

AI Agents vs. Agentic AI: AI Agent in Oncology

Nat Tangchitnob
as Commentator

The paper

Article | [Open access](#) | Published: 06 June 2025

Development and validation of an autonomous artificial intelligence agent for clinical decision-making in oncology

[Dyke Ferber](#), [Omar S. M. El Nahhas](#), [Georg Wölflein](#), [Isabella C. Wiest](#), [Jan Clusmann](#), [Marie-Elisabeth Leßmann](#), [Sebastian Foersch](#), [Jacqueline Lammert](#), [Maximilian Tschochohei](#), [Dirk Jäger](#), [Manuel Salto-Tellez](#), [Nikolaus Schultz](#), [Daniel Truhn](#) & [Jakob Nikolas Kather](#) 

[Nature Cancer](#) (2025) | [Cite this article](#)

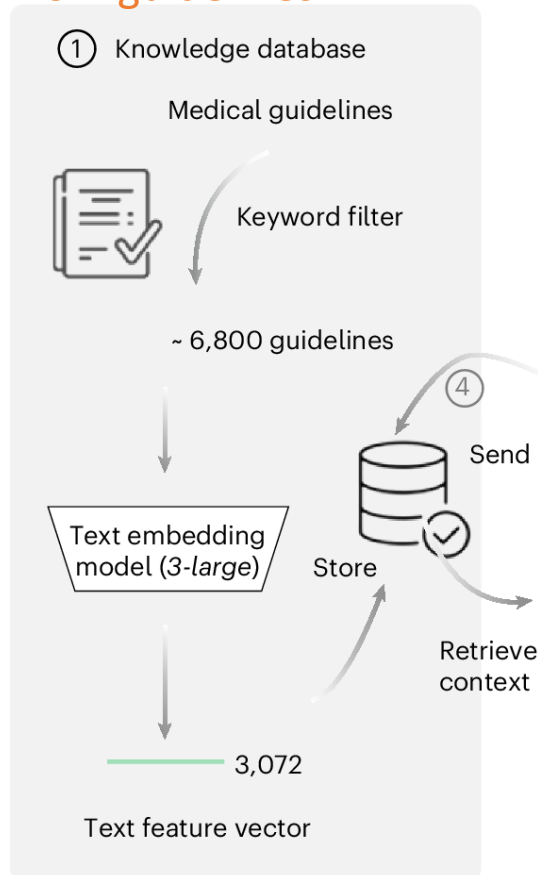
19k Accesses | **232** Altmetric | [Metrics](#)

Objectives

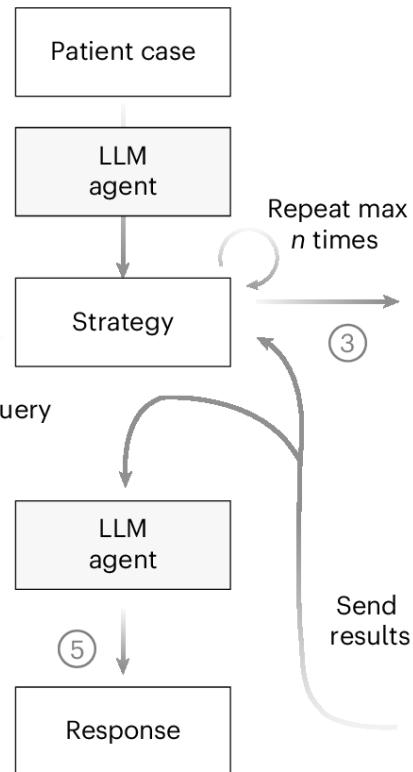
- Build and evaluate an AI agent tailored to interact with and draw conclusions from multimodal patient data through tools in oncology
- The tools includes:
 1. Vision model API to **generate radiology reports from MRI and CT**
 2. MedSAM: medical **image segmentation** model
 3. An in-house vision transformer models trained to detect genetic alterations from **histopathology** slides.
 4. Basic calculator
 5. **Web searching** tool for Google, PubMed, OncoKB (precision oncology database)
 6. Knowledge database: 6,800 **oncology guidelines**

