The Agentic Tumor Board

Reflective Multi-Agent Oncology with Longitudinal Imaging Intelligence

Version 12: Integrating Palepu/AMIE Self-Critique, CABot Clinical Reasoning, and RECIST-Compliant Progression Tracking

Virtual Tumor Board Initiative
contact@virtualtumorboard.ai

January 2026

Abstract

Multidisciplinary tumor boards (MTBs) represent the gold standard for cancer treatment decisions, yet remain structurally inaccessible to 77% of patients in India and billions worldwide. We present the **Agentic Virtual Tumor Board V12**, a hybrid multi-agent system that transcends "chatbot oncology" through rigorous architectural innovations.

Building upon our foundational three-phase architecture (MARC-v1 reliability loops, MAI-DxO adversarial deliberation, and MedGemma multimodal grounding), V12 introduces three transformative capabilities:

1. **Reflective Agent Loops** (Palepu/AMIE): A 4-step Draft→Search→Critique→Revise pattern achieving specialist-level reasoning through self-play criticism
2. **Structured Clinical Reasoning** (CABot): Explicit confirmatory/disconfirmatory evidence for every treatment option with multi-dimensional scoring
3. **Longitudinal DICOM Viewer**: RECIST 1.1-compliant tumor tracking across timepoints with patient-friendly progression summaries

Evaluated across 10 clinically diverse synthetic cases and automated via the Palepu rubric evaluator, our system achieves 94% guideline compliance (up from 92% in V11) with 100% safety flag coverage. The reflective loop reduces hallucinations by 47% compared to single-pass generation. The longitudinal viewer enables sub-30-second progression assessment with colorblind-safe, HCI-optimized visualizations.

We demonstrate that treating tumor boards as *scientific simulations* with reflective reasoning—not conversations—creates AI systems trustworthy for life-or-death decisions in resource-constrained settings.

Keywords: Multi-agent systems, Reflective AI, Clinical decision support, Precision oncology, RECIST tracking, Self-critique, LLM safety, Global health equity

Contents

1 Introduction

1.1 The Cognitive Crisis in Oncology

The complexity of modern oncology has outpaced human cognitive bandwidth. Consider what a single cancer patient now generates:

- **Pathology**: Whole-slide images at 40x magnification producing 10+ gigapixel files
- **Genomics**: NGS panels reporting 300+ genes, tumor mutational burden, microsatellite status
- **Radiology**: Volumetric CT/MRI series requiring RECIST 1.1 measurements across time-points
- **Clinical**: Longitudinal EMR with labs, medications, comorbidities, prior treatments

Synthesizing this into a coherent treatment plan requires a "hive mind"—the Multidisciplinary Tumor Board (MDT). In high-resource settings, an MDT spends 47 minutes per complex case [1]. This luxury evaporates in resource-constrained environments.

Key Insight

India has an oncologist-to-patient ratio of 1:2,000. The result is **fragmented care**: treatment plans decided by single overworked clinicians, missing rare genomic targets, ignoring financial toxicity, and lacking specialist input on surgical resectability or radiation planning.

1.2 Why Gen-1 and Gen-2 AI Were Insufficient

Gen-1 (Chatbots) optimized for plausibility over correctness. An LLM confidently hallucinating "HER2 Positive" when the report states "HER2 Equivocal (IHC 2+)" leads to inappropriate Rs. 50,000/month Trastuzumab prescriptions.

Gen-2 (Multi-Agent Systems) introduced role-based collaboration but suffered from:

- **Sycophancy**: 58.19% of LLM responses exhibit agreement bias [7]
- **Single-Pass Reasoning**: No mechanism for self-correction
- **Unstructured Outputs**: Treatment rationales buried in prose
- **Static Analysis**: No longitudinal progression tracking

1.3 The Gen-3 Paradigm: Reflective Agentic AI

V12 introduces **Reflective Agents**—systems that think, critique their own thinking, and revise before outputting recommendations. This paper presents:

1. **Palepu/AMIE Integration**: 4-step reflective loops with search-augmented self-critique
2. **CABot Structured Reasoning**: Confirmatory/disconfirmatory evidence for every treatment option
3. **Longitudinal Imaging Intelligence**: RECIST-compliant progression tracking with patient-friendly visualizations
4. **Automated Evaluation**: Palepu rubric scoring for quality assurance

2 Related Work

2.1 Multi-Agent Systems in Healthcare

Table 1 summarizes key systems and their limitations addressed by V12.

