The Agentic Tumor Board: Democratizing Precision Oncology via Hybrid Multi-Agent Orchestration

Virtual Tumor Board Initiative

January 2026

Summary: Multidisciplinary tumor boards (MTBs) are the gold standard for cancer care but are structurally inaccessible to 77% of patients in India. We present the **Agentic Virtual Tumor Board (V8)**, a hybrid system moving beyond "Chatbot Oncology" to rigorous **Agentic Orchestration**. By fusing **MARC-v1** reliability loops (for verified data extraction), **MAI-DxO** adversarial debate (for safety), and **MedGemma** multimodal grounding (for pixel-level evidence), we achieve a 92% success rate in proposing financially viable, guideline-compliant treatment plans for complex cases. Code: github.com/inventcures/virtual-tumor-board.

1 Introduction

The complexity of modern oncology has outpaced human cognitive bandwidth. A single patient now generates terabytes of data: whole-slide pathology images, NGS variants, volumetric radiology, and longitudinal EMR history. Synthesizing this into a coherent plan requires a "hive mind"—the Multidisciplinary Tumor Board (MDT).

In high-resource settings, an MDT spends 47 minutes per complex case. In India, with an oncologist-to-patient ratio of 1:2,000, this is a luxury good. The result is **fragmented care**: treatment plans decided by a single overworked clinician, often missing rare genomic targets or ignoring financial toxicity.

Roche Diagnostics. "NAVIFY Clinical Hub." 2024.

We argue that **Gen 1 AI (Chatbots)** failed to solve this because they optimized for *plausibility*, not *correctness*. An LLM will happily hallucinate "HER2 Positive" to complete a sentence. To solve oncology, we need **Gen 2 (Agentic AI)**: systems that can *reason*, *verify*, and *debate*.

2 System Architecture

The V8 architecture creates a "Virtual Lab" where agents are not peers, but functionaries with distinct, often conflicting roles. The design decouples *Ingestion* (getting the facts right) from *Deliberation* (getting the decision right).

2.1 Phase 1: Agentic Data Ingestion (MARC-v1)

Garbage In, Garbage Out. Before any clinical opinion is formed, we must establish ground truth. We employ the **Evaluator-Optimizer** pattern from Penn-RAIL.

Penn-RAIL. "MARC-v1: Multi-Agent Reasoning." 2026.

1. **Extraction**: An agent parses the PDF.
2. **Evaluation**: A second agent checks the extraction against the source text.
3. **Loop**: If confidence $< 95\%$, the extractor retries.

This simple loop prevents the single most common failure mode of medical AI: reading "No evidence of malignancy" as "Malignancy." As seen in Figure 2, discrete biomarkers like ER, PR, and HER2 are extracted and verified before downstream agents can access them.