High-level overview of LLM agent pipeline

2. Oncology knowledge from guidelines



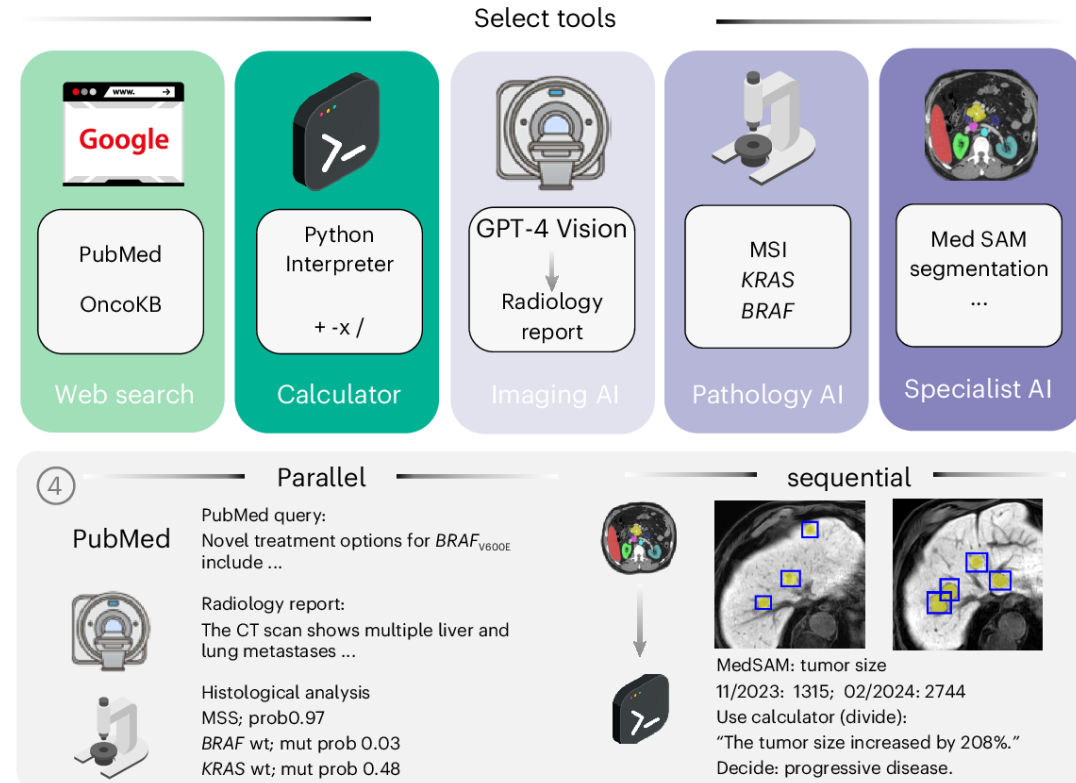
1. Main pipeline



3. Callable tools

Agent pipeline

② General purpose and domain-specific medical tools



Example case: Patient A

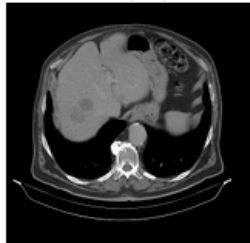
Case Description

Mrs. Adams (DOB: 03/05/1969, ID: A-CC-38), a 55-year-old female, has been battling left-sided colorectal cancer for an extended period. Four years after her initial diagnosis, in November 2023, a relapse was detected, characterized by multiple metastases to the liver, as evidenced by the imaging study dated 11/11/2023 (CTA_11_11_2023.png). Subsequently, she commenced a chemotherapy regimen comprising FOLFOX and bevacizumab at an external hospital in Berlin. The latest imaging, captured on 02/24/2024 (CTA_02_24_2024.png), is stored in the patient's folder within the Images path of the system. To date, there are no additional metastatic sites identified beyond the liver. Mrs. Adams has generally tolerated the chemotherapy well. However, she has recently reported an increase in blood pressure, with readings consistently in the range of 170-190 mmHg. Apart from elevated blood pressure, she has not experienced any other symptomatic manifestations of her condition.

Doctor: How did the disease develop? What are the next steps in her cancer treatment plan?

Appendix¹:

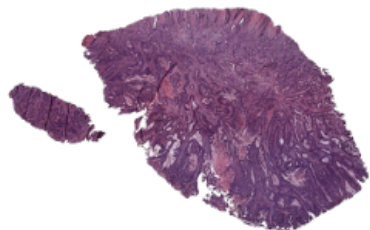
CT-A, 11/11/2023



CT-A, 02/24/2024



Histology



Model response: Strategy, Cited Response, Suggestion

Given Mrs. Adams' stable liver metastases as evidenced by the comparison of CTA_11_11_2023.png and CTA_02_24_2024.png [Tool], alongside her genetic profile indicating MSS, and wild-type status for both BRAF and KRAS [Tool], her disease appears to be in a controlled state under the current regimen of FOLFOX and bevacizumab. The stability of liver lesions suggests that the chemotherapy, in conjunction with bevacizumab, is effective in managing her metastatic colorectal cancer (mCRC) [Tool].

The development of hypertension, a known side effect of bevacizumab [11][12][13], necessitates a multidisciplinary approach to manage this condition without compromising her cancer treatment efficacy. Given the stable disease and the absence of BRAF and KRAS mutations [Tool], Mrs. Adams remains a candidate for continued anti-EGFR therapy, considering her left-sided mCRC, which is typically more responsive to such treatments [2][8].

Next Steps in Treatment Plan:

- 1. Management of Hypertension:** Immediate consultation with a cardiologist is recommended to manage her newly developed hypertension, potentially related to bevacizumab [22][23][26][27]. Adjustments to her antihypertensive therapy may be necessary, with continuous monitoring of blood pressure to prevent cardiovascular complications [68][69][70][71].
- 2. Ongoing Cancer Treatment:** Given the stable disease, continuation of the current regimen is advisable. However, the patient's tolerance, especially concerning the development of hypertension, should be closely monitored [1][4]. If hypertension cannot be controlled, a dose adjustment or a switch to an alternative antiangiogenic agent with a potentially different side effect profile, such as aflibercept or ramucirumab, could be considered [9][86].
- 3. Re-evaluation for Surgical Intervention:** Despite the stable metastatic disease, re-evaluation for the possibility of liver metastasectomy or ablation should be considered, as these could offer a chance for prolonged survival or even cure in selected cases [33][34][35].

Example case: Patient A

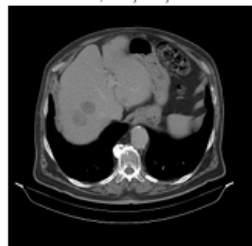
Case Description

Mrs. Adams (DOB: 03/05/1969, ID: A-CC-38), a 55-year-old female, has been battling left-sided colorectal cancer for an extended period. Four years after her initial diagnosis, in November 2023, a relapse was detected, characterized by multiple metastases to the liver, as evidenced by the imaging study dated 11/11/2023 (CTA_11_11_2023.png). Subsequently, she commenced a chemotherapy regimen comprising FOLFOX and bevacizumab at an external hospital in Berlin. The latest imaging, captured on 02/24/2024 (CTA_02_24_2024.png), is stored in the patient's folder within the Images path of the system. To date, there are no additional metastatic sites identified beyond the liver. Mrs. Adams has generally tolerated the chemotherapy well. However, she has recently reported an increase in blood pressure, with readings consistently in the range of 170-190 mmHg. Apart from elevated blood pressure, she has not experienced any other symptomatic manifestations of her condition.

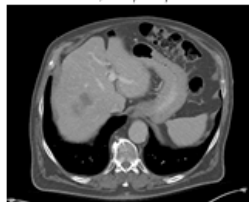
Doctor: How did the disease develop? What are the next steps in her cancer treatment plan?

Appendix¹:

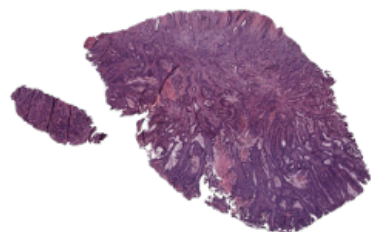
CT-A, 11/11/2023



CT-A, 02/24/2024



Histology



Model response (cont.)

4. **Maintenance Therapy:** If surgery is not feasible and once a maximum response to the current regimen is achieved, transitioning to maintenance therapy with a fluoropyrimidine (e.g., capecitabine) with or without bevacizumab could be considered to reduce toxicity while maintaining disease control [4][85].
5. **Genetic and Molecular Marker Re-evaluation:** Although initial genetic predictions indicate MSS, and wild-type BRAF and KRAS, a broader panel re-evaluation might uncover other actionable mutations or markers that could open up additional targeted therapy options [33][42][46].
6. **Clinical Trials and Novel Therapies:** Exploration of clinical trials focusing on novel therapies or combinations that could be beneficial for her specific genetic and molecular profile is recommended [6][33].
7. **Palliative Care Integration:** Early integration of palliative care to manage symptoms, improve quality of life, and support Mrs. Adams and her family through treatment decisions is crucial [126][127][129].
8. **Regular Monitoring:** Continuous assessment of her disease with imaging studies and tumor markers is essential to evaluate the response to treatment and adjust the treatment plan as necessary [38][80].

This comprehensive approach, tailored to Mrs. Adams' specific medical history, current condition, and potential treatment pathways, aims to optimize her quality of life while effectively managing her mCRC. Regular reassessment of her condition and treatment tolerance will guide any necessary adjustments to her treatment plan. To further refine Mrs. Adams' treatment plan, it would be helpful to know the specifics of her blood pressure readings over time, to assess the severity and progression of her hypertension. Utilizing my array of resources, I can further assist by identifying optimal blood pressure management strategies that align with her ongoing cancer treatment. Additionally, detailed information regarding her response to previous treatments and any side effects experienced would be invaluable. This would enable a more personalized approach to considering the integration of anti-EGFR therapy, ensuring it complements her current regimen while minimizing potential adverse effects.