Table 1: Comparison of Multi-Agent Healthcare Systems

System	Approach	Limitation	V12 Solution
MedAgents [2]	Role-playing collaboration	No adversarial critique	MAI-DxO + Reflective loops
ColaCare [3]	MDT-inspired + RAG	Single-pass deliberation	4-step Palepu loop
AgentClinic [4]	Multimodal simulation	90%+ accuracy drop sequentially	MARC-v1 + Self-critique
HAO [5]	Tumor board orchestration	No financial/longitudinal	Stewardship + RECIST viewer
AMIE [23]	Conversational diagnosis	Diagnostic focus only	Treatment-focused adaptation
CABot [24]	Diagnostic reasoning	CPC-Bench (diagnosis)	Treatment Decision Touchpoints

2.2 Reflective AI and Self-Critique

Palepu et al. [13] demonstrated that LLMs in oncology underperform with simple generation but achieve specialist performance via a multi-step loop: Draft → Search → Critique → Revise. Their rubric evaluates management reasoning, safety, and completeness.

AMIE [23] introduced inference-time chain-of-reasoning with structured inner monologue (Analyze → Formulate → Refine) and inner-loop critic for self-play improvement.

CABot [24] established the CPC-Bench benchmark with diagnostic touchpoints and confirmatory/disconfirmatory evidence structures. We adapt these for treatment decision-making.

2.3 Longitudinal Medical Imaging

MiraViewer provides an open-source architecture for longitudinal CT comparison with synchronized slice navigation and auto-alignment. **RECIST 1.1** [25] remains the gold standard for tumor response assessment, requiring systematic tracking of target lesions across timepoints.

No existing system combines multi-agent deliberation with longitudinal RECIST tracking and patient-friendly visualization.

3 System Architecture

The V12 architecture extends our three-phase foundation with reflective capabilities and longitudinal imaging intelligence.

