ACCSCIENCE PUBLISHING

Eurasian Journal of Medicine and Oncology

PERSPECTIVE ARTICLE

Artificial intelligence as a copilot, not autopilot: A framework for responsible integration in pediatric hematology-oncology

Ioannis Kyriakidis*

Department of Pediatric Hematology-Oncology & Autologous Hematopoietic Stem Cell Transplantation Unit, University Hospital of Heraklion, and Laboratory of Blood Diseases and Childhood Cancer Biology, School of Medicine, University of Crete, Heraklion, Greece

Abstract

Artificial intelligence (AI) is transforming pediatric hematology-oncology. Current applications include pattern recognition in blood and marrow images, inference of tumor biology, relapse risk prediction, integration of complex clinical data, and support for writing, learning, and communication. Although often framed as a single technological revolution, these tools differ in purpose, evidentiary basis, and risk profile, yet they share a central consequence: the redistribution of cognitive work and clinical authority. Adoption should follow a "copilot, not autopilot" model in which AI extends perception and expertise while preserving accountable clinical judgment, rigorous formative reasoning in trainees, confidentiality, and equitable access to validated benefits. This demands pediatric- and setting-specific validation, transparent disclosure of AI use, systematic source verification, post-deployment monitoring, and graduated trust proportional to clinical stakes. AI may widen access to knowledge and reduce workload, but its value depends on disciplined verification and the clinician's authority to override or decline its use.

Keywords: Artificial intelligence; Machine learning; Decision support systems; Pediatric hematology-oncology; Education; Medical; Health inequities

***Corresponding author:**
Ioannis Kyriakidis
(kyriakidis@med.uoc.gr)

Citation: Kyriakidis I. Artificial intelligence as a copilot, not autopilot: A framework for responsible integration in pediatric hematology-oncology. *Eurasian J Med Oncol*. 2026;10(5):026290355. doi: 10.36922/EJMO026290355

Received: July 14, 2026

Revised: August 12, 2026

Accepted: August 14, 2026

Published online: August 19, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction: from prediction to partnership

Over the past decade, artificial intelligence (AI) in pediatric hematology-oncology has evolved from specialized relapse-prediction studies to systems that recognize disease-associated morphology, infer molecular biology from images, estimate individual risk, and generate text that now accompanies everyday clinical and academic work.[1] The speed of this transition is easy to celebrate and difficult to govern. AI is already changing what clinicians can see, what they can predict, how they access knowledge, and how future specialists learn to reason.

Clinicians trained before generative AI became routine now experience this transition from both sides: these tools can organize complex information, improve scientific communication, and reveal overlooked connections, yet fluent output can easily be mistaken for genuine understanding. The key question is no longer whether AI will enter the field, but how far to delegate cognitive tasks and clinical authority.

The appropriate stance is neither celebratory nor defensive. Pediatric hematology-oncology should adopt AI as a copilot: a system that extends perception, memory, and analytic capacity while remaining subordinate to an accountable professional who understands the task and can reject the output. An autopilot culture begins when recommendations are accepted simply because they are fast, persuasive, or usually correct. In a field of rare diseases, small biological subgroups, consequential decisions, and decades of survivorship, "usually correct" is not an adequate standard.

2. One label, several revolutions

AI encompasses a range of technologies rather than a single solution. Predictive models, computer vision, automated image analysis, and large language models (LLMs) each address different questions, generate distinct types of evidence, and have unique limitations. Viewing them as interchangeable both underestimates their potential and overlooks their individual risks.

Clinical advances are substantial (Figure 1). Early machine-learning work in childhood acute lymphoblastic leukemia (ALL) showed that multidimensional data could predict relapse.[2] A systematic review later mapped applications across leukemia, central nervous system tumors, and solid tumors, highlighting limited external validation and uncertain clinical applicability.[3] In adult hematologic disease, computational analysis of peripheral blood smears identified disease-associated cytomorphologies, a transferable approach that still requires pediatric validation.[4] MRI-based models can infer medulloblastoma molecular subgroups preoperatively.[5] Deep learning applied to Giemsa-stained marrow aspirates can predict B- or T-lineage and *ETV6*::*RUNX1* status, suggesting that morphology contains biological information beyond conventional visual classification.[6] In a held-out test set, PREPARE-ALL achieved higher relapse sensitivity than three clinicians using multicenter childhood ALL data.[7]

These advancements transcend mere rapid classification. The blood smear evolves into a source of computational phenotypes; the marrow image serves as a proxy for genotype; and the longitudinal record represents a risk trajectory. AI has the potential to expand the scope of questions we can ask, beyond merely accelerating existing inquiries. It may also facilitate the integration of clinical, imaging, morphological, and molecular data layers that no single observer could consistently unify.

