

AI IN ONCOLOGY

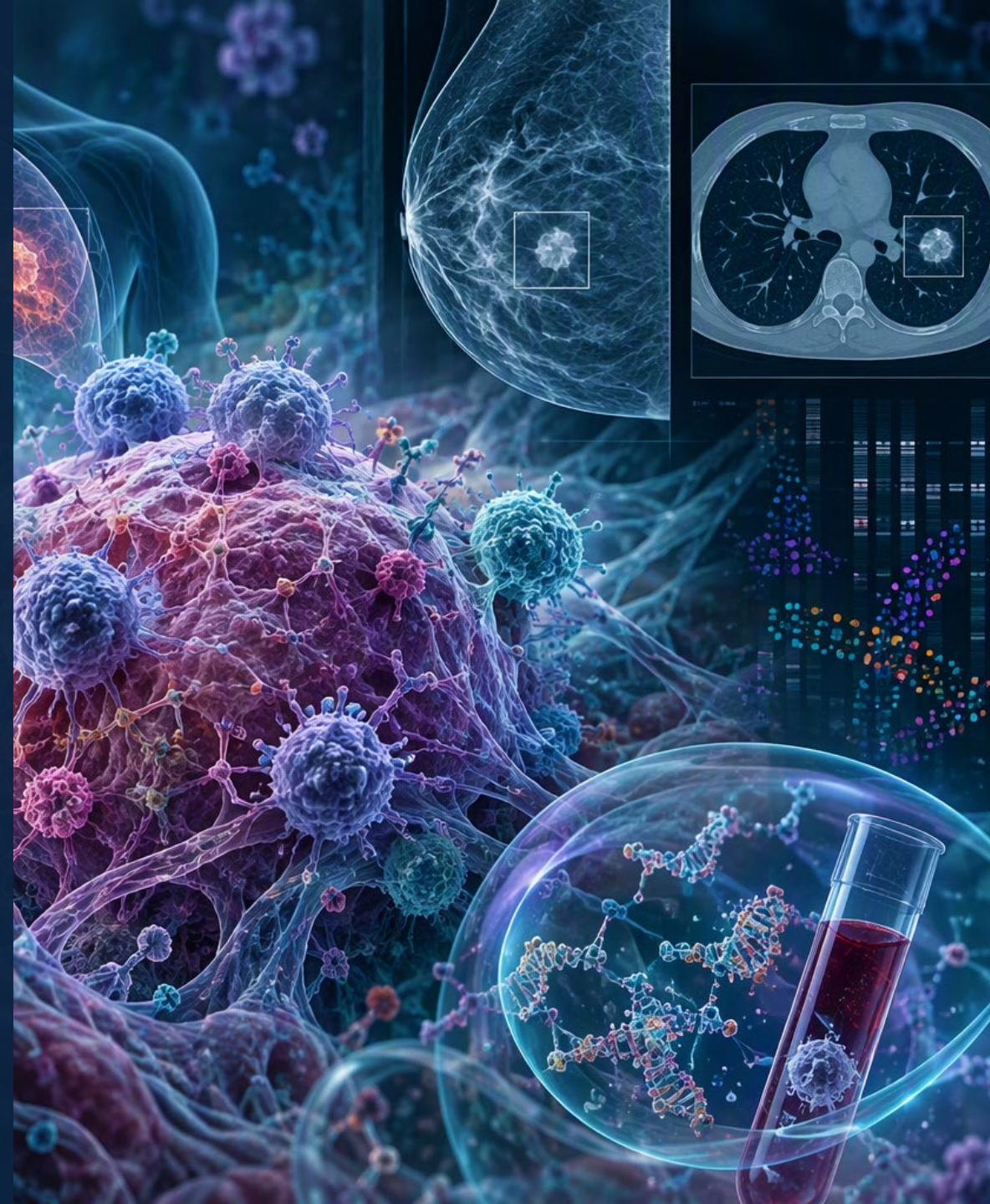
Artificial Intelligence in Oncology: How AI is changing Cancer Detection and Treatment.

From point solutions to learning systems of care

Jason Hill, MD, MMM, MS

September 2026

System Vice Chair, Hospital Medicine and Innovation



The story in one slide

AI is moving from “find the lesion” to “orchestrate the journey.”



Bottom line: the competitive advantage is not the model. It is the integrated clinical workflow, evidence loop, and accountable owner.