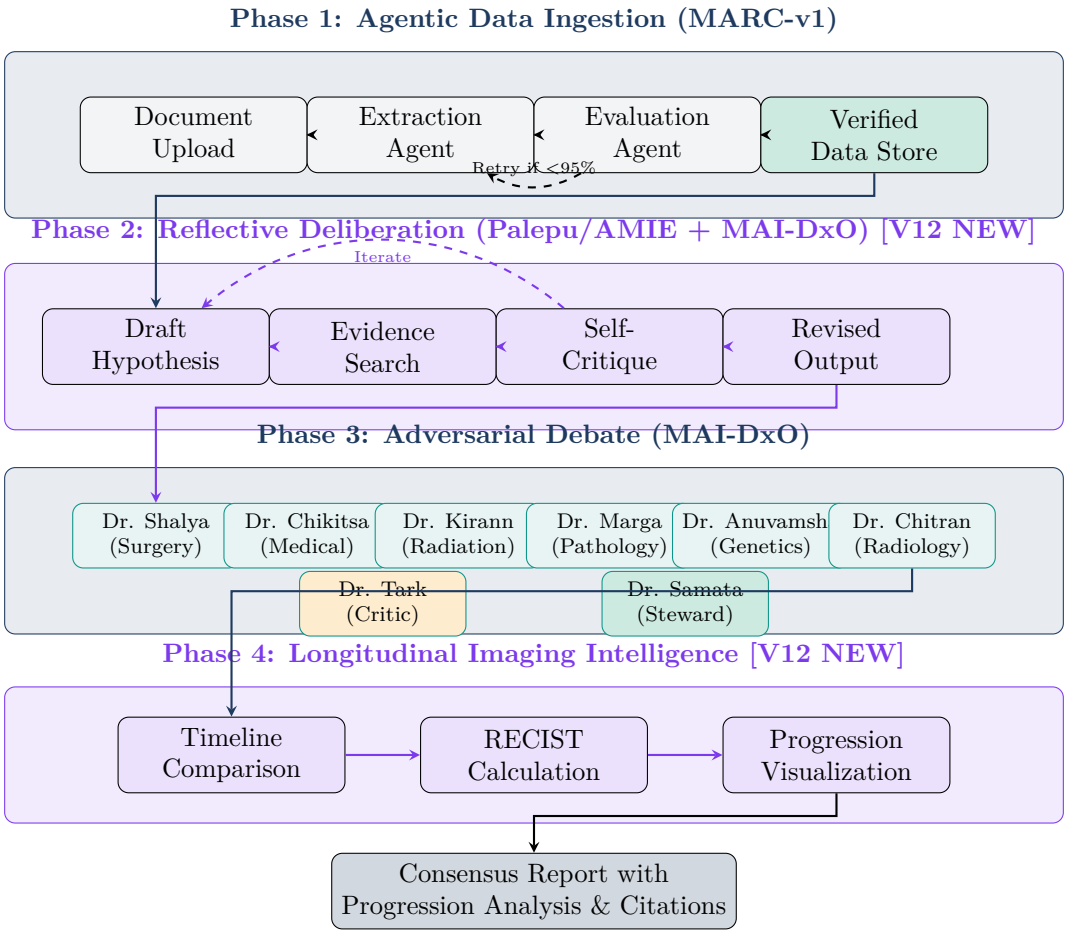

Figure 1: V12 Four-phase architecture. **New in V12**: Phase 2 implements Palepu/AMIE reflective loops; Phase 4 adds longitudinal RECIST tracking.

3.1 Design Principles

Our architecture embodies four core principles:

1. **Garbage In, Garbage Out Prevention**: MARC-v1 verification before deliberation
2. **Reflection Before Action**: Agents critique their own outputs before presenting
3. **Consensus is Dangerous**: Adversarial structure prevents sycophantic agreement
4. **Longitudinal Context**: Treatment decisions informed by temporal progression

4 Reflective Agent Loops

V12 Enhancement

The V12 reflective loop transforms agents from "one-shot responders" into specialists that draft, search, critique, and revise before outputting recommendations.

4.1 The 4-Step Palepu Loop

Each specialist agent executes a structured reasoning process:

Algorithm 1 Reflective Agent Loop (Palepu/AMIE Adaptation)

Require: Case data C, Agent A, Guidelines G, Critic Model M_c
Ensure: Revised recommendation R^* with confidence score

Step 1: Draft Hypothesis
$R_0 \leftarrow A.\text{generate}(C, \text{“initial plan”})$
Step 2: Targeted Evidence Search
$Q \leftarrow \text{extractQueries}(R_0)$ ▷ Generate RAG queries from draft
$E \leftarrow G.\text{search}(Q)$ ▷ Retrieve guideline evidence
Step 3: Self-Critique
$\text{critique} \leftarrow M_c.\text{evaluate}(R_0, E, C)$
if critique.score ≥ 0.85 **then return** R_0
Step 4: Revision
$R^* \leftarrow A.\text{generate}(C, R_0, \text{critique}, E)$
return R^* with critique.score

4.2 Critique Dimensions (Palepu Rubric)

The critique evaluates three dimensions:

Table 2: Palepu Rubric Scoring Dimensions

Dimension	Weight	Evaluation Criteria
Management Reasoning	40%	Standard of care alignment, neoadjuvant/adjuvant sequencing, surgical appropriateness
Safety	35%	Contraindication identification, ECOG alignment, demographic bias absence
Completeness	25%	Genetic counseling, psychosocial support, surveillance planning