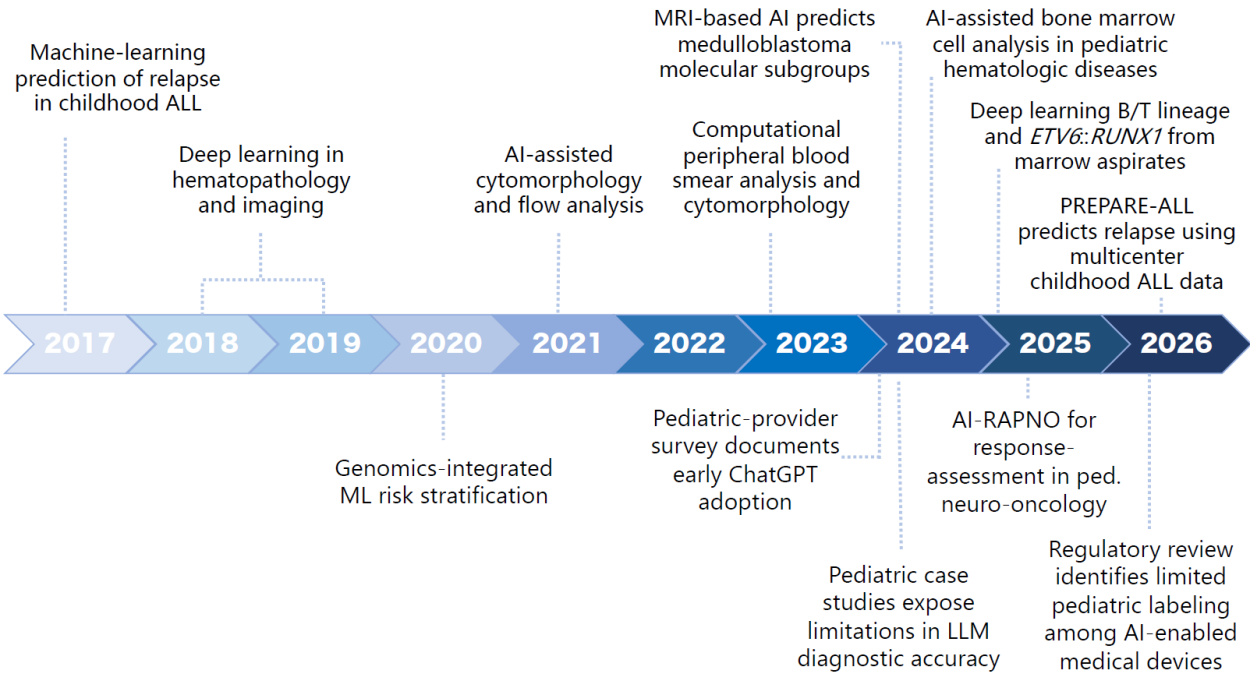


Figure 1. From prediction to partnership: Selected developments in artificial intelligence relevant to pediatric hematology-oncology, 2017–2026. The timeline combines publication-specific milestones with broader phases of development and is selective rather than exhaustive.[1-10,13]
Abbreviations: AI: artificial intelligence; AI-RAPNO: Artificial Intelligence for Response Assessment in Pediatric Neuro-Oncology; ALL: Acute lymphoblastic leukemia; LLM: Large language model; MRI: Magnetic resonance imaging.