Detection: strong signal, uneven clinical value

MASAI, PRAIM, ACCEPT, and NHS-Galleri: what the 2025-2026 evidence supports

Mammography AI

MASAI (n=105,934): interval cancer non-inferior, 1.55 vs 1.76 per 1,000 (PR 0.88)

Sensitivity 80.5% vs 73.8% (p=0.031); specificity identical at 98.5%

PRAIM (n=463,094): detection 6.7 vs 5.7 per 1,000 (+17.6%); PPV 17.9% vs 14.9%

Reduction in screen reading workload by 44%

Clinical posture

DEPLOY + MONITOR

Colonoscopy CAdE

CAdE raises adenoma detection, but the gain is workflow-dependent

ACCEPT: unassisted ADR fell 28.4% to 22.4% after AI exposure (p=0.0089)

Deskilling is a measured risk, not a theoretical one

Track unassisted performance, not just AI flags

SCALE SELECTIVELY

Blood-based MCED

NHS-Galleri (~142,000 adults, 50-77) missed its primary endpoint

Stage III/IV incidence unchanged: IRR 1.03 (0.92-1.14), p=0.63

Stage IV alone lower: IRR 0.86 (0.74-1.00) - hypothesis-generating

Outcome evidence still required before adoption

RESEARCH ONLY

IMAGING AI

The implementation question is no longer “Does it detect?”

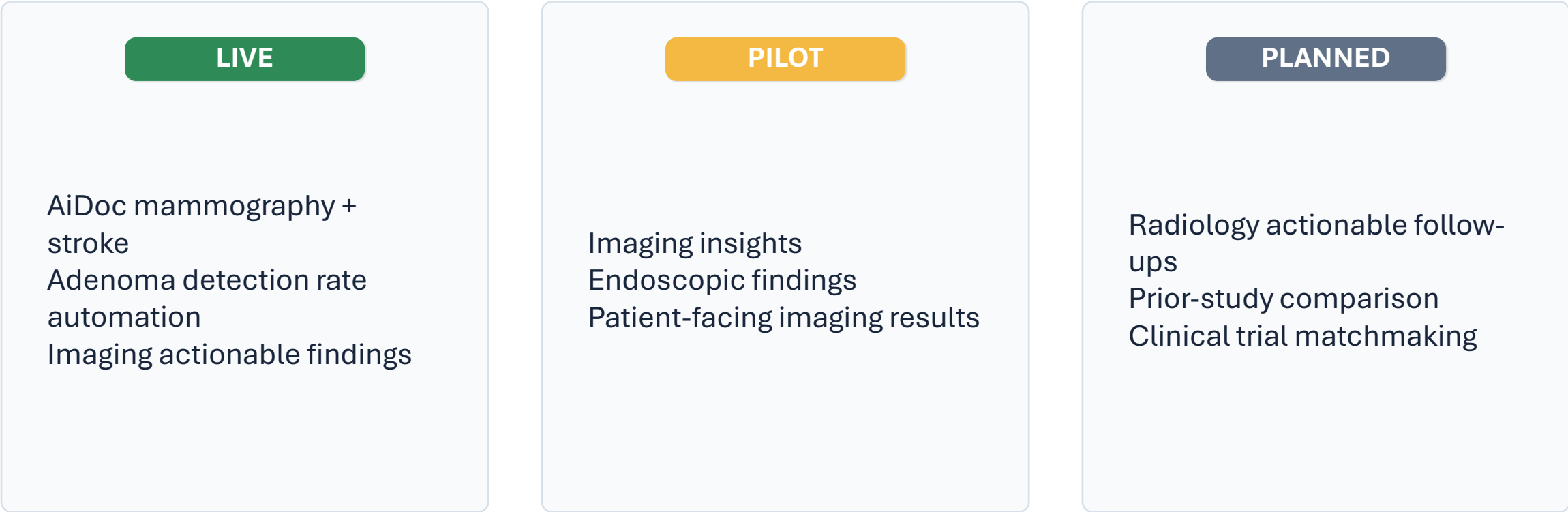
It is “What changes at the workstation, in the reading list, and downstream?”

- Prioritization without alert fatigue**
- Standardized measurement and follow-up**
- Site-by-site validation across equipment and populations**
- Explicit ownership of false negatives and drift**



Ochsner reality: a portfolio, not a pilot

The enterprise roadmap already spans detection, workflow, and follow-up.



Strategic implication: consolidate around shared workflow infrastructure, measurement, and governance. Do not fund disconnected algorithms.