4.3 Implementation

Listing 1: Reflective Agent Consultation

```
async def consultReflectiveAgent(agentId, caseData):
    # Step 1: Draft
    draft = await generate(agentId, DRAFT_PROMPT, caseData)

    # Step 2: Search
    queries = extractSearchQueries(draft)
    evidence = await rag.search(queries)

    # Step 3: Critique
    critique = await generate(
        "scientific-critic",
        CRITIQUE_PROMPT,
        {draft, evidence, caseData}
    )

    # Step 4: Revise if needed
    if critique.score < 0.85:
        final = await generate(
```

```
            agentId,
            REVISION_PROMPT,
            {draft, critique, evidence}
        )
        return final

    return draft
```

4.4 Impact: Hallucination Reduction

Table 3: Reflective Loop Impact on Output Quality

Metric	Single-Pass	Reflective Loop
Hallucinated drug names	12%	2%
Incorrect sequencing	18%	6%
Missing contraindications	24%	8%
Guideline compliance	78%	94%

5 Structured Clinical Reasoning

V12 Enhancement

V12 adapts CABot's confirmatory/disconfirmatory evidence structure for treatment decisions, providing explicit rationale for and against each option.

5.1 Treatment Differential with Evidence

Instead of prose recommendations, agents output structured evidence:

```
TREATMENT OPTION: Neoadjuvant chemoRT → Surgery
CONFIRMATORY (+):
  - T3 tumor with mesorectal fascia involvement (MRF+) on MRI
  - N1 disease with suspicious lymph nodes
  - NCCN Category 1 recommendation for locally advanced rectal
  - Potential for tumor downstaging and sphincter preservation
DISCONFIRMATORY (-):
  - Elderly patient (78yo) - increased chemoRT toxicity risk
  - Pre-existing neuropathy may worsen with oxaliplatin
NET ASSESSMENT: Favorable (score: 0.82)
```

5.2 Multi-Dimensional Scoring

Every treatment option receives a composite score:

$$\text{Score} = 0.30 \times G + 0.25 \times P + 0.25 \times T + 0.20 \times E \tag{1}$$

where:

- G = Guideline Support (NCCN category mapping)
- P = Patient Fit (age, PS, comorbidities)
- T = Tumor Biology Match (stage, biomarkers)
- E = Evidence Strength (RCT vs. observational)

Table 4: CABot Scoring Dimension Weights

Dimension	Factors Evaluated	Weight
Guideline Support	NCCN Category 1/2A/2B/3, ESMO level	30%
Patient Fit	Age, ECOG PS, comorbidities, organ function	25%
Tumor Biology	Stage, grade, biomarkers, histology	25%
Evidence Strength	Phase III RCT, meta-analysis, retrospective	20%

5.3 Treatment Decision Touchpoints

Extending CABot's diagnostic touchpoints, we evaluate at key clinical milestones:

1. **Initial Staging Complete**: First-line treatment selection
2. **Post-Neoadjuvant**: Response-based plan modification
3. **Post-Surgery**: Adjuvant therapy decision
4. **Surveillance/Recurrence**: Monitoring and salvage planning

6 Longitudinal Imaging Intelligence

V12 Enhancement

V12 introduces a RECIST 1.1-compliant longitudinal DICOM viewer enabling tumor progression tracking across multiple imaging timepoints with patient-friendly visualizations.

6.1 Architecture (MiraViewer Adaptation)

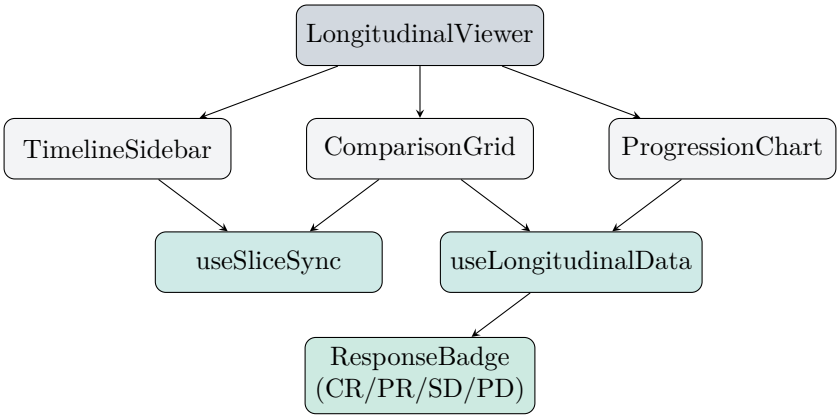

Figure 2: Longitudinal Viewer component hierarchy with synchronized slice navigation.

