IEEE Standards Association. Accessed August 26, 2026. https://standards.ieee.org/ieee/7001/6929/.
[138] "IEEE SA - IEEE 7002-2022." n.d. IEEE Standards Association. Accessed August 26, 2026. https://standards.ieee.org/ieee/7002/6898/.
[139] "IEEE SA - P7003." n.d. Standards.Ieee.Org. Accessed August 26, 2026. https://standards.ieee.org/ieee/7003/11357/.
[140] "IEEE SA - IEEE 2801-2022." n.d. IEEE Standards Association. Accessed August 26, 2026. https://standards.ieee.org/ieee/2801/7459/.
[141] "IEEE Standards Association." 2022. IEEE Standards Association. 2022. https://standards.ieee.org/ieee/2802/7460.
[142] "IEEE Standards Association." 2024. IEEE Standards Association. 2024. https://standards.ieee.org/ieee/2894/11296/.
[143] Aldea, M, M Salto-Tellez, A Marra, R Umeton, A Stenzinger, M Koopman, A Prelaj, et al. 2026. "ESMO Basic Requirements for AI-Based Biomarkers in Oncology (EBAI)." Annals of Oncology : Official Journal of the European Society for Medical Oncology 37 (3): 414–30. https://doi.org/10.1016/j.annonc.2025.11.009.

Appendix

Annex 1. Technical Workshop. Trust and usage of digital health solutions (including LLMs) 2025

Digital Health Monitoring: Workshop Findings

A technical workshop was organised in Lerești, Romania, to evaluate the trust and usage of digital tools for health in rural populations. For two hours, 9 participants (4 male, 5 female), with ages between 40 and 76 y.o., personal income ranging between 500 and 1800 euro/ month, 5 have completed higher education and 4 have completed post-secondary education, 7 are married, 3 work for the mayors' office, 1 is a former history teacher and 1 is a former army employee, all residing in the rural area, participated in the event.

Findings:

- **Usage and familiarity with digital tools and LLMs for health-related activities:** Overall, participants were familiar with the existence of such tools, even though they have not used them; they see them as better suited for occasional rather than continuous use, as frequent monitoring may heighten anxiety or stress.
- **Concerns related to using digital tools for monitoring health:** Participants raised several concerns about AI-based tools: some feared that devices using AI might disturb the body's "internal energy"; others felt that increasing reliance on such technologies could contribute to the subtle "robotization" of individuals; and a third emphasized that the real risk lies not in the tools themselves but in insufficient regulation, arguing that strong governance is essential to ensure their safe and beneficial use.
- **Trust in AI-based tools for monitoring health:** Participants expressed mixed views on digital health-monitoring tools. Some showed low trust, noting discrepancies between device readings and medical measurements, while others highlighted that although new technologies often face initial resistance, they can ultimately become valuable and beneficial as society adapts to them.
- **The role of healthcare professionals in placing trust in digital health solutions:** Participants highlighted the essential role of healthcare professionals in a technologically evolving system. One emphasized the need for experts who understand both medicine and digital technologies to use new tools effectively. Another stressed that, despite technological progress, traditional medical skills must be preserved, illustrating that foundational skills remain indispensable when technology fails.

- **Perceived risks in relation to providing personal data to digital tools:** Participants did not perceive sharing personal data with a LLM as a major risk, noting that they already provide similar information to other platforms, and emphasizing that usefulness of results matters more than data concerns.
- **General perceptions in relation to where LLM-enabled physical devices for health monitoring would be acceptable:** Participants had varied views on whether they would accept having LLM-enabled physical devices for health monitoring in their life. Most felt uncomfortable having it at home - especially in private spaces - because of the built-in camera, though a few (women) said they would accept it in a workplace setting. None initially suggested using it in a doctor's office, expressing concerns that it might distract clinicians, who should confirm rather than rely on the device's output. Several participants stressed that the device should be operated by someone trained to interpret its results, noting that it has limited value at home without such expertise and would be more useful in supporting family doctors' diagnostic work.
- **General perceptions in relation to LLM-enabled physical devices for health monitoring:** Participants expressed mixed perceptions of such devices. Some viewed the technology as necessary, emphasizing its advanced capacity to process large amounts of information through multiple sensors and cameras. Others highlighted that for the device to be truly reliable, its components must be well-integrated and programmed to ensure accurate and consistent results. In contrast, one participant opposed the idea altogether, fearing that the device could deliver alarming health information and noting that people vary in their psychological ability to cope with such messages.
- **Perceived risks in relation to being monitored:** Participants expressed concern about being digitally monitored. Some pointed out that targeted ads already appear immediately after offline interactions, suggesting constant surveillance. Others compared such tools to clairvoyants and felt that modern technologies not only listen but also "guide" user behaviour, reinforcing their worry that digital monitoring is already pervasive and intrusive.
- **How they would prefer to receive information from LLMs:** Participants preferred the device to communicate politely and conversationally, offering supportive explanations rather than blunt or alarming messages.