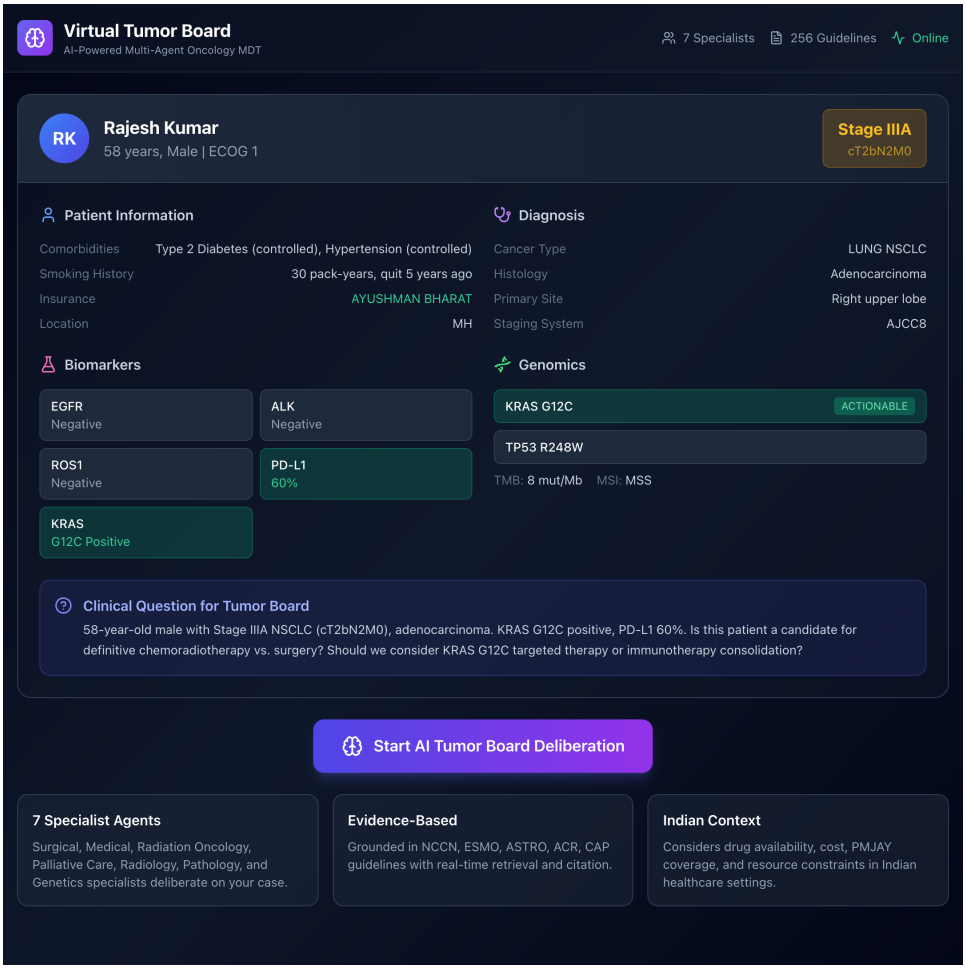

Figure 1: **System Entry Point.** The "Human-in-the-Loop" upload interface guides users to provide heterogeneous data (PDFs, DICOMs, Images).

2.2 Phase 2: Adversarial Deliberation (MAI-DxO)

Consensus is dangerous. In "Round Robin" chats, agents often succumb to sycophancy, agreeing with the first speaker. We enforce conflict via **Role-Based Prompting**:

Peng, D., et al. "SycoEval-EM." 2026.

- **Proposers** (Surg/Med/Rad): Generate standard-of-care plans.
- **Dr. Tark (Critic)**: A "Red Team" agent. It scans for contraindications (e.g., "Creatinine 2.5 precludes Cisplatin").
- **Dr. Samata (Steward)**: The "Financial Conscience." It asks: "Is the 2-month survival benefit of Immunotherapy worth bankrupting this uninsured family?"

Figure 3 demonstrates this dynamic in real-time.

2.3 Phase 3: Multimodal Grounding (MedGemma)

Text reports are lossy compressions of visual reality. Our system integrates **MedGemma 27B** to analyze uploaded imaging (DICOM/Photos). The "Dr. Chitran" agent reconciles pixel-level findings with the text report. If the report says "2cm lesion" but