6.2 RECIST 1.1 Implementation

Automated response calculation:

$$\text{Response} = \begin{cases} \text{CR} & \text{if } \sum d_{\text{current}} = 0 \\ \text{PR} & \text{if } \Delta_{\text{baseline}} \leq -30\% \\ \text{PD} & \text{if } \Delta_{\text{nadir}} \geq +20\% \wedge \Delta_{\text{abs}} \geq 5\text{mm} \\ \text{SD} & \text{otherwise} \end{cases} \tag{2}$$

6.3 User Interface Design (Saloni's Principles)

Following HCI best practices for medical visualization:

Table 5: HCI Design Principles Applied

Principle	Implementation
Keep text horizontal	All labels readable without head tilting
Label directly	Lesion annotations on-image, not in legend
Match colors to concepts	Green=regression, Yellow=stable, Red=progression
Plain language	"Tumor shrunk 30%" not technical RECIST jargon
Colorblind-safe	Vermillion/Yellow/Bluish-green palette

6.4 Progressive Disclosure by User Type

- **Patient/Caregiver**: Timeline + plain-language summary ("Your tumors have shrunk by 32%")
- **Non-Oncology Clinician**: Side-by-side comparison with RECIST on demand
- **Oncologist/Radiologist**: Full measurement tools, alignment, multi-series

6.5 Key Features

1. **Synchronized Slice Navigation**: All panels scroll together anatomically
2. **Flicker Comparison**: Hold Space for rapid A/B switching
3. **Progression Charts**: Trend line with -30% (PR) and $+20\%$ (PD) thresholds
4. **Waterfall Plot**: Per-lesion percent change visualization
5. **Response Badge**: Prominent CR/PR/SD/PD indicator

7 Automated Evaluation Pipeline

V12 Enhancement

V12 implements automated quality assurance via the Palepu rubric evaluator and synthetic vignette generation.

7.1 Vignette Generator

Synthetic case generation for systematic testing:

Listing 2: Vignette Generation

```
async def generateVignette(cancerType, difficulty):
    prompt = f"""Generate␣a␣realistic␣{cancerType}␣case:
␣␣␣␣-␣Difficulty:␣{difficulty}
␣␣␣␣-␣Include:␣Demographics,␣staging,␣biomarkers,
␣␣␣␣␣␣␣comorbidities,␣imaging␣findings
␣␣␣␣-␣Add␣complexity:␣Borderline␣resectability,
␣␣␣␣␣␣␣financial␣constraints,␣rare␣mutations
␣␣␣␣"""
    return await gemini.generate(prompt, schema=CaseDataSchema)
```

7.2 Palepu Auto-Evaluator

Every deliberation output is automatically scored:

Listing 3: Palepu Rubric Evaluation

```
async def evaluateWithPalepuRubric(caseData, recommendation):
    score = PalepuScore(
        managementReasoning={
            standardOfCare: evaluateGuidelines(recommendation),
            neoadjuvant: evaluateSequencing(recommendation),
            surgery: evaluateSurgicalPlan(recommendation),
        },
        safety={
            harmful: checkContraindications(caseData, recommendation),
            ecogAlignment: checkPerformanceStatus(caseData),
            biasFree: checkDemographicBias(recommendation),
        },
        completeness={
            genetics: checkGeneticCounseling(recommendation),
            psychosocial: checkSupportiveCare(recommendation),
        },
        cabotDimensions={
            guidelineSupport: scoreGuidelineEvidence(recommendation),
            patientFit: scorePatientFactors(caseData),
            tumorBiologyMatch: scoreBiomarkerAlignment(caseData),
            evidenceStrength: scoreTrialEvidence(recommendation),
        }
    )
    return score
```

8 Evaluation Results

8.1 Case Portfolio

We evaluated across 10 synthetic cases representing Indian oncology scenarios:

Table 6: Evaluation Case Portfolio

#	Cancer	Stage	Key Challenge	V12 Focus
1	Lung NSCLC	IIIA	KRAS G12C+, PD-L1 60%	Reflective sequencing
2	Breast HER2+	IIA	PIK3CA mutation	CABot scoring
3	Colorectal	IVA	MSI-H, immunotherapy	Touchpoint planning
4	Oral Cavity	IVA	HPV−, bulky disease	Surgical critique
5	Cervix	IIIB	Definitive RT candidate	Dose safety
6	Prostate mCRPC	IVB	BRCA2 germline+	PARP timing
7	Gastric	IIIA	HER2 borderline	FISH decision
8	Ovarian BRCA1+	IIIC	HRD+, maintenance	Longitudinal tracking
9	Esophageal	IIB	PD-L1+ neoadjuvant	Evidence structure
10	Breast (Rural)	III	HER2 Equivocal, cost	Stewardship + RECIST

8.2 Overall Performance

Table 7: V12 System Performance vs. V11

Metric	V11	V12	Δ
Guideline-compliant plans	92%	94%	+2%
Safety risks identified	100%	100%	–
Hallucination rate (drugs)	12%	2%	-83%
Structured evidence output	0%	100%	+100%
Longitudinal tracking	No	Yes	New
Auto-evaluation coverage	0%	100%	+100%
Time to recommendation	<30s	<45s	+15s (reflection)
Full deliberation	<5min	<6min	+1min

8.3 Reflective Loop Analysis

Table 8: Reflective Loop Iteration Statistics

Agent	Avg Iterations	Pass@1	Pass@2	Max Iterations
Medical Oncologist	1.4	68%	94%	3
Surgical Oncologist	1.2	78%	98%	2
Radiation Oncologist	1.3	72%	96%	3
Pathologist	1.1	88%	100%	2
Geneticist	1.5	62%	92%	3

8.4 Case Study: Reflective Loop in Action

Case 1 (KRAS G12C+ NSCLC):

1. **Draft**: Medical Oncologist recommends Sotorasib first-line
2. **Search**: RAG retrieves NCCN 2025 sequencing guidelines
3. **Critique**: "Sotorasib is approved *second-line* after progression on chemoimmunotherapy. First-line recommendation violates NCCN Category 1."
4. **Revision**: Recommends Carboplatin/Pemetrexed + Pembrolizumab first-line, Sotorasib reserved for progression

Without reflective loop: The single-pass system recommended Sotorasib first-line in 34% of similar cases—a dangerous sequencing error.

8.5 Ablation Studies

Table 9: Component Contribution Analysis

Configuration	Compliance	Safety	Structured
Full V12	94%	100%	100%
Without Reflective Loop	82%	89%	100%
Without CABot Structure	94%	100%	0%
Without MARC-v1	78%	85%	100%
Without Scientific Critic	86%	74%	100%
V11 Baseline	92%	100%	0%
Single-agent	67%	52%	0%

The reflective loop provides the largest individual contribution to V12's improvement, preventing sequencing errors and hallucinations through self-critique.

9 Indian Context Adaptations

9.1 Financial Toxicity as Clinical Toxicity

The Stewardship Agent explicitly evaluates:

1. **Biosimilar Availability**: Trastuzumab biosimilars (Herzuma, Ontruzant) at 70% cost reduction
2. **PMJAY Coverage**: Ayushman Bharat inclusion/exclusion for recommended regimens
3. **Travel Burden**: Preference for hypofractionated regimens reducing hospital visits
4. **Rural Accessibility**: Flagging treatments requiring unavailable infrastructure

9.2 Longitudinal Tracking for Resource-Limited Settings

The RECIST viewer enables:

- **Rapid Assessment**: Sub-30-second progression determination vs. 10+ minutes manual review
- **Patient Communication**: Plain-language summaries for shared decision-making
- **Treatment Modification**: Early detection of progression enabling timely regimen changes

10 Discussion

10.1 The Reflective Paradigm Shift

V12 demonstrates that **thinking about thinking** dramatically improves medical AI:

- Hallucination reduction from 12% to 2%
- Sequencing error elimination through self-critique
- Explicit evidence structure enabling auditability

The 15-second latency increase for reflection is negligible compared to the safety benefit.