Conclusions:
Participants reported using various digital tools to monitor their health, acknowledging their usefulness while noting that frequent monitoring can provoke anxiety, especially when readings appear abnormal. Views on digital health technologies were mixed: some feared that close-to-body devices might disrupt the body's "internal energy," others believed such tools contribute to the subtle "robotization" of individuals, and some

stressed that the real risk lies in weak regulation rather than in the technologies themselves.

Trust in LLMs varied considerably, as some participants questioned their accuracy when compared with clinical measurements, whereas others recognized that new technologies often face initial resistance but eventually prove beneficial; trust in the tool in particular would increase if recommendations came from healthcare professionals (or, for some, from figures such as priests or clairvoyants). Participants highlighted that integrating such innovations requires professionals who understand both medicine and technology, while also preserving traditional skills essential when technology fails.

Perceptions of LLM-enabled devices ranged from seeing them as advanced and necessary, to doubting its reliability without proper sensor integration, to fearing that it could deliver frightening health information that users might not be psychologically prepared to receive. Most participants felt uncomfortable placing such devices in their homes, especially private spaces, though some would accept them at work; none initially recommended placing them in a doctor's office, arguing that clinicians should not rely on them at the expense of direct interaction, and several emphasized that only trained individuals should interpret their results, making them more suitable as a tool to support family doctors than as a household device. Although sharing personal data with digital tools was not perceived as a major risk, digital monitoring in general raised concerns, with participants noting pervasive online tracking and likening modern technologies to "guiding" rather than simply observing users.

Finally, participants expressed a preference for LLMs to communicate politely, conversationally, and supportively, offering optional clarifications rather than blunt or alarming messages.

Annex 2. Expert Focus Group

Methodology and Expert Panel

On May 11, 2026, INOMED organized a 2-hour online focus group within the CURTAIN project consortium to co-create and validate the guidelines for the use of AI in cancer literacy. The session brought together a multidisciplinary panel of 12 experts from key domains relevant to the safe integration of AI in healthcare, including academia, AI development, research and policy, data science, and clinical practice.

Prior to the session, participants received a structured overview of the project background and objectives. The discussion was facilitated using Mentimeter to present

guiding questions and capture initial responses, followed by an open, in-depth group discussion.

The set of 17 questions guiding the consultation was developed based on a triangulated approach, drawing on findings from the scientific literature, observed real-world practices, and insights collected from participants during the November 2025 Living Lab workshop. The questions were organized into three levels of implementation: Level 1 (Patients and Citizens), Level 2 (Healthcare Professionals), and Level 3 (Organizations). A final question (Q18) addressed future steps related to testing and validation of the framework.

Context, Questions and Interpretation of Results

Level 1: Guidelines for Patients and Citizens

Q1: Considering this rapid, AI-driven pattern of consuming medical information, out of 100 citizens and cancer patients, how many do you estimate currently have the health literacy required to critically evaluate the information they encounter and distinguish between a legitimate medical recommendation and a potentially harmful deepfake?

Q2: How should a red flag checklist look like to help citizens and cancer patients to spot a health-related deepfake or AI-generated hallucination?

Q3: How important is that the guidelines teach citizens and cancer patients to ask the proper questions (for example, by offering simple examples or templates), so they receive clearer, safer, and more useful answers from AI tools?

Q4: What information should be clearly communicated so people understand they are interacting with an AI system and that the output is not a medical diagnosis?

Q5: How can local community leaders be involved to help people better understand and appropriately interpret the risks and benefits of these technologies?

Level 2: Guidelines for Healthcare Professionals

Q6: What verification steps should be included in the guidelines to ensure doctors can quickly and reliably check the accuracy of AI-generated summaries?