However, the most compelling performance metric is not always the most clinically meaningful one. High discrimination in a retrospective cohort does not equate to calibration in a new population, benefits at the point of care, or improvements in patient outcomes. A model that flags slides for review is not synonymous with one that provides a diagnosis; similarly, a model estimating relapse risk is not the same as one recommending treatment. Prior to clinical deployment, systems must demonstrate external and pediatric validation, subgroup performance, transparent handling of missing data, and prospective evaluation within the workflow where their output will influence decision-making. Implementation is not marked by model availability but by the clinical team's ability to understand its function, limitations, and accountability.

Generative AI demands careful consideration. Unlike specialized, validated models, an LLM produces plausible text by prediction. Its flexibility is what makes it valuable for tasks like literature mapping, drafting, coding, summarization, communication, and education, but this same flexibility also makes its errors hard to control. The interface can improve a paragraph, generate a citation, simplify an explanation for a family member, or even create a convincingly false clinical statement. Fluency is not a substitute for evidence.

3. Training must protect the formation of judgment

The educational benefits and potential harms of generative AI stem from the same core function. It can assist trainees in clarifying complex concepts, comparing differential diagnoses, enhancing scientific English, preparing presentations, or identifying search terms for literature reviews. Survey data already indicate significant use of LLMs among pediatric healthcare providers, especially trainees.[9] Such support has the potential to reduce unnecessary clerical burdens and increase access to advanced educational resources.

However, expertise is not solely gained from receiving correct answers. It is developed through constructing problem representations, identifying relevant information, tolerating uncertainty, comparing alternative explanations, consulting original evidence, and recognizing when a coherent conclusion does not apply to the child in front of us. The effort to reach an answer is not wasted; it constitutes part of the learning process. If AI handles these steps initially, trainees might become adept at editing conclusions without understanding how to generate and test them.

A human-centered sequence is therefore needed. The learner should first formulate the question, review primary material, and reach a preliminary interpretation. AI can then identify omissions, challenge assumptions, generate counterarguments, or improve communication. The trainee should explain what was accepted or rejected and why. Used this way, AI supports metacognition rather than replacing reasoning—a critical distinction given the challenges of pediatric cases in which fluent LLM responses remained diagnostically insufficient.[10]

Academic governance should follow the same principles. Detector-based policing is unreliable and can be unfair; widely used detectors have shown bias against non-native English writing.[11] In an international discipline, concise or simple English should not be regarded as misconduct. The more appropriate standard is auditability. When AI significantly contributes to a manuscript, presentation, or educational product, its use must be disclosed; all claims and citations should be verified against original sources; confidential patient information must never be entered into unsecured public systems; and the designated author should be able to justify their reasoning and final wording.

This boundary is neither permissive nor punitive. AI may aid in expression, organization, comparison, and critique. It should not generate evidence, draw conclusions that the user cannot independently defend, or hide gaps in understanding. In the era of AI, the most competent clinician will be the one who recognizes what still needs verification, rather than the fastest to obtain answers.

4. Equity requires more than access

The assertion that AI will democratize expertise warrants careful examination. Merely providing access to the same interface does not ensure safe and competent use. Successful implementation relies on robust infrastructure, local data, language capabilities, cybersecurity, ongoing maintenance, regulatory support, and access to specialists capable of identifying unlikely outputs. An NIH portfolio analysis of AI-for-health research relevant to low- and middle-income countries identified major gaps in locally grounded development and evaluation.[12]

These limitations are particularly pronounced in pediatric settings. Pediatric hematology-oncology is not simply a scaled-down version of adult oncology; factors such as age-related physiology, rare molecular subtypes, evolving treatment protocols, limited datasets, and extended survivorship significantly influence the applicability of findings. The availability of regulated AI-enabled medical devices for pediatric use remains limited, reflecting the tendency for children to inherit technologies primarily developed and tested for adults.[13] A system trained in a well-resourced center may perform differently in environments with varying laboratory

platforms, imaging systems, supportive care, population demographics, or therapeutic options.

Generative AI has the potential to mitigate certain inequities in healthcare. It can facilitate translation, clarify unfamiliar terminology, assist in developing search strategies, and offer clinicians a structured initial framework. However, it can generate fabricated references[14] and plausible misinformation, with risks magnified where expert verification is scarce. Therefore, equity should be defined not just by access to AI models, but by ensuring access to validated, secure, and contextually appropriate benefits.