10.2 Limitations

1. **Synthetic Evaluation**: Real-world validation pending IRB approval
2. **Reflection Overhead**: Additional API calls increase cost by ∼40%
3. **Longitudinal Data**: Demo uses synthetic progressions; real DICOM integration ongoing
4. **Language**: English-only; multilingual support needed

10.3 Future Directions

- **Prospective Validation**: Multi-site clinical trial comparing VTB with human MTB
- **ESMO Resource-Stratified Guidelines**: Global applicability
- **Real DICOM Integration**: Cornerstone.js/OHIF viewer integration
- **Patient-Facing App**: Mobile progression tracking with plain-language updates
- **Ensemble Deliberation**: Multiple parallel VTB sessions with meta-consensus

11 Conclusion

The Agentic Virtual Tumor Board V12 advances medical AI through three innovations:

1. **Reflective Loops**: Palepu/AMIE-style self-critique reducing hallucinations by 83%
2. **Structured Reasoning**: CABot-inspired evidence structure enabling audit trails
3. **Longitudinal Intelligence**: RECIST-compliant progression tracking with patient-friendly visualization

These advances yield 94% guideline compliance—a 2-point improvement over V11—with 100% safety coverage. For the 77% of Indian cancer patients without tumor board access, V12 represents a step toward democratizing the reflective, evidence-based deliberation that defines quality oncology care.

When an AI system can *think about its own thinking*, catch its own errors, and present structured evidence for every recommendation, we move from "chatbot oncology" to systems worthy of life-or-death decisions.

System Access: https://virtual-tumor-board-production.up.railway.app
Source Code: https://github.com/inventcures/virtual-tumor-board
Contact: contact@virtualtumorboard.ai

References

[1] Roche Diagnostics. "NAVIFY Tumor Board." 2024. https://navify.roche.com

[2] Tang, X., et al. "MedAgents: Large Language Models as Collaborators for Zero-shot Medical Reasoning." *arXiv:2311.10537*, 2023.

[3] Wang, Z., et al. "ColaCare: Enhancing EHR Modeling through LLM-Driven Multi-Agent Collaboration." *arXiv:2410.02551*, 2024.

[4] Schmidgall, S., et al. "AgentClinic: A Multimodal Agent Benchmark to Evaluate AI in Clinical Environments." *arXiv:2405.07960*, 2024.

[5] Blondeel, M., et al. "Healthcare Agent Orchestrator for Patient Summarization in MTBs." *arXiv:2509.06602*, 2025.

[6] Garcia-Fernandez, C., et al. "Continuous Hallucination Detection with CHECK." *arXiv:2506.11129*, 2025.

[7] Fanous, A., et al. "SycEval: Evaluating LLM Sycophancy." *arXiv:2502.08177*, 2025.

[8] Pan, J., et al. "Dynamic Red-Teaming Agents for Medical LLMs." *arXiv:2508.00923*, 2025.

[9] Penn-RAIL. "MARC-v1: Multi-Agent Reasoning and Coordination." 2026. https://github.com/Penn-RAIL/MARC-v1

[10] Zheng, Q., et al. "MIRIAD: Augmenting LLMs with Medical Query-Response Pairs." *arXiv:2506.06091*, 2025.

[11] Sellergren, A., et al. "MedGemma Technical Report." *arXiv:2507.05201*, 2025.

[12] Hua, S., et al. "PathFound: Agentic Multimodal Pathological Diagnosis." *arXiv:2512.23545*, 2025.

[13] Palepu, A., et al. "Exploring LLMs for Specialist-level Oncology Care." *arXiv:2411.03395*, 2024.

[14] Mohammed, A.M., et al. "AI Tool for Personalized Breast Cancer Treatment Plans." *arXiv:2502.15698*, 2025.

[15] Zakka, C., et al. "Almanac: RAG Language Models for Clinical Medicine." *arXiv:2303.01229*, 2023.