Compare to standard LLM

a Example responses from GPT-4 alone in comparison to GPT-4 with tool use and RAG

GPT-4

As an AI developed by OpenAI, I'm unable to access or analyze real patient data, images, or external databases directly, nor can I use tools for precise measurement of lesions described in radiology reports. Therefore, I cannot provide real-time analysis or visual assessment of Mrs. Chen's radiology images (February 3 2024.png and June 25 2024.png) or perform precise measurements of the lesions. However, based on the provided information, I can offer a hypothetical and educational response to the query regarding Mrs. Chen's current state of disease and a detailed treatment plan.

Given Mrs. Zhang's disease progression despite treatment with both FOLFOX + Panitumumab and FOLFIRI + Bevacizumab, and considering her tumor is KRAS and BRAF wild-type, the next step in her treatment should be guided by the latest clinical guidelines and the molecular profile of her tumor. The pathology report from the recent biopsy, once available, may provide additional insights into potential targeted therapies based on any new mutations or biomarkers identified.

Current State of Disease: Based on the most recent follow-up and imaging data available up to March 2024, Mrs. Patel shows no evidence of disease (NED). This status is after a successful RO resection of a stage IIIb colorectal adenocarcinoma and subsequent adjuvant chemotherapy. Her current health condition is reported to be good, with no relevant health issues noted.

GPT-4



Tools



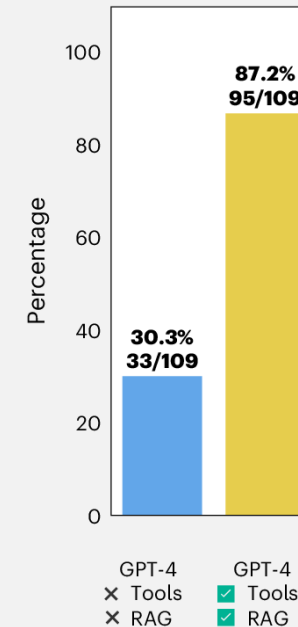
RAG

Based on the provided data and tool results, Mrs. Chen's metastatic colorectal cancer has demonstrated progression from February to June 2024. The significant increase in the size of lung metastases, as quantified by the segmentation tool results showing an increase in the overall area of the lesions from 702 to 2798 square units, indicates that the disease is not adequately controlled under the current regimen of FOLFIRI and bevacizumab [Tool]. Furthermore, the genetic mutation analysis revealing MSS status and the absence of BRAF and KRAS mutations provides valuable information for guiding future treatment decisions [Tool]. Given the progression of disease despite current systemic therapy, and considering Mrs. Chen's stable ECOG performance status of 1, which suggests she remains in relatively good condition to tolerate treatment, the following treatment plan is proposed: ...

Based on the comprehensive analysis of Mrs. Zhang's medical history, radiological data, and genetic testing, we have a clear picture of her current oncologic status and can make informed recommendations for her ongoing treatment strategy. (...)
-The quantitative analysis of liver metastases from CT scans in February 2024 and June 2024 shows a significant decrease in the size of liver metastases from 5980 square units to 2181 square units, indicating a substantial response to the FOLFIRI + Bevacizumab treatment [Tool].
-Genetic analysis indicates MSI-High status (...)

Based on the most recent imaging and clinical data, Mrs. Patel's disease has progressed from no evidence of disease (NED) to metastatic disease involving the liver and potentially the spleen, as evidenced by the July 2024 CT scan showing multiple hypodense lesions [Tool]. This progression necessitates a comprehensive reassessment of her treatment strategy to address the metastatic spread.

b Completeness



How it is done: Patient X



Clinical Context

Ms. Xing, a 33-year-old patient, was diagnosed with cholangiocellular carcinoma in early 2023. Following her diagnosis, she underwent a complete surgical resection (R0) and was treated with adjuvant chemotherapy using capecitabine for six months. Subsequent imaging showed no evidence of disease recurrence until September of the same year, when MRI imaging ("September2023.png") revealed a new, solitary metastasis.

In response, a treatment regimen of gemcitabine and cisplatin was initiated. Considering Ms. Xing's young age, a liver biopsy was performed for comprehensive panel diagnostics. Both imaging results and panel findings are made available for review:

Radiology Report (* ...)

Liver: There is one single hypointense lesion in the left lobe of the liver, measuring approximately 0.4 cm in its greatest dimension (Location: [475, 250, 490, 275]). Highly suggestive for recurrence of the known cholangiocellular carcinoma.

(* ...)

Lymph Nodes: There is no enlargement of the abdominal or pelvic lymph nodes.

Peritoneum: No evidence of peritoneal implants or significant fluid accumulation is present.

Strong suspicion of a recurrence of the known CCC due to a solitary metastasis in the left liver.

Molecular report summary:

Microsatellite-Instability High, Mutation in BRAF V600E and CD74-ROS1 variant fusion.

No alterations in NTRK or FGFR2.

Currently, the patient experiences recurrent ascites with notable abdominal tension.

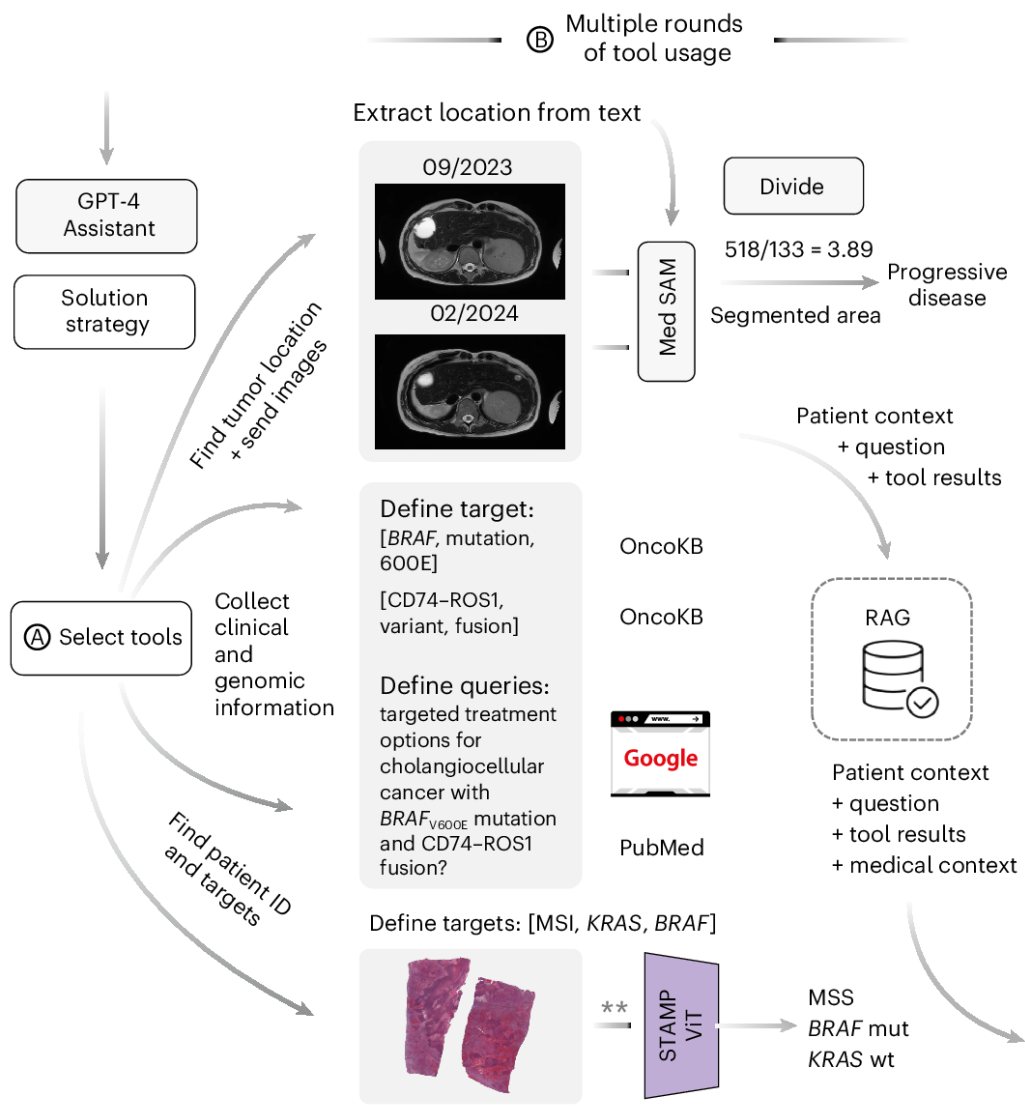
Also, a new MRI-scan was performed yesterday ("February2024.png") by our in house radiologists.