Q7: How important is that the AI shows the logic process behind the decision?

Q8: How can AI tools be integrated into daily clinical workflows so that healthcare professionals benefit from their efficiency without becoming overly dependent on them?

Q9: How should the guidelines address situations where AI recommendations differ from a doctor's initial clinical judgment?

Q10: What specific information must an AI tool disclose (e.g., source of medical data, date of last update) for it to be trustworthy?

Q11: How should healthcare professionals be guided to respond when patients present self-generated "ChatGPT diagnoses" or other AI-generated health interpretations during clinical consultations?

Level 3: Guidelines for Organizations

Q12: How important is that healthcare organizations involve patients and citizens in the ongoing governance of AI, rather than just asking for their opinion once?

Q13: What role should a healthcare organization play in 'vouching' for specific AI tools? Would a 'certified safe' badge from a cancer society change a citizen's level of caution?

Q14: What level of personalization should organizations use in their AI systems? Should AI chatbots (used for appointments, medical information etc.) have access to the person's electronic health records and give recommendations based on it?

Q15: What governance rules should organizations establish to ensure that AI systems are transparent, explainable, reliable and embedded with strict data protection measures?

Q16: Who do you believe should be held responsible if an AI tool provides a recommendation that leads to an unsafe health choice?

Q17: What specific rules or practices should organizations follow when implementing automated health features at work (like 50-minute meeting limits or AI break reminders) so that employees feel supported rather than spied on?

Q18: How should we test these guidelines in a controlled environment?

Annex 3. Recommendations Factsheet used for validation

Citizens and Patients

Limited ability to identify AI-generated misinformation

Only a small percentage of citizens and cancer patients can distinguish between trustworthy medical information coming from experts, and AI-generated content.
Estimates ranged from fewer than 5% to approximately 30% of the population.

Key contributing factors:

- Low health literacy;
- Limited digital literacy;
- Emotional distress and anxiety;
- Exposure to misinformation through social media;
- Increasingly realistic AI-generated content.

Recommendation:

Look for clear AI warnings. In line with EU rules on AI transparency, AI tools should clearly explain that their answers are not medical diagnoses and may contain mistakes, outdated or even false information. These warnings should be easy to see and understand, permanently visible on the screen and not hidden in small text.

Citizens need practical warning signs

There is a growing need for simple and easy-to-use "red flag checklists" to help people recognize potentially misleading AI-generated health information.

Recommendation:

Learn to recognize warning signs ("red flags"). Citizens and patients should be able to identify potentially unsafe or misleading AI content. Warning signs include:

- The AI does not encourage consultation with a qualified doctor (*e.g. "Just follow these steps to manage your symptoms" omitting any recommendation to speak with the oncologist or your general practitioner*);
- It promotes supplements or paid services (*e.g. "Have you tried this immunity supplement?"*);
- It uses emotional, alarming, or urgent language (*e.g. "Act now before it is too late"*);
- It promises "miracle cures" (*e.g. "Follow this diet and the cancer won't spread in your body"*);
- It does not provide reliable medical sources (*e.g. cannot give the proper link to scientific papers, real oncology societies or World Health Organization, or it may hallucinate scientific references*);
- It shows unnatural images or voices (*e.g. a social media video shows a medical practitioner presenting themselves as an expert in their field, actively giving health advice, but when searching their medical background online, there is no other information to be found apart from several other similar videos*);
- It encourages long conversations mainly for commercial purposes (*e.g. the chatbot keeps asking questions to prolong the chat like "Tell me a bit more about your daily habits so I can tailor your profile"*).

AI should support understanding, not diagnosis

AI can help explain medical reports, procedures, and health information in simpler and easier language, so people can better understand them. However, AI should not be used for diagnosing diseases, understanding symptoms, making treatment decisions, or starting or stopping medications. AI should only be used as a tool for learning and explaining, not as a medical expert. Clear information is very important, and users must always know when they are talking to AI. AI systems should clearly say they are not doctors, explain that they can make mistakes or give wrong information, show warnings in a clear and visible way, and remind users to talk to qualified healthcare professionals.