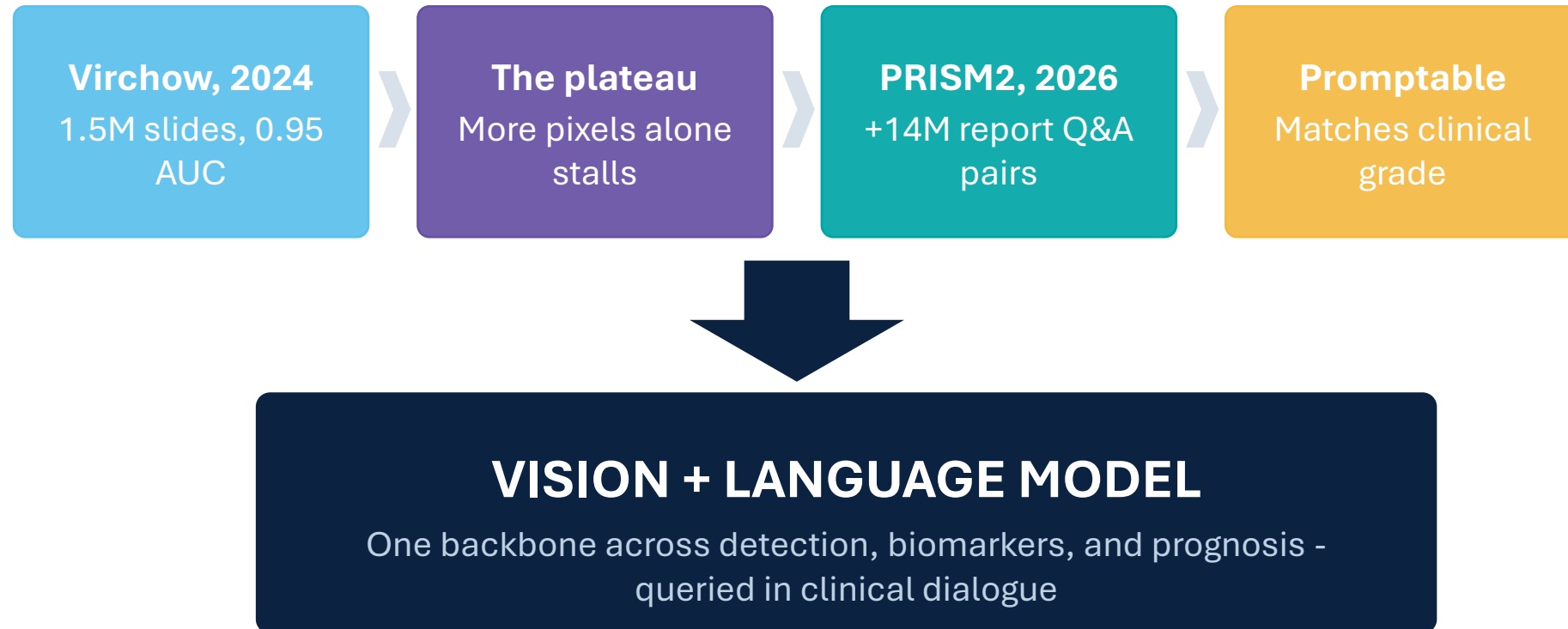
Detection is only the front door



The next wave is multimodal: pathology, imaging, genomics, notes, and longitudinal response interpreted together.

Diagnosis: multimodal AI changes the unit of analysis

Virchow to PRISM2: from scaling pixels to fusing the pathology report



Equity gap: 3.0% breast, 10.9% lung, 16.0% IDH1 AUC differences by race; TCGA is 82% White. Stratify before deployment.

Prostate Biomarkers: Separating Prognosis from Predictive Treatment Utility

The MMAI (ArteraAI) digital-pathology biomarker as a template — and its limits

1 • The core distinction

Prognostic — risk of an outcome (metastasis, death). Relative benefit is ~constant across risk; only absolute benefit (NNT) shifts.

Predictive — who derives differential benefit from a specific therapy. The higher-value, higher-bar claim (NCCN v7.2026).

2 • What the MMAI biomarker is

Deep learning on digitized H&E biopsy slides + routine clinical data (age, PSA, Gleason/Grade Group, T stage). Prognostic model: 5 NRG/ROG phase 3 trials (~5,654 pts, ~11.4 yr f/u); beat NCCN risk groups (AUC gain ~9–15%).

3 • The predictive story — intermediate risk (ROG 9408, n=1,594)

Model-positive (~34%): adding ST-ADT cut 15-yr distant metastasis 14.4% → 4.0% (sHR ~0.34).