According to the latest radiology report, the MRI scan performed on the patient with the system ID X-09.22 and documented as "February2024.png" revealed a single metastatic lesion in the left lobe of the liver, delineated by the coordinates 455, 270, 505, 320. The report emphasizes the importance of a comprehensive review by comparing these findings with the images from a previous examination in September 2023 during the forthcoming tumor board meeting for a detailed analysis. Also, the imaging findings suggest the presence of peritoneal carcinomatosis. Her system ID is X-09.22.

Question

Please investigate the current state of the disease. What does the new MRI scan show?

Did the metastasis grow? In case of progress, what (targeted) treatment options could we suggest for the molecular tumor board tomorrow? Be aware that the patient is extremely young and we need all options we could find. Also check any options you can find on google or pubmed.



Ⓒ Write the answer

The latest MRI scan from February 2024 reveals that the metastatic lesion in Ms. Xing's left liver lobe has notably enlarged, growing from an area of 133 in September 2023 to 518, signifying a progression in her condition. This significant increase highlights the critical need to consider further treatment avenues for Ms. Xing. Given Ms. Xing's medical history and the molecular characteristics of her tumor, including the BRAF V600E mutation and CD74-ROS1 variant fusion, several targeted therapy options emerge:

- Dabrafenib and Trametinib:** This combination is recommended for BRAF V600E-mutated tumors, including cholangiocellular carcinoma, based on their effectiveness in other cancers with similar genetic alterations [Tool]. The combination has FDA approval for BRAF V600E mutant unresectable or metastatic melanoma and other solid tumors excluding colorectal cancer, with noted effectiveness in biliary tract cancer [Tool].
- Crizotinib, Entrectinib, and Repotrectinib:** These drugs are identified treatments for ROS1-positive tumors, primarily in NSCLC. Given the CD74-ROS1 fusion in Ms. Xing's tumor, these options could be explored for their potential applicability, despite the primary association with NSCLC [Tool].
- Encorafenib and Binimetinib:** This combination, highlighted through a Google search, could be considered based on their effectiveness in cancers with BRAF V600E mutations [Tool]. Although not specifically indicated for cholangiocellular carcinoma, their use in similar genetic contexts suggests potential applicability.

Regarding chemotherapy options, the current guidelines for cholangiocellular carcinoma recommend systemic chemotherapy as the primary treatment for advanced disease. Given Ms. Xing's prior treatment with gemcitabine and cisplatin, alternative regimens could include:

- FOLFOX (fluorouracil, leucovorin, and oxaliplatin) as a second-line treatment option, especially for patients who have progressed on a gemcitabine-based regimen [55].

For Ms. Xing, considering her young age and the urgent need for effective management strategies, participation in clinical trials exploring novel agents or combinations should be strongly considered. This includes trials for targeted therapies based on her tumor's molecular profile or emerging immunotherapy options for cholangiocellular carcinoma.

Given the conflicting information on microsatellite status, further clarification is needed. The 'check mutations' tool indicated Microsatellite Stability (MSS), however, the patient's molecular report specified that there was Microsatellite Instability (MSI). If MSI-High status is confirmed, immunotherapy with pembrolizumab could be considered, as it has shown effectiveness in MSI-High tumors across various cancer types [57].

In summary, for the tumor board discussion, the recommended approach includes exploring targeted therapy options such as Dabrafenib and Trametinib for the BRAF V600E mutation and potentially Crizotinib or Entrectinib for the CD74-ROS1 fusion. Additionally, considering second-line chemotherapy options and the possibility of clinical trial participation are crucial steps. The significant progression of Ms. Xing's disease necessitates a prompt and comprehensive evaluation of these treatment strategies to optimize her care.

To refine our approach and ensure the most accurate and personalized treatment recommendations for Ms. Xing, could you please provide the latest comprehensive genomic profiling of her tumor? This includes any recent sequencing data that might reveal additional actionable mutations or variants not previously identified. Additionally, detailed information regarding her response to past treatments, including any side effects experienced and the duration of response, would greatly enhance our ability to tailor future therapy options. This data is crucial for leveraging our resources to their fullest potential in guiding Ms. Xing's treatment plan.