Recommendation:

Understand the limits of prompting and know what is safe to ask. Current AI tools do not have built-in medical safety systems. Even well-formulated questions may lead to misleading answers if important medical details are missing. Safe use does not depend only on how the user formulates or writes a question. In line with EU recommendations for high-risk AI, these tools should act only as educational support to help users understand complex information (e.g. "explain what this medical term means"). They must never replace a medical professional or be used for decision-making, such as

diagnosing symptoms, or choosing to stop or change a cancer treatment (e.g. "should I take this medication").

Community leaders can support health literacy

Trusted local figures such as community representatives, patient advocates, religious leaders, and patient organizations can help citizens better understand health information when they receive proper training and support. They can play an important role in explaining both the benefits and risks of using AI for health information and in guiding people towards safer and more informed use of digital health tools.

Recommendation:

Seek guidance from trusted and trained community resources. Citizens and patients are encouraged to discuss the use of digital health tools with trained community representatives, patient organizations, or trusted local support networks. These resources can help people better understand both the benefits and risks of using AI for health information.

Healthcare Professionals

Verification remains essential

Any AI-generated medical summaries, references, or citations must always be checked against trusted clinical guidelines and peer-reviewed scientific literature before being used in practice. A key risk identified is that AI can produce convincing but completely false references, including invented authors and publications, making independent verification essential.

Recommendation:

Verify everything with real evidence. In line with the European Commission guidance, clinicians should carefully check any AI-generated summaries, references, or citations against trusted clinical guidelines and peer-reviewed medical literature. Professionals must remember that AI often invents highly realistic but completely fake medical references and authors.

AI can support administrative and information tasks in healthcare

AI can be useful for supporting, however not for replacing, clinical work. Appropriate uses include summarizing medical literature, extracting evidence, organizing information, administrative assistance, and voice-to-text transcription. In no case AI must not be used as a substitute for professional clinical judgment.

Recommendation:

Understand the limits of AI in clinical workflows. Reflecting EU rules on high-risk AI use, healthcare providers should only use general AI tools for well-defined administrative support, like extracting evidence, summarizing long texts, or transcribing voice-to-text dictations. These tools must be clearly distinguished from certified clinical decision-support systems, and they should never replace professional medical judgment.

Human responsibility must remain central

The final responsibility for clinical decisions must always remain with healthcare professionals. When AI outputs differ from clinical judgment, additional verification should be performed. AI can make mistakes, just as clinicians can make mistakes, and decisions should always be based on evidence.

Recommendation:

Maintain clinical responsibility and final decision-making. Following the EU principle that humans hold ultimate responsibility, any disagreement between AI output and clinical judgment should trigger further review and verification. AI recommendations should not be accepted automatically, but differences should also not be ignored. Clinical decisions must always remain under professional control.

AI conversations can become opportunities for trust-building with patients

Patients increasingly use AI tools before consultations and may present AI-generated interpretations of symptoms or diagnoses. Rather than dismissing these inputs, clinicians should use them to support communication, acknowledge patient engagement, and improve understanding. Clear explanations of AI limitations and guidance toward reliable sources can strengthen trust and health literacy.

Recommendation:

Communicate with empathy and openness. Healthcare professionals should be prepared for patients who arrive with AI-generated information or self-diagnoses. Instead of dismissing these inputs, clinicians should use them as an opportunity to build trust (*for example, by saying "Let's look at this together"*). Patients' efforts should be acknowledged, while clearly and calmly explaining the limits of AI in simple terms and guiding them toward reliable medical sources. When time is limited, patients should be advised to seek a second medical opinion where appropriate.

Organizations

Patient involvement should be continuous

Both citizens and patients should be involved in AI governance as an ongoing process rather than through one-time consultations.

Recommendation:

Make sure patients are continuously involved in decisions. Unique to this framework is the recommendation that organizations actively and continuously involve patients and citizens in AI governance. Healthcare institutions should integrate public feedback into ongoing decision-making processes rather than treating it as a single, one-time consultation.

Avoid creating over-trust in AI

Healthcare organizations officially endorsing specific AI tools are prone to raise concerns. Labels such as "certified safe" may unintentionally create a false sense of security and reduce critical evaluation by users.

Recommendation:

Avoid endorsing specific AI tools. To prevent users from developing a false sense of security, institutions should refrain from officially endorsing open-access AI tools or

applying absolute labels like "certified safe." Such endorsements can create over-trust and discourage patients and staff from critically evaluating the AI-generated information they encounter.