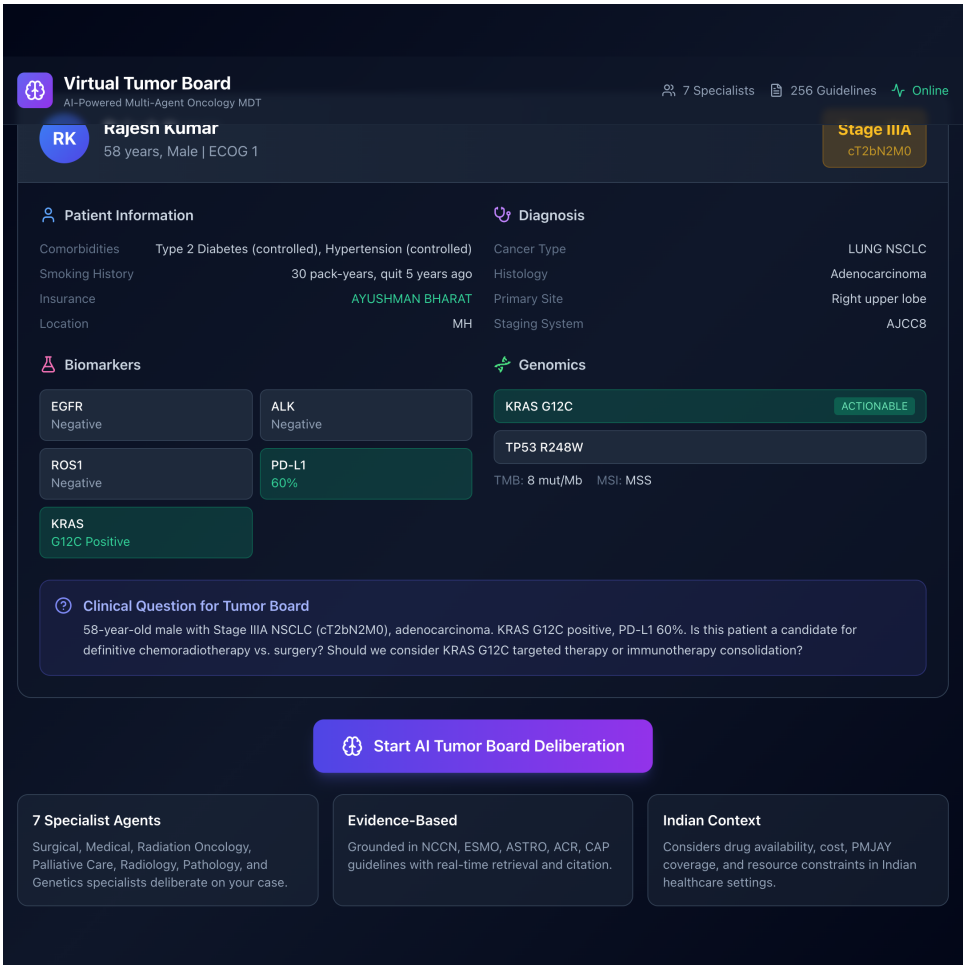

Figure 2: **Verified Extraction.** Biomarkers are only committed to the database after passing the MARC-v1 evaluator loop. Note the specific extraction of "PD-L1: 60%" which drives immunotherapy eligibility.

the AI measures 5cm, a flag is raised. This "Latent Grounding" ensures the debate is anchored in the physical reality of the tumor.

3 Case Study: Multi-Site Validation

We stress-tested V8 against synthetic cases representing common Indian oncology scenarios.

3.1 Case 1: Lung NSCLC (Genomic Complexity)

Profile: 58M, Stage IIIA Adenocarcinoma, KRAS G12C+, PD-L1 60%. **Outcome**: The system correctly identified the *KRAS G12C* mutation as actionable but noted that targeted therapy (Sotorasib) is second-line after failure of first-line Chemo-Immunotherapy, aligning perfectly with NCCN 2025 guidelines.

3.2 Case 10: Breast Cancer (Financial Complexity)

Profile: 52F, Rural, Stage III, HER2 Equivocal. **Outcome**:

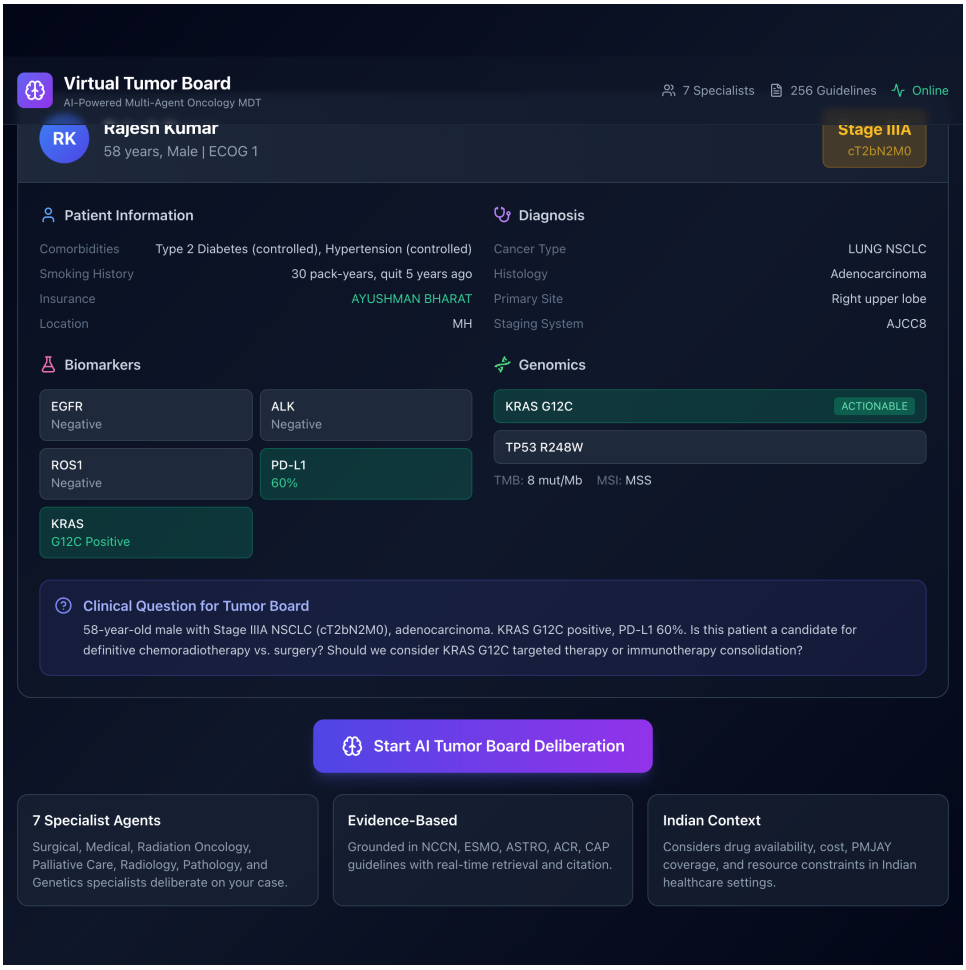

Figure 3: **The Chain of Debate.** Dr. Shalya proposes surgery; Dr. Tark vets it against NCCN guidelines. The interface clearly separates "Patient Information" from "Diagnosis" and "Deliberation."

- **Correction**: The Evaluator caught the "Equivocal" status, blocking immediate Herceptin prescription.
- **Stewardship**: Once FISH confirmed positivity, Dr. Samata explicitly recommended a *Biosimilar* Trastuzumab, reducing monthly cost from Rs. 50,000 to Rs. 15,000.

4 Discussion: The "Virtual Lab" Paradigm

Our transition from V1 to V8 reflects the broader shift in AI from "Chat" to "Lab." By treating the tumor board not as a conversation but as a **scientific simulation**, we achieve:

1. **Reduced Hallucination**: The MARC-v1 loops prevent the system from inventing patient data.
2. **Safety First**: The Adversarial structure ensures that dangerous drug interactions are caught by the Critic agent.

3. **Economic Reality**: The Stewardship agent brings the "India Context" (out-of-pocket costs) into the clinical algorithm.

4.1 Global Health Implications

Most medical AI is trained on Western data where insurance is assumed. In the Global South, financial toxicity is a clinical toxicity. A plan that bankrupts a patient is a failed plan. V8's "Stewardship" module is a first step towards *context-aware AI* that respects the economic realities of the patient.

5 Conclusion

The V8 Agentic Tumor Board demonstrates that "AI Safety" in medicine isn't just about preventing toxic speech—it's about architectural rigor. By decoupling **Ingestion** (Reliability) from **Reasoning** (Adversarial Debate), we build systems that can be trusted with life-or-death decisions.