Agents in details

Guideline database and RAG

Tools

Combine, Retrieve, and Generate responses

Guideline database and RAG

- Six sources
 - Clinical scores: MDCalc (<https://www.mdcalc.com/>)
 - General-purpose medical recommendations:
 - UpToDate
 - MEDITRON
 - Clinical Practice Guidelines in Oncology
 - American Society of Clinical Oncology
 - European Society of Medical Oncology
 - German and English subset of Onkopedia guidelines from German Society for Hematology and Medical Oncology
- Preprocessed by
 - GROBID for conversion of PDF into XML
 - Data cleansing to remove extraneous and irrelevant information
 - Reformat and standardize text from all sources
 - Add metadata and archive in JSON format for subsequent processing

Guideline database and RAG

- Embeddings
 - Model: OpenAI's 'text-embedding-3-large'
- Indexing
 - Text segments of varying lengths (512, 256, 128 tokens)
 - 50-token overlapping
- Retrieval
 - Vector database: Chroma
 - Lookup operation: cosine similarity

Tools: Web search

- Google API
 - Responses underwent text extraction and purification
 - And integrated directly as context (prompt) within the model
- Pubmed
 - Formulate custom PubMed queries
 - Responses were processed akin to above-described RAG in separate database
- OncoKB
 - Sending data through API
 - HUGO symbol
 - The change of interest (mutation, amplification or variant)
 - Specific alteration of interest (such as *BRAF*^{V600E}) if applicable
 - Returns
 - potential FDA-approved or investigational drug options including evidence levels

Tools: Imaging (CT or MRI scans)

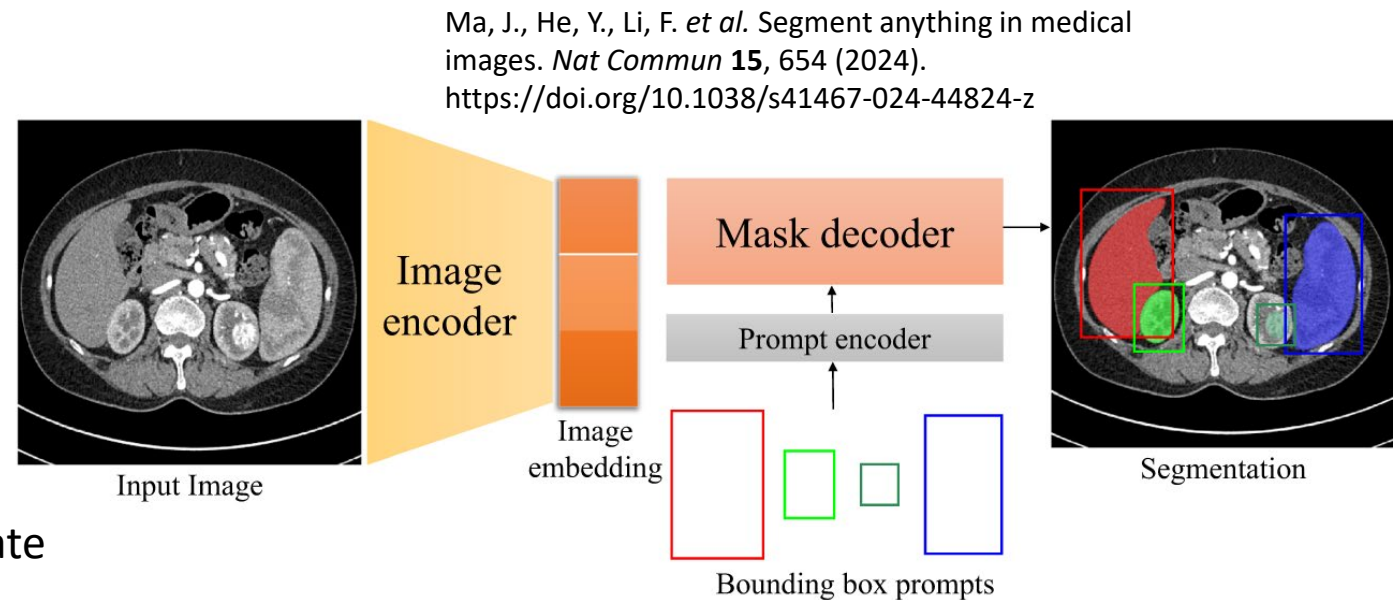
2 options:

1. GPT-4 Vision model

- Generate a comprehensive, detailed and structured report from provided CT or MRI

2. MedSAM

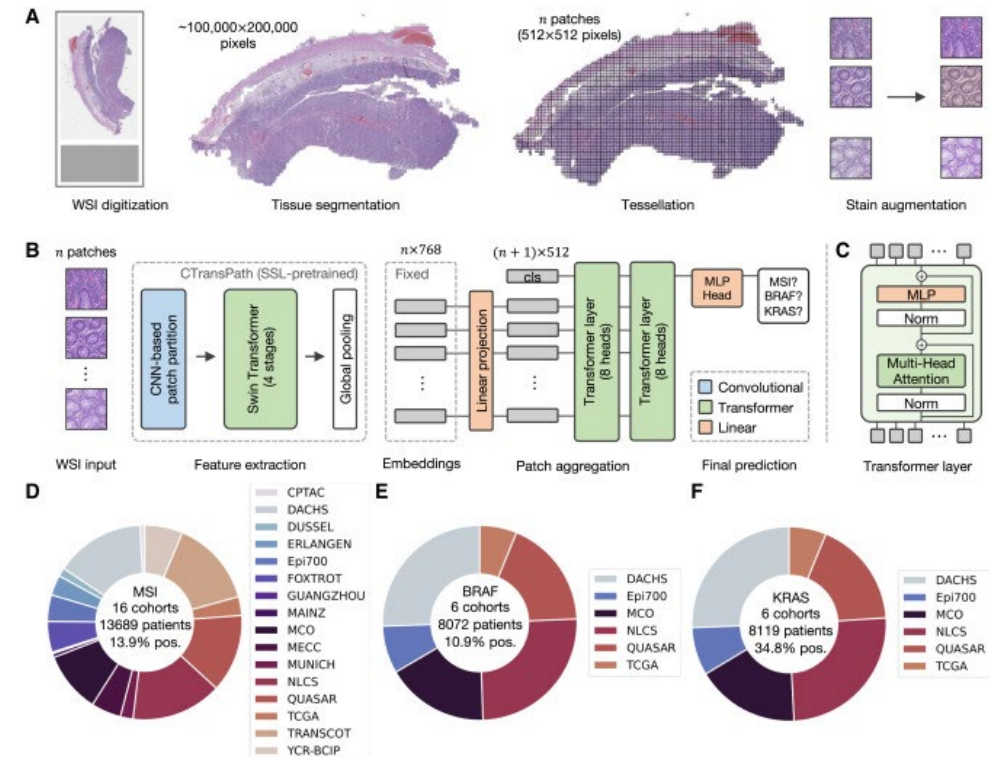
- Measure size of lesions using Calculator (local python)
- In scenarios involving multiple images from different date
 - First investigates and reports on each image separately
 - Then synthesizes a comparative analysis
 - GPT-4 can track results from MedSAM back to original image date



Tools: Histopathology

- In-house training a vision classifier model
- Input
 - Histopathology features extracted from colorectal cancer tissue image in The Cancer Genome Atlas (TCGA) program
 - CTranspath for feature extraction
- Output: binary prediction
 - MSI vs MSS
 - KRAS mutation vs Wild type
 - BRAF mutation vs Wild type
 - + Mutation probability in percentage

Wagner, Sophia J., et al. "Transformer-based biomarker prediction from colorectal cancer histology: A large-scale multicentric study." *Cancer Cell* 41.9 (2023): 1650-1661. <https://doi.org/10.1016/j.ccell.2023.08.002>



Main pipeline: Strategy!