Model-negative: no significant ADT benefit — candidates for de-escalation and toxicity avoidance.

Significant treatment-by-biomarker interaction = the hallmark of a true predictive marker.

4 • The template is expanding

LT- vs ST-ADT in high risk (ROG 9202): 14% absolute 15-yr DM gain if positive, none if negative · abiraterone in STAMPEDE (interaction p=0.02) · post-RP BCR model (ROG 9601/0534).

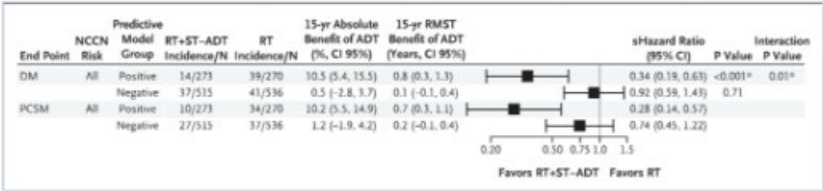
5 • The ambiguity

NCCN: prognostic model at category 2A; predictive biomarker only in a footnote, pending further validation.

All pivotal validations are post-hoc on archived tissue; no completed prospective biomarker-directed RCT; cost-effectiveness unestablished.

ROG 9408: ADT benefit by predictive-model status

Distant metastasis and prostate-cancer mortality



Suggested: Spratt et al., NEJM 2024, Figure 3

RT+ST-ADT vs RT alone by predictive-model status — benefit confined to model-positive patients.

6 • Before local use

Define eligibility · ordering pathway · result interpretation · turnaround · financial workflow.

Prospective trials pending: ASTuTE and four other studies now enrolling.

Raise the evidence threshold when a model changes treatment, not merely prognosis — decision support, not decision maker.

AI-Guided Drug Discovery: The Pipeline Is Real, the Clinical Proof Is Early

Two truths, stated plainly — a large and growing pipeline, and genuinely early clinical evidence

Truth 1 • The pipeline is real

117 AI-enabled assets across 63 companies are in Phase 1–3 interventional trials.
69 (59%) oncology · 82% small molecules.

The distinction that matters

No AI-designed drug has won a Phase 3 oncology outcome trial, and none has reached full regulatory approval in any indication.
Early-phase success does not carry through: ~80–90% Phase 1 success falls to ~40% at Phase 2 — comparable to historic industry averages.

Ecosystem context: a translational bottleneck

AI-oncology publication-to-interventional-trial ratio is ~220:1 (2015–2025); interventional trials are 0.45% of publication volume.
In GU oncology, only 7% of AI/ML trials were therapeutic — and none of the published trials were randomized therapeutic.

Truth 2 • The clinical proof is early

51% completed Phase 1 — only **7%** completed Phase 2.
As of Dec 1, 2025: 60 assets past Phase 1; 8 past Phase 2.

The high-water mark is non-oncologic

Rentosertib (ISM001-055), an AI-designed TNIK inhibitor from Insilico Medicine, reached positive Phase 2a in idiopathic pulmonary fibrosis.
A real milestone for AI-enabled discovery — but not cancer, and not late phase.

Moving target – refresh before delivery

Isomorphic Labs: first-in-human oncology programs discussed publicly with end-of-2026 timing; not yet in peer-reviewed literature.

Trial entry is a real milestone — but value is defined at the outcome. AI is a discovery accelerant still awaiting late-phase proof.

Treatment: automation is already clinically useful

The best near-term wins remove repetitive expert labor



Measure what stays human

- Final contour approval
- Exception handling
- Clinical intent and tradeoffs
- Patient-specific adaptation

Measure what changes

- Planning time
- Inter-observer variability
- Rework and delay
- Safety events

Benchmark before bedside: LLMs in oncology treatment decisions

High concordance is not safe use — anchoring leaves a small, dangerous tail of fluent-but-wrong advice.

5%

Residual unsafe-recommendation rate, best-anchored configuration

26% baseline GPT-4
17% literature-anchored
5% NCCN-anchored

Not benign: contraindicated therapy or missed context, delivered fluently.
Yost et al., ASCO 2026 — 316 cases

Retrieval anchoring works

- 57% → 84% correct with RAG (Ferber, NEJM 2024)
- 0.68 → 0.85 anchored to NCCN (Yost, ASCO 2026)
- Retrieval moves the needle

The right frame: assistant, not decision-maker

- ESMO ELCAP: oversight for all clinical use (2025)
- No unsupervised or unvalidated deployment
- Retrieve, summarize, build the tumor-board packet
- Surface discordance for human adjudication

AI patient–trial matching: enrollment is the bottleneck

Prospective oncology evidence — measure success by enrollment and outcomes, not matching accuracy alone.