A responsible equity agenda in medical AI should encompass local and multilingual validation, the use of representative datasets governed appropriately, and comprehensive training for users and supervisors. Procurement standards must mandate subgroup performance evaluation and post-deployment monitoring. Global partners should actively participate in defining clinical problems and share in developing capabilities, rather than merely supplying data for externally developed systems. AI implementation should aim to distribute capacity broadly, avoiding the creation of dependencies or widening gaps between institutions capable of interrogating models and those expected to rely on them. Maintaining these principles is essential to ensure equitable and effective deployment of AI in healthcare.

5. Graduated trust as an operating principle

These concerns about evidence, training, and equity converge on a practical alternative to blanket acceptance or prohibition: graduated trust. Low-stakes uses (such as language editing, administrative drafting, coding assistance, literature mapping, educational question generation, and non-diagnostic image preprocessing) can be implemented earlier, provided confidentiality is maintained and outputs are verified. Moderate-stakes applications, including case prioritization, longitudinal summarization, or differential diagnosis suggestions, require validated workflows, explicit acknowledgment of uncertainty, and routine expert review. High-stakes uses (such as diagnosis, prognosis, treatment selection, trial eligibility, and response assessment) must meet the highest evidentiary standards and should never obscure the identity of the final decision-maker.

Oversight should increase proportionally with both the potential consequences of an error and the difficulty of detecting it. While an awkward sentence is straightforward to correct, a plausible false citation presents a greater challenge. Some segmentation errors may be visually apparent, whereas a convincing risk score could subtly influence care, especially when the patient lies outside the training-data distribution. Oncologists have already emphasized the importance of explainability, bias mitigation, consent, and accountability in AI-supported healthcare systems.[15] Therefore, clinically meaningful transparency should encompass details such as the intended population, development and validation settings, uncertainty measures, subgroup performance, principal failure modes, and the necessary actions when model outputs conflict with clinical evidence.

Before deploying an AI system in clinical care, four critical questions must be addressed: What specific task is being delegated? Has the system been validated in the target population? How will errors, inequities, or model drift be detected? Who has the authority to override or suspend its use? Continuous governance is essential after deployment, as protocols, data distributions, laboratory platforms, and user behaviors evolve. All near misses and overrides should be documented, and use should be halted if performance declines. The AI-RAPNO initiative in pediatric neuro-oncology exemplifies this approach, emphasizing not only technical progress but also the shift from admiring capabilities to establishing standards for disciplined integration.[8]

Finally, the human clinician should not be retained merely as a ceremonial signatory. A "human in the loop" who lacks sufficient time, information, expertise, or authority to contest outputs does not provide genuine oversight; instead, they assume liability. Effective copilot practice requires dedicated time for review, clear avenues for challenging outputs, and institutional recognition that sometimes abstaining from use is the safest course.

6. Concluding remarks

Taken together, these principles form a clinical and educational argument, not merely a technical one. The key skill in the coming decade will be knowing when to trust, challenge, or decline AI output. Clinicians trained before this transition should not defend a fading pre-AI paradigm but should help identify which forms of expertise must endure.

AI can broaden access to knowledge, diminish avoidable cognitive load, and identify patterns that human observers might overlook. However, it cannot independently develop expertise, address systemic inequities, or ensure safety in under-resourced systems. Achieving these goals requires pediatric and local validation, transparent application, graduated trust, and accountable human authority. In pediatric hematology-oncology, AI should be regarded as legitimate when it is verified, useful when supervised, and powerful when functioning as a copilot rather than an autopilot. The future of this field will depend not on

the frequency of AI utilization, but on how rigorously clinicians understand when not to rely on it.

Acknowledgments

None.

Funding

None.

Conflict of interest

The author declares no conflicts of interest.

Author contributions

This is a single-authored article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

Further disclosure

Grammarly Pro was used to enhance language, grammar, and readability. The author reviewed and edited the output to ensure scientific integrity and takes full responsibility for the content.