LLM library = DSPy

Base Model = OpenAI's 'gpt-4-0125-preview'

Temperature = 0.2 and 0.1 during RAG

Vision Model = OpenAI's 'gpt-4-vision-preview'

Embedding Model = OpenAI's 'text-embedding-3-large'

Reranking Model = Cohere Rerank 3 English

PROMPT =

"""

You are a medical AI assistant trained by OpenAI, based on the GPT-4 model.

You will receive medical information about a patient and a question from a medical doctor.

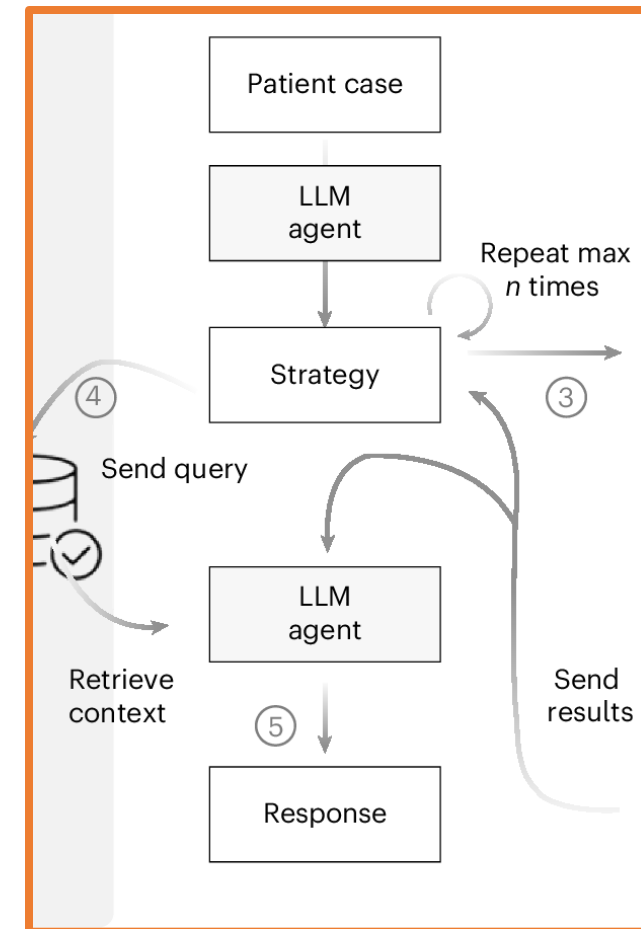
Lets think step by step.

First think about the information you received. Then check your available tools. Develop a strategy to get all relevant information using multiple rounds of tools if necessary. You can also combine tool outputs and inputs.

Then, run all tools that you consider useful.

Finally, do NOT answer the user question. Instead, summarize the new information we have received from the tools and draw conclusions. Include every detail.

"""



Main pipeline: combine, retrieve, response!

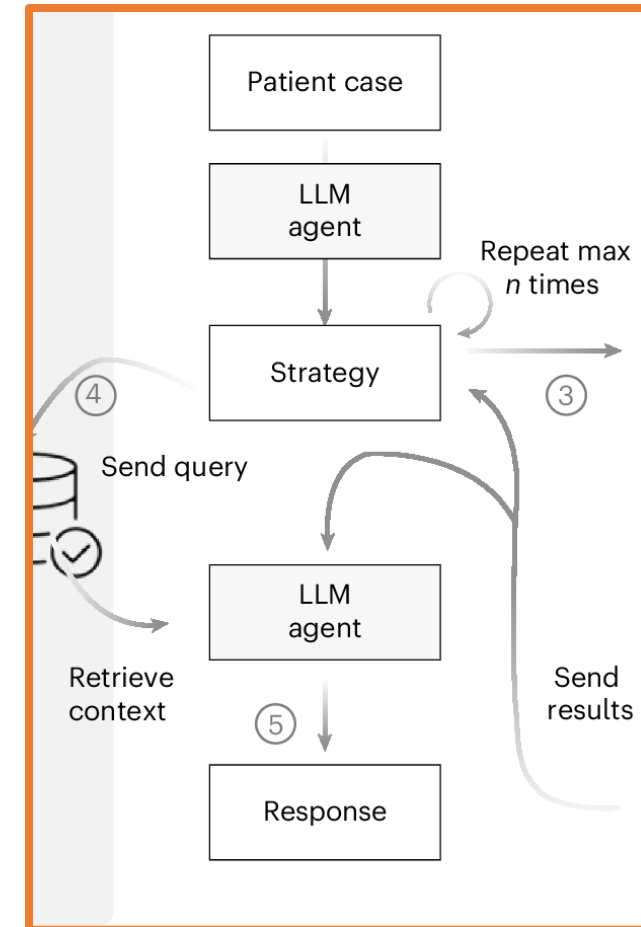
Algorithm 1 Response generation algorithm. This algorithm generates a the model's final response after the agent has invoked the necessary tools in response to the question. It involves several calls to the LLM which are denoted as functions of the form **Generate***(\cdot); these functions invoke the LLM with a templated prompt into which the input arguments are inserted. CoT = Chain-of-Thought, P = Predict (akin to `dspy.Predict`).

Require:

- Patient context $C_{patient}$ ▷ Contains clinical information about the patient
- Question Q ▷ Question(s) or instruction(s) related to the patient
- Agent tools A_{tools} ▷ List of all available tools
- Tool outputs T_{out} ▷ Free-text summary of the results of the tools used by the agent
- Documents D ▷ Collection of medical documents (guidelines, textbooks, etc.)

Ensure: Comprehensive and accurate response R

- 1: $C_Q \leftarrow$ empty list ▷ Question context (will contain passages relevant to Q)
- 2: $Subqueries \leftarrow$ **GenerateSubqueries**_{CoT}(C, Q, T_{out}) ▷ Generate 10-14 sub-questions for RAG
- 3: **for each** subquery q **in** $Subqueries$ **do**
- 4: $P \leftarrow$ **Retrieve**_n(D, q) ▷ Retrieve $n = 40$ passages relevant to the subquery q
- 5: $P \leftarrow$ **Rerank**(P, q) ▷ Re-rank the retrieved passages based on relevance to q
- 6: $P \leftarrow$ **Top**_k(P) ▷ Keep only the top $k = 10$ passages
- 7: $C_Q \leftarrow C_Q \cup P$ ▷ Add the passages to the question context
- 8: **end for**
- 9: $C_Q \leftarrow$ **Deduplicate**(C_Q) ▷ Remove duplicate passages
- 10: **for** i **in** $1, \dots, |C_Q|$ **do** ▷ Add numbered source information to each passage
- 11: $C_Q[i] \leftarrow$ **Concat**("Source ", $i, C_Q[i]$)
- 12: **end for**
- 13: $Strategy \leftarrow$ **GenerateAnswerStrategy**_{CoT}($C, Q, A_{tools}, T_{out}, C_Q$)
- 14: $CitedResponse \leftarrow$ **GenerateCitedResponse**_P($C, Q, A_{tools}, T_{out}, C_Q, Strategy$)
- 15: $Suggestions \leftarrow$ **GenerateSuggestions**_{CoT}($CitedResponse, A_{tools}, T_{out}$)
- 16: $R \leftarrow$ **Concat**($CitedResponse, Suggestions$) ▷ Final output
- 17: **return** R