The catch

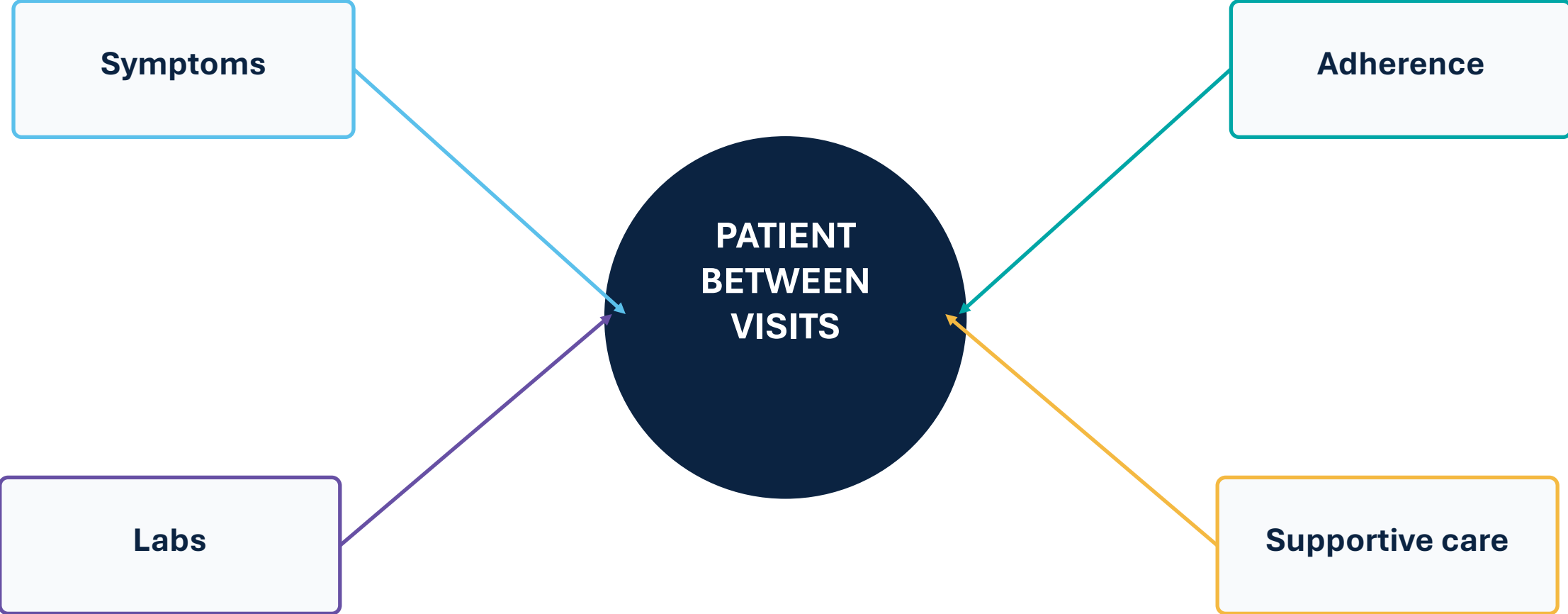
Efficiency gains do not reliably raise enrollment (Villa, Nat Rev Clin Oncol 2026)
Slots, travel, consent and clinician judgment stay rate-limiting
Fluent-but-wrong outputs: missed exclusions, misread lab thresholds
Inappropriate referrals → human-in-the-loop review is mandatory

Governance frame

ESMO ELCAP: validation, transparency, continuous monitoring (Wong, Ann Oncol 2025)
Explicit warning against unsupervised or unvalidated use
TrialGPT: >90% recall, 87.3% criterion accuracy, –42.6% screening time
Assistant, not autonomous decision-maker

Monitoring: AI must extend beyond the clinic visit

Treatment adherence and toxicity management are an operational system.



Owner → threshold → escalation → documentation → outcome

Governance: assign accountable ownership at the use-case level

Vendor monitoring is necessary. Clinical ownership is non-transferable.