Strong governance and accountability are required

There is a crucial need for clear legal frameworks, human accountability, consistent regulation, and continuous human oversight. AI systems should not operate independently in healthcare environments and should be continuously monitored.

Recommendation:

Keep strict data boundaries. In line with EU data protection and cybersecurity rules, as well as the responsible use of generative AI in healthcare, organizations must clearly separate administrative AI tools from regulated clinical systems. Open AI tools must never access sensitive patient records or other protected health data. Any system that processes personal health information must meet strict security and certification standards, with strong safeguards against data breaches, unauthorized access, and cyberattacks, and must ensure clear human responsibility and oversight.

Workplace AI should support employees, not monitor them

AI-based wellness tools should be transparent, voluntary (opt-in), and protect employee privacy while avoiding surveillance practices. Data generated through these tools should not be accessible to managers or human resources departments.

Recommendation:

Implement ethical workplace features. When using AI-based tools to support staff well-being (for example, reminders for breaks or workload management), organizations must ensure full transparency. These features should be optional and not enabled by default. To avoid any form of pressure or monitoring, institutions must clearly state that there are no consequences for ignoring these prompts, and that human resources or management cannot access any individual-level wellness data.

Key Messages

- ✓ Limited health and digital literacy, makes it difficult for many citizens and patients to identify reliable medical information and distinguish it from AI-generated content.

- ✓ AI should be used only to support understanding of health information, not for diagnosis, treatment decisions, or medication management.
- ✓ Continuous human judgment, oversight and verification are required, with final responsibility always remaining with healthcare professionals.
- ✓ Clear transparency, guidance on misinformation risks, and strong governance are essential for safe and responsible use of AI in healthcare.

Disclaimer

CURTAIN (Beating Cancer Inequalities through Literacy in Europe) is a project co-funded by the European Union's EU4Health programme under grant agreement No 101219448. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

Acknowledgements

The authors would like to thank all CURTAIN Consortium partners involved in Work Package 3 (WP3) and the Citizen's Jury of Lerești Living Lab for their valuable contributions, provision of expertise and guidance throughout the development of the guidelines.

Annex 4. Validation Results of the Recommendations Across Participant Groups

This annex summarizes the feedback collected during the validation of the recommendations across four participant groups: students (health communication, psychology, social sciences), pharmacists, older adults living in rural areas, and older adults. Students and pharmacists assessed the full set of recommendations, while the two older-adult groups evaluated only the section addressed to citizens and patients. Before providing their feedback, participants received a Microsoft Forms questionnaire, a Romanian-language factsheet, and the version of the guideline available at the time of validation. These materials were provided to ensure that participants had a clear understanding of the recommendations, could review the relevant information in an accessible format, and could assess the guideline based on the same reference documents. The questionnaire and supporting materials were distributed online to the four participant groups through designated representatives, who facilitated their dissemination within the respective groups. Responses were collected electronically and

analyzed by participant category. The main findings from each group are presented below.

Participant group	Number of responses
Students	7
Pharmacists	2
Older adults	9
Older adults living in rural areas	4

Students:
Within the student group, the language used in the recommendations was generally considered easy to understand, and the visual structure of the document was evaluated positively, although opinions regarding its length were more varied. The list of "warning signs" was perceived as relatively clear; however, several improvements were suggested, such as including the tendency of AI to confirm or agree with the user, emphasizing more clearly that AI systems may provide inaccurate or unpredictable information, and distinguishing AI-specific warning signs from more general indicators of misleading content. Several participants considered that the recommendations could be difficult to follow for people with limited technological experience, suggesting the need for wording that is as simple and accessible as possible. For healthcare professionals, the recommendations were generally considered applicable, with the main barriers mentioned being lack of time, work overload, high patient volumes, and the risk of placing too much trust in AI-generated responses in situations of clinical uncertainty. The need to carefully verify sources provided by AI was also highlighted, including checking whether the generated claims are actually supported by the content of the cited articles. At the organizational level, responses were more heterogeneous, with difficulties mentioned regarding the integration of the recommendations into existing procedures, transparency, leadership, and the feasibility of their practical implementation.