Evaluation

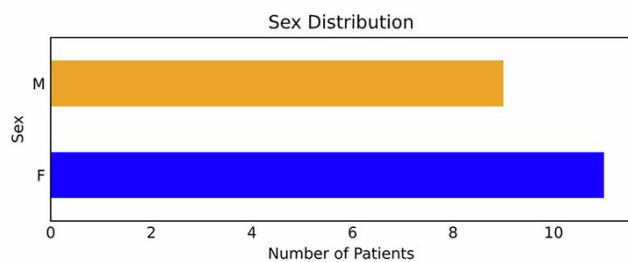
Dataset

N = 20 patient cases

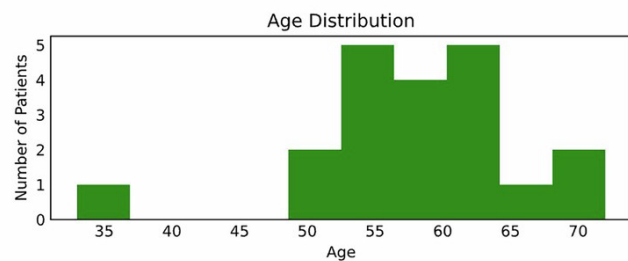
15 random permutations of sex, age, ethnicity

20*15 = 300 in total

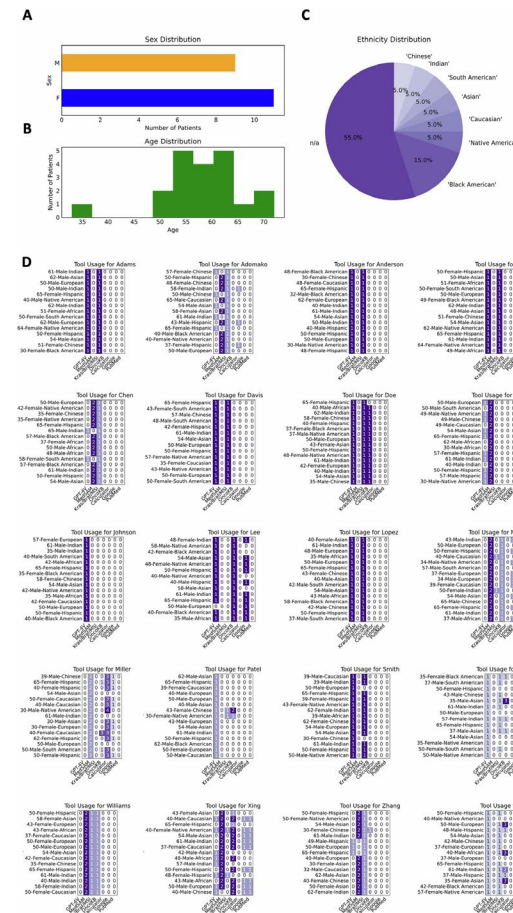
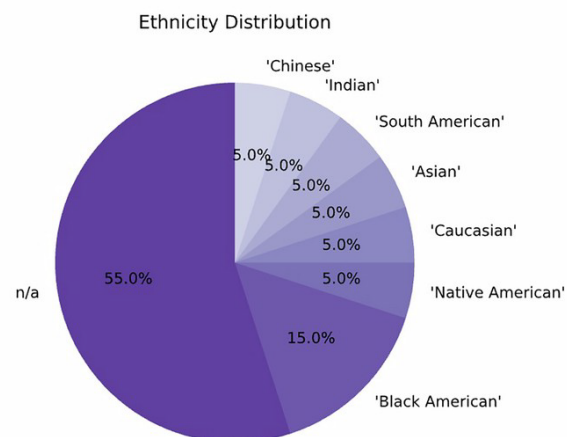
A



B



C



Human results evaluation

3 primary aspects evaluated independently by 4 certified clinicians with expertise in oncology using majority vote.

1. Use of tools

- Ratio of actual versus required uses
- Required tool use is defined as
 - Directly instructed to use a tool OR
 - Specific output of a tool is needed to answer the question

2. Quality of text output

- Correctness
 - Correctness: proportion of correct replies relative to all output
 - Incorrectness: hallucinations of diagnostic procedures or patient information
 - Harmful: hallucinations of suboptimal or contraindicated treatments
- Completeness
 - Essential information that oncologists would anticipate in a human-generated answer under similar conditions
 - Specific keywords and terms for each scenario
 - Expected relevant and specific interventions (for examples, “FOLFOX and bevacizumab” instead of “chemotherapy and antiangiogenic drugs”)
- Helpfulness
 - Ratio of answers to all questions given by the user

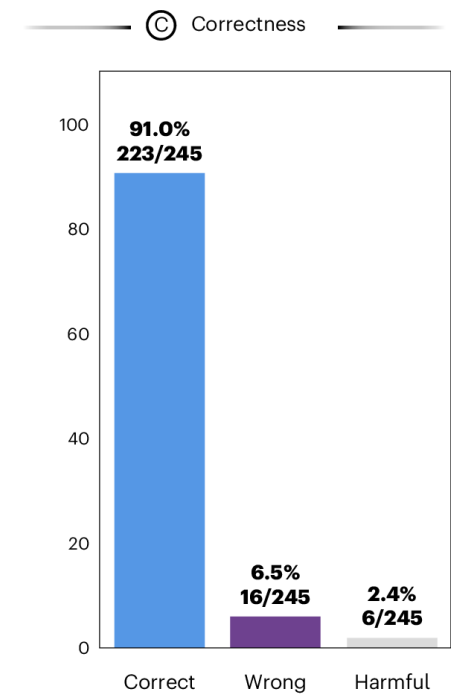
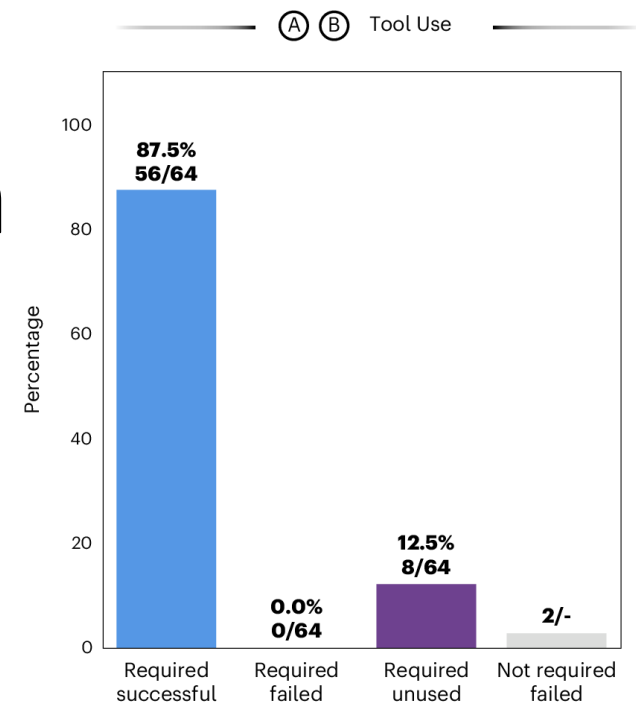
3. Accurate citation through RAG