MODEL	WORKFLOW	CLINICAL	VALUE
Performance, drift, version	Adoption, burden, exceptions	Safety, benefit, subgroup effects	Capacity, margin, access
Named owner Defined stop rule Review cadence	Named owner Defined stop rule Review cadence	Named owner Defined stop rule Review cadence	Named owner Defined stop rule Review cadence

One dashboard is not enough. Each use case needs a clinical owner, operating owner, technical owner, and executive sponsor.

The test for every oncology AI use case

A six-question screen before pilot, procurement, or scale

1 Which decision or task changes?

2 Who owns the resulting work?

3 What is the comparator?

4 What can fail silently?

5 Which subgroup could be harmed?

6 What evidence unlocks scale?

If the answer is “the algorithm is impressive,” the use case is not ready.

THE TAKEAWAY

Oncology AI will not be won by the best standalone model.

It will be won by the health system that can connect detection, diagnosis, treatment, monitoring, and learning without adding friction or losing accountability.

Start with workflow. Prove value. Scale responsibly.

Jason Hill, MD, MMM, MS

Bibliography: screening and early detection

Consolidated reference list of all articles reviewed, grouped by topic.

AI in cancer screening (imaging & endoscopy)

Gommers J, Hernström V, Josefsson V, et al. Interval cancer, sensitivity and specificity comparing AI-supported mammography screening with standard double reading (MASAI). *Lancet*. 2026;407(10527):505-514.

Eisemann N, Bunk S, Mukama T, et al. Nationwide real-world implementation of AI for cancer detection in population-based mammography screening (PRAIM). *Nat Med*. 2025;31:917-924.

Budzyń K, Romańczyk M, Kitala D, et al. Endoscopist deskillng risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. *Lancet Gastroenterol Hepatol*. 2025;10:896-903.

Ray K. AI-assisted colonoscopy and risk of endoscopist deskillng. *Nat Rev Gastroenterol Hepatol*. 2025.

Multi-cancer early detection (NHS-Galleri & PATHFINDER 2)

Sasieni P, Johnson P, Round T, et al. Effect of screening with a multicancer early-detection test on late-stage cancer diagnosis (NHS-Galleri). *N Engl J Med*. 2026.

Neal RD, Dolly S, Johnson P, et al. Performance of a multi-cancer early detection test in the randomized controlled NHS-Galleri trial. *Nat Med*. 2026.

Nabavizadeh N, McDonnell C, Kurbegov D, et al. Performance and safety of a multi-cancer early detection test: the PATHFINDER 2 study. *Nat Med*. 2026.

Sasieni P, Swanton C, Neal RD. The NHS-Galleri multicancer screening trial: explanation and justification of unique and important design issues. *J Natl Cancer Inst*. 2026;118(4):573-578.

Neal RD, Johnson P, Clarke CA, et al. Cell-free DNA-based multi-cancer early detection test in an asymptomatic screening population (NHS-Galleri): trial design. *Cancers (Basel)*. 2022.

Marinac CR, Rebbeck TR, O'Donnell EK. Some answers, more questions for multi-cancer early detection tests. *Nat Med*. 2026.

Hunter DJ. A step in the search for the elusive holy grail of early detection of cancer. *N Engl J Med*. 2026.

Landon BV, Annapragada AV, Niknafs N, Velculescu VE, Anagnostou V. Liquid biopsies across the cancer care continuum. *Nat Med*. 2025.

Items marked (not retrievable) could not be located or verified; details reproduced only as supplied.

Bibliography: pathology and radiotherapy

Consolidated reference list of all articles reviewed, grouped by topic.

Computational pathology foundation models

Vorontsov E, Bozkurt A, Casson A, et al. A foundation model for clinical-grade computational pathology and rare cancers detection (Virchow). *Nat Med*. 2024;30(10):2924-2935.

Vorontsov E, Shaikovski G, Casson A, et al. End-to-end multimodal pathology foundation model with clinical dialogue (PRISM2). *Nat Med*. 2026;32(9):3203-3213.

Vaidya A, Chen RJ, Williamson DFK, et al. Demographic bias in misdiagnosis by computational pathology models. *Nat Med*. 2024;30:1174-1190.

Lipkova J, Kather JN. The age of foundation models. *Nat Rev Clin Oncol*. 2024.

Chen RJ, Wang JJ, Williamson DFK, et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nat Biomed Eng*. 2023.

AI auto-contouring / auto-segmentation in radiotherapy

Doolan PJ, Charalambous S, Roussakis Y, et al. A clinical evaluation of the performance of five commercial artificial intelligence contouring systems for radiotherapy. *Front Oncol*. 2023.