References

1. Wyatt KD, Alexander N, Hills GD, *et al*. Making sense of artificial intelligence and large language models-including ChatGPT-in pediatric hematology/oncology. *Pediatr Blood Cancer*. 2024;71(9):e31143.

   doi: 10.1002/pbc.31143

2. Pan L, Liu G, Lin F, *et al*. Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia. *Sci Rep*. 2017;7(1):7402.

   doi: 10.1038/s41598-017-07408-0

3. Ramesh S, Chokkara S, Shen T, *et al*. Applications of Artificial Intelligence in Pediatric Oncology: A Systematic Review. *JCO Clin Cancer Inform*. 2021;5:1208-1219.

   doi: 10.1200/CCI.21.00102

4. de Almeida JG, Gudgin E, Besser M, *et al*. Computational analysis of peripheral blood smears detects disease-associated cytomorphologies. *Nat Commun*. 2023;14(1):4378.

   doi: 10.1038/s41467-023-39676-y

5. Wang YRJ, Wang P, Yan Z, *et al*. Advancing presurgical non-invasive molecular subgroup prediction in medulloblastoma using artificial intelligence and MRI signatures. *Cancer Cell*. 2024;42(7):1239-1257.e7.

   doi: 10.1016/j.ccell.2024.06.002

6. Piven A, Shamai G, Elitzur S, *et al*. Prediction of B/T Subtype and ETV6-RUNX1 Translocation in Pediatric Acute Lymphoblastic Leukemia by Deep Learning Analysis of Giemsa-Stained Whole Slide Images of Bone Marrow Aspirates. *Pediatr Blood Cancer*. 2025;72(8):e31797.

   doi: 10.1002/pbc.31797

7. Saravanan S, Rengaswamy R, Narula G, *et al*. PREPARE ALL: An Artificial Intelligence Tool for Predicting Relapse in Children With Acute Lymphoblastic Leukemia. *JCO Clin Cancer Inform*. 2026;10(1):e2500222.

   doi: 10.1200/CCI-25-00222

8. Kazerooni AF, Familiar AM, Aboian M, *et al*. Artificial Intelligence for Response Assessment in Pediatric Neuro-Oncology (AI-RAPNO), part 2: challenges, opportunities, and recommendations for clinical translation. *Lancet Oncol*. 2025;26(11):e607-e618.

   doi: 10.1016/S1470-2045(25)00489-9

9. Kisvarday S, Yan A, Yarahuan J, *et al*. ChatGPT Use Among Pediatric Health Care Providers: Cross-Sectional Survey Study. *JMIR Form Res*. 2024;8:e56797.

   doi: 10.2196/56797

10. Barile J, Margolis A, Cason G, *et al*. Diagnostic Accuracy of a Large Language Model in Pediatric Case Studies. *JAMA Pediatr*. 2024;178(3):313-315.

    doi: 10.1001/jamapediatrics.2023.5750

11. Liang W, Yuksekgonul M, Mao Y, Wu E, Zou J. GPT detectors are biased against non-native English writers. *Patterns (N Y)*. 2023;4(7):100779.

    doi: 10.1016/j.patter.2023.100779

12. Forsyth AD, Povlich LK, Kilmarx PH. Use of artificial intelligence for health science in low- and middle-income countries: NIH portfolio landscape, gaps and opportunities. *BMJ Open*. 2026;16(3):e112394.

    doi: 10.1136/bmjopen-2025-112394

13. Zapotoczny G, Goyal A, Christmas M, Qazi S, Carroll M, Espinoza J. FDA-Regulated AI-Enabled Medical Devices With Pediatric Indications. *JAMA Netw Open*. 2026;9(3):e262636.

    doi: 10.1001/jamanetworkopen.2026.2636

14. Moskatel LS, Zhang N. Comparative prevalence and characteristics of fabricated citations in large language models in headache medicine. *Headache*. 2024;64(1):93-95.

    doi: 10.1111/head.14638

15. Hantel A, Walsh TP, Marron JM, *et al*. Perspectives of Oncologists on the Ethical Implications of Using Artificial Intelligence for Cancer Care. *JAMA Netw Open*. 2024;7(3):e244077.

doi: 10.1001/jamanetworkopen.2024.4077