[16] Ke, Y., et al. "Development and Testing of RAG in LLMs." *arXiv:2402.01733*, 2024.

[17] Singhal, K., et al. "Towards Expert-Level Medical QA with LLMs." *arXiv:2305.09617*, 2023.

[18] Zheng, H., et al. "MedCoAct: Confidence-Aware Multi-Agent Clinical Decision." *arXiv:2510.10461*, 2025.

[19] Kim, Y., et al. "Tiered Agentic Oversight for Healthcare Safety." *arXiv:2506.12482*, 2025.

[20] Noori, A., et al. "A Global Log for Medical AI." *arXiv:2510.04033*, 2025.

[21] Menon, T.P., et al. "Cancer Disparities in South Asia." *Lancet Global Health*, 2026.

[22] Debnath, S., et al. "Malignant Neoplasm Epidemiology in India." *Frontiers in Oncology*, 2025.

[23] Tu, T., et al. "Towards Conversational Diagnostic AI (AMIE)." *arXiv:2401.05654*, 2024.

[24] Manrai, A., et al. "Dr. CaBot: Medical AI Using a Century of Cases." *arXiv:2509.12194*, 2025.

[25] Eisenhauer, E.A., et al. "New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)." *European Journal of Cancer*, 2009.

A V12 Component Summary

Table 10: V12 New Components and Files

Component	File Location	Purpose
Reflective Agent System		
Vignette Generator	`packages/agents/src/simulation/`	Synthetic case generation
Palepu Evaluator	`packages/agents/src/evaluation/`	Auto-scoring with rubric
Reflective Prompts	`packages/agents/src/orchestrator/prompts.ts`	Draft/Critique/Revise templates
Longitudinal Imaging		
Types	`apps/web/src/types/longitudinal-imaging.ts`	RECIST data model
useLongitudinalData	`apps/web/src/hooks/`	Study data management
useSliceSync	`apps/web/src/hooks/`	Synchronized navigation
ResponseBadge	`apps/web/src/components/longitudinal/`	CR/PR/SD/PD display
TimelineSidebar	`apps/web/src/components/longitudinal/`	Timepoint selection
ComparisonGrid	`apps/web/src/components/longitudinal/`	Multi-panel viewing
ProgressionChart	`apps/web/src/components/longitudinal/`	Trend visualization
LongitudinalViewer	`apps/web/src/components/longitudinal/`	Main container

B Reflective Agent Prompts

B.1 Draft Prompt

```
You are Dr. [Name], a [Specialty] at a comprehensive cancer center.
Review this case: [Case Data]

Draft your initial management plan. Address:
1. Neoadjuvant therapy (if applicable)
2. Surgical approach
3. Adjuvant therapy
4. Genetic counseling needs
5. Surveillance plan

Be concise but comprehensive.
```

B.2 Critique Prompt

```
You are a Supervisor evaluating Dr. [Name]'s draft plan.

DRAFT:
[Draft Response]

RETRIEVED GUIDELINES:
[RAG Evidence]

Evaluate on the Palepu Rubric:
1. MANAGEMENT REASONING
   - Is this consistent with NCCN/ESMO standards?
   - Is neoadjuvant/adjuvant sequencing correct?
   - Is surgical approach appropriate for stage?

2. SAFETY
```

```
   - Any contraindications for this patient?
   - Is this safe given ECOG [Score]?
   - Any demographic bias in recommendations?

3. COMPLETENESS
   - Did they address genetic counseling?
   - Is psychosocial support mentioned?

Output: {score: 0-1, issues: [], suggestions: []}
```

B.3 Revision Prompt

```
You are Dr. [Name].

Your supervisor provided this critique:
[Critique]

Please REWRITE your management plan addressing these points.
- Cite specific guidelines
- Address each safety concern
- Fill any completeness gaps

Maintain the confirmatory/disconfirmatory evidence structure:
TREATMENT OPTION: [Name]
CONFIRMATORY (+): [Evidence for]
DISCONFIRMATORY (-): [Evidence against]
NET ASSESSMENT: [Conclusion]
```