Young T, Butterworth V, Misson S, et al. Real-world performance evaluation of commercial autocontouring software for head and neck cancer radiotherapy. *Br J Radiol*. 2025;98:1632.

Yuan L, Chen Q, Al-Hallaq H, et al. Quantitative evaluation of AI-based organ segmentation across multiple anatomical sites using eight commercial software platforms. *Pract Radiat Oncol*. 2025.

Podobnik G, Borg C, Debono CJ, Mercieca S, Vrtovec T. Geometric, dosimetric and psychometric evaluation of three commercial AI software solutions for OAR auto-segmentation in head and neck radiotherapy. *Sci Rep*. 2025.

Niu G, Guan Y, Zhang Y, et al. A prospective multicenter trial of deep learning auto-segmentation for organs at risk in thoracic radiotherapy. *Nat Commun*. 2026.

Wang T, et al. *J Appl Clin Med Phys*. 2025;26:e14620. (not retrievable — details as supplied)

Items marked (not retrievable) could not be located or verified; details reproduced only as supplied.

Bibliography: trial matching and decision support

Consolidated reference list of all articles reviewed, grouped by topic.

AI/LLM clinical trial matching & prescreening

Wornow M, Lozano A, Dash D, et al. Zero-shot clinical trial patient matching with LLMs. NEJM AI. 2025;2:Alcs2400360.

Jin Q, Wang Z, Floudas CS, et al. Matching patients to clinical trials with large language models (TrialGPT). Nat Commun. 2024.

Unlu O, Shin J, Mailly CJ, et al. Retrieval-augmented generation-enabled GPT-4 for clinical trial screening (RECTIFIER). NEJM AI. 2024.

Gupta S, Basu A, Nieves M, et al. PRISM: patient records interpretation for semantic clinical trial matching using large language models (OncoLLM). NPJ Digit Med. 2024.

Parikh RB, Kolla L, Beothy EA, et al. Human-AI teaming to improve accuracy and efficiency of eligibility criteria prescreening for oncology trials. Nat Commun. 2026. (NCT06561217)

Villa A, Eadie AL, Synnott D, et al. AI-based augmentation of oncology clinical trials. Nat Rev Clin Oncol. 2026.

Ören Varol T, Varol M. Closing the translational gap: closed-loop AI discovery frameworks in cancer therapeutics. Cancer Med. 2026. (reports the Loaiza-Bonilla prospective neuro-symbolic matching data)

Loaiza-Bonilla A, Thaker N, Chung C, et al. Driving knowledge to action: building a better future with AI-enabled multidisciplinary oncology. ASCO Educ Book. 2025;45(3):e100048.

Loaiza-Bonilla A, et al. ESMO Real World Data Digit Oncol. 2026. (primary article not retrievable — data available only via reference 30)

LLM decision support, tumor boards & governance

Ferber D, Wiest IC, Wölflein G, et al. GPT-4 for information retrieval and comparison of medical oncology guidelines. NEJM AI. 2024.

Ferber D, El Nahhas OSM, Wölflein G, et al. Development and validation of an autonomous artificial intelligence agent for clinical decision-making in oncology. Nat Cancer. 2025.

Hernández-Flores LA, López-Martínez JB, Rosales-de-la-Rosa JJ, et al. Assessment of challenging oncologic cases: ChatGPT, Gemini and a multidisciplinary tumor board. J Surg Oncol. 2025.

Griewing S, Knitza J, Boekhoff J, et al. Evolution of publicly available large language models for complex decision-making in breast cancer care. Arch Gynecol Obstet. 2024.

Wong EYT, Verlingue L, Aldea M, et al. ESMO guidance on the use of large language models in clinical practice (ELCAP). Ann Oncol. 2025.

Patel A, Khand Y, Vallamchetla SK, et al. Fine-tuning, retrieval-augmented generation and hybrid adaptation of language models for clinical decision-making: systematic review. J Med Internet Res. 2026.

Jun H, Tanaka Y, Johri S, et al. A context-augmented large language model for accurate precision oncology medicine recommendations. Cancer Cell. 2026.

Yost C, Palumbo P, Tripathi N, et al. Sources matter: evaluating LLMs for oncology treatment plans when anchored to an oncology-specific database. ASCO 2026 (abstr 1617).

Yost C, Leyfman Y, Tripathi N, et al. Safety-aligned evaluation of large language models for oncology clinical decision support across disease subtypes. 2026.

Items marked (not retrievable) could not be located or verified; details reproduced only as supplied.