Pharmacists:
Within the pharmacist group, the language and structure of the recommendations were generally evaluated favorably, and the respondents considered the length of the document appropriate for a quick and efficient read. The list of "warning signs" was positively assessed both in terms of clarity and its ability to help identify potentially misleading content, with no additional specific warning signs proposed. Respondents considered that the document clearly distinguishes, to a large extent, between situations in which AI can be used safely and those in which a medical opinion is necessary, and that the recommendations address real situations encountered by citizens and patients. With regard to healthcare professionals, the recommendations were considered realistic and applicable, the distinction between administrative tasks and clinical decision support was viewed as clear, and the need to verify AI-generated results and maintain

professional responsibility was well emphasized. At the organizational level, responses were more varied regarding the integration of the recommendations, continuous collection of feedback, and the recommendation not to officially endorse specific AI tools, while data protection, governance, and transparency were evaluated favorably.

Older adults:
Within the group of retired participants, the material intended for citizens and patients was generally evaluated favorably, although there were some differences regarding ease of understanding. Some participants considered the language easy or very easy to understand, while others rated it as neither difficult nor easy, whereas the visual structure was evaluated from moderately to very favorably. The length of the document was considered appropriate or very appropriate by all respondents, and what users should expect when interacting with an AI tool was considered clear or very clear. The list of "warning signs" was generally viewed as easy to understand or neutral in terms of difficulty, while its usefulness in identifying misleading content was predominantly evaluated to a moderate or large extent. Regarding possible additional warning signs, most participants did not provide specific examples, although a few considered that other elements might be identified at a later stage. The distinction between the use of AI and situations in which a medical opinion is required was evaluated to a moderate or large extent. Responses were more varied regarding how easily people with limited technological experience could follow the recommendations, with assessments ranging from difficult to easy. The recommendations were generally considered relevant to real-life situations, and most participants did not propose any additional recommendations.

Older adults living in rural areas:
For the older adults living in rural areas, the results indicate an overall favorable evaluation of the material intended for citizens and patients. The language was considered easy or very easy to understand by all participants, while the visual structure and length of the document were viewed as appropriate. The clarity regarding what users should expect when interacting with an AI tool was evaluated positively by most respondents, although one participant expressed a neutral opinion. The list of "warning signs" was generally considered clear, and its ability to help identify misleading content was rated from moderate to very favorable. No additional "red flags" were proposed. The distinction between situations in which AI can be used and those in which a medical opinion is required was considered moderately to clearly presented. The greatest variation in responses concerned how easily a person with limited technological experience could follow the recommendations, with assessments ranging from "difficult" to "very easy." Overall, the recommendations were perceived as relevant to real situations encountered by citizens and patients, with no additional recommendations suggested.

Frequently used AI terms (plain-language glossary)

- **Artificial Intelligence (AI):** Technology that simulates human intelligence, enabling computers to learn, reason, and perform tasks typically requiring human intellect [118].
- **Prompting:** Giving instructions or questions to an AI model to guide its output [119].
- **Artificial Narrow Intelligence (ANI):** "Weak AI" specializing in one task or a narrow domain, unable to transfer skills elsewhere [120].
- **Generalist Artificial Intelligence (GAI):** AI using complex machine learning to generate original text, images, or other content by identifying deep structural patterns [121].
- **Transformers:** Deep learning architecture using "self-attention" to relate every word in a sentence to every other, capturing context across long passages [122].
- **Pre-training:** Initial training on large datasets to learn general language/data patterns before task-specific adaptation [123].
- **Fine-tuning:** Further training a pre-trained model on a smaller, specific dataset to specialize it [124].
- **Prompting Paradigms:** Zero-shot (no examples), few-shot (guided by examples), chain-of-thought (step-by-step reasoning) [125].
- **Large Language Model (LLM):** AI systems trained to read, understand, and generate human language for tasks like answering questions or translation [126].
- **Retrieval-Augmented Generation (RAG):** Framework that retrieves verified external data before generating a response, improving accuracy over static pre-trained parameters [127].
- **Agentic AI Systems:** Autonomous architectures using coordinated AI agents to pursue complex objectives with minimal human oversight [128].
- **Chatbots:** Software simulating human conversation via text or voice [129].
- **AI Hallucination:** When a generative model produces confident but factually inaccurate or fabricated output [130].
- **Medical Digital Twin (MedDTs):** A dynamic virtual replica of a patient, continuously updated to simulate health changes and predict treatment response [131].
- **Medical Deepfake:** AI-generated synthetic audio, video, or images imitating real medical content or professionals, potentially used for misinformation [132].