Bibliography: decision support and governance

Consolidated reference list of all articles reviewed, grouped by topic.

LLM decision support, tumor boards & governance (cont.)

Kaiser KN, Hughes AJ, Yang AD, et al. Accuracy and consistency of publicly available large language models as clinical decision support tools for colon cancer. *J Surg Oncol*. 2024.

Kaiser KN, Hughes AJ, Yang AD, et al. Use of large language models as clinical decision support tools for pancreatic adenocarcinoma using NCCN guidelines. *Surgery*. 2025.

Yasser M, Barakat G, Awany S, et al. Performance of leading large language models in adhering to clinical guidelines for anaplastic thyroid cancer. *Sci Rep*. 2026.

Harkos C, Hadjigeorgiou AG, Voutouri C, et al. Using mathematical modelling and AI to improve delivery and efficacy of therapies in cancer. *Nat Rev Cancer*. 2025.

Ong JCL, Chang SY, William W, et al. Medical ethics of large language models in medicine. *NEJM AI*. 2024.

Qazi IA, Ali A, Khawaja AU, et al. Automation bias in LLM-assisted diagnostic reasoning among physicians trained in AI literacy — a randomized clinical trial. *NEJM AI*. 2026.

Nenadic I, Loring Z, Noseworthy PA, et al. From technical performance to clinical readiness: a phase-based framework for evidence standards in clinical AI. *NEJM AI*. 2026.

Hao Y, Qiu Z, Holmes J, et al. Large language model integrations in cancer decision-making: a systematic review and meta-analysis. *NPJ Digit Med*. 2025.

LLM decision support, tumor boards & governance (cont.)

Prabhakaran S, Bell S, Lee JC, Kong JCH. The utility of large language models in oncological multidisciplinary team meetings: a systematic review. *Eur J Surg Oncol*. 2026.

Ivashchuk O, Hovornyan S. Prospective multi-centre evaluation of guideline-based artificial intelligence to streamline multidisciplinary tumour board. *Front Oncol*. 2026.

Zoller J, Kalz M, Wu X, et al. Clinical decision support in hematological malignancies using a case-grounded AI agent (HemaGuide). *Nat Med*. 2026.

Elemento O, Khozin S, Sternberg CN. The use of artificial intelligence for cancer therapeutic decision-making. *NEJM AI*. 2025.

Konzmann B, Henkel M, Breit C, Wetterauer C. Applications of large language models in tumor boards: a systematic review. *Oncol Rev*. 2025.

Gould A, Aseem A, Naffakh N, et al. Impact of NCCN-guided training on large language model performance in mNSCLC. 2026.

Zhang SK, Rodman A. Artificial intelligence to revolutionize surgical decision-making is still just around the corner. *Surgery*. 2026.

Wei Y, Tian Y, Liu X, et al. Large language model treatment-pathway outputs in mid- and low rectal cancer: concordance with multidisciplinary team decisions. *Eur J Surg Oncol*. 2026.

Items marked (not retrievable) could not be located or verified; details reproduced only as supplied.

Bibliography: AI-guided drug discovery

Consolidated reference list of all articles reviewed, grouped by topic.

AI-guided drug discovery: pipeline and clinical evidence

Jayatunga MKP, Ayers M, Bruens L, Jayanth D, Meier C. How successful are AI-discovered drugs in clinical trials? A first analysis and emerging lessons. *Drug Discov Today*. 2024;29(6):104009.

Zitnik M, et al. Artificial intelligence in drug discovery and development. *Nat Med*. 2025.

Seyhan AA, Carini C. Translational challenges and the promise of AI-enabled drug discovery in oncology. *Crit Rev Oncol Hematol*. 2026.

Weitzner B, et al. Generative and structure-based approaches to AI-guided small-molecule design in precision oncology. *npj Precis Oncol*. 2026.

Translational bottlenecks, oncology signals and caveats

Dharmasivam M, et al. Artificial intelligence in drug discovery: promise, progress and unresolved limitations. *Pharmacol Rev*. 2025.

Loaiza-Bonilla A, et al. AI-enabled target and compound identification in oncology drug development. *J Clin Oncol*. 2026;44(suppl):abstr 11072.

AI-designed therapeutic candidates entering oncology first-in-human study. *ASCO*. 2026 (abstr e13663).

AI-guided candidate selection in genitourinary oncology. *ASCO Genitourinary Cancers Symposium*. 2026 (abstr 337).

Conference abstracts are preliminary and not peer-reviewed; verify before citing.