- Correct: model’s text mirror the content of original document
- Irrelevant: instances where model’s assertions are NOT substantiated by the source material
- Wrong: discrepancies of the information and its actual content

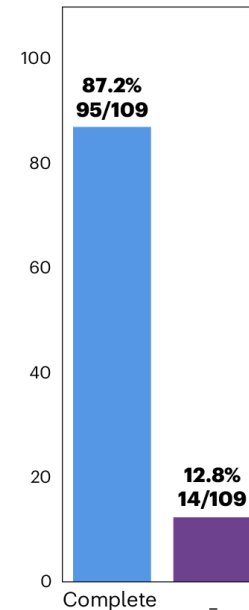
Human results evaluation

3 primary aspects evaluated independently by 4 certified clinicians with expertise in oncology using majority vote.

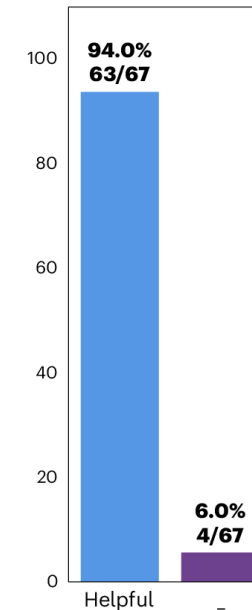
1. Use of tools
 - Ratio of actual versus required uses
2. Quality of text output
 - Correctness
 - Completeness
 - Helpfulness
3. Accurate citation through RAG
 - Correct
 - Irrelevant
 - Wrong



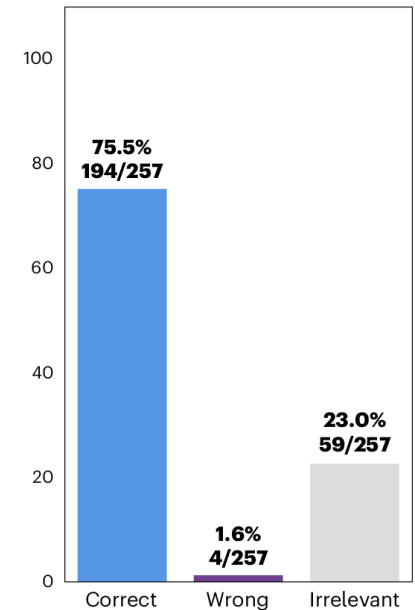
(C) Completeness



(C) Helpfulness



(C) Citation



Comparing models: with and without tools

a Example responses from GPT-4 alone in comparison to GPT-4 with tool use and RAG

GPT-4

As an AI developed by OpenAI, I'm unable to access or analyze real patient data, images, or external databases directly, nor can I use tools for precise measurement of lesions described in radiology reports. Therefore, I cannot provide real-time analysis or visual assessment of Mrs. Chen's radiology images (February 3 2024.png and June 25 2024.png) or perform precise measurements of the lesions. However, based on the provided information, I can offer a hypothetical and educational response to the query regarding Mrs. Chen's current state of disease and a detailed treatment plan.

Given Mrs. Zhang's disease progression despite treatment with both FOLFOX + Panitumumab and FOLFIRI + Bevacizumab, and considering her tumor is KRAS and BRAF wild-type, the next step in her treatment should be guided by the latest clinical guidelines and the molecular profile of her tumor. The pathology report from the recent biopsy, once available, may provide additional insights into potential targeted therapies based on any new mutations or biomarkers identified.

Current State of Disease: Based on the most recent follow-up and imaging data available up to March 2024, Mrs. Patel shows no evidence of disease (NED). This status is after a successful R0 resection of a stage IIIb colorectal adenocarcinoma and subsequent adjuvant chemotherapy. Her current health condition is reported to be good, with no relevant health issues noted.

GPT-4



Tools



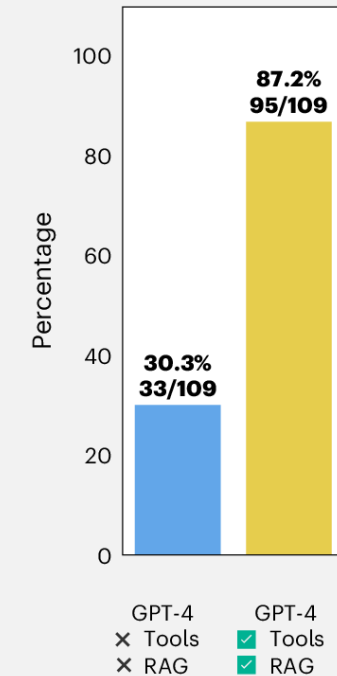
RAG

Based on the provided data and tool results, Mrs. Chen's metastatic colorectal cancer has demonstrated progression from February to June 2024. The significant increase in the size of lung metastases, as quantified by the segmentation tool results showing an increase in the overall area of the lesions from 702 to 2798 square units, indicates that the disease is not adequately controlled under the current regimen of FOLFIRI and bevacizumab [Tool]. Furthermore, the genetic mutation analysis revealing MSS status and the absence of BRAF and KRAS mutations provides valuable information for guiding future treatment decisions [Tool]. Given the progression of disease despite current systemic therapy, and considering Mrs. Chen's stable ECOG performance status of 1, which suggests she remains in relatively good condition to tolerate treatment, the following treatment plan is proposed: ...

Based on the comprehensive analysis of Mrs. Zhang's medical history, radiological data, and genetic testing, we have a clear picture of her current oncologic status and can make informed recommendations for her ongoing treatment strategy. (...)
-The quantitative analysis of liver metastases from CT scans in February 2024 and June 2024 shows a significant decrease in the size of liver metastases from 5980 square units to 2181 square units, indicating a substantial response to the FOLFIRI + Bevacizumab treatment [Tool].
-Genetic analysis indicates MSI-High status (...)

Based on the most recent imaging and clinical data, Mrs. Patel's disease has progressed from no evidence of disease (NED) to metastatic disease involving the liver and potentially the spleen, as evidenced by the July 2024 CT scan showing multiple hypodense lesions [Tool]. This progression necessitates a comprehensive reassessment of her treatment strategy to address the metastatic spread.

b Completeness



Comparing models: close and open source

Llama-3 70B (llama3-70b-8192) Mixtral 8x7B (mixtral-8x7b-32768)

Temp = 0.2, Max output token = 4096

a Example tool results from three state-of-the-art LLMs

LLama

✗ Calling „segment_image“: Error: No such file: ...
 ✗ Calling „segment_image“: Error: No such file: ...
 ✗ Calling „calculate“: "The sum of 75.3 and 10.5 is 85.8."
 It seems like we've reached the end of the tool chain. I'll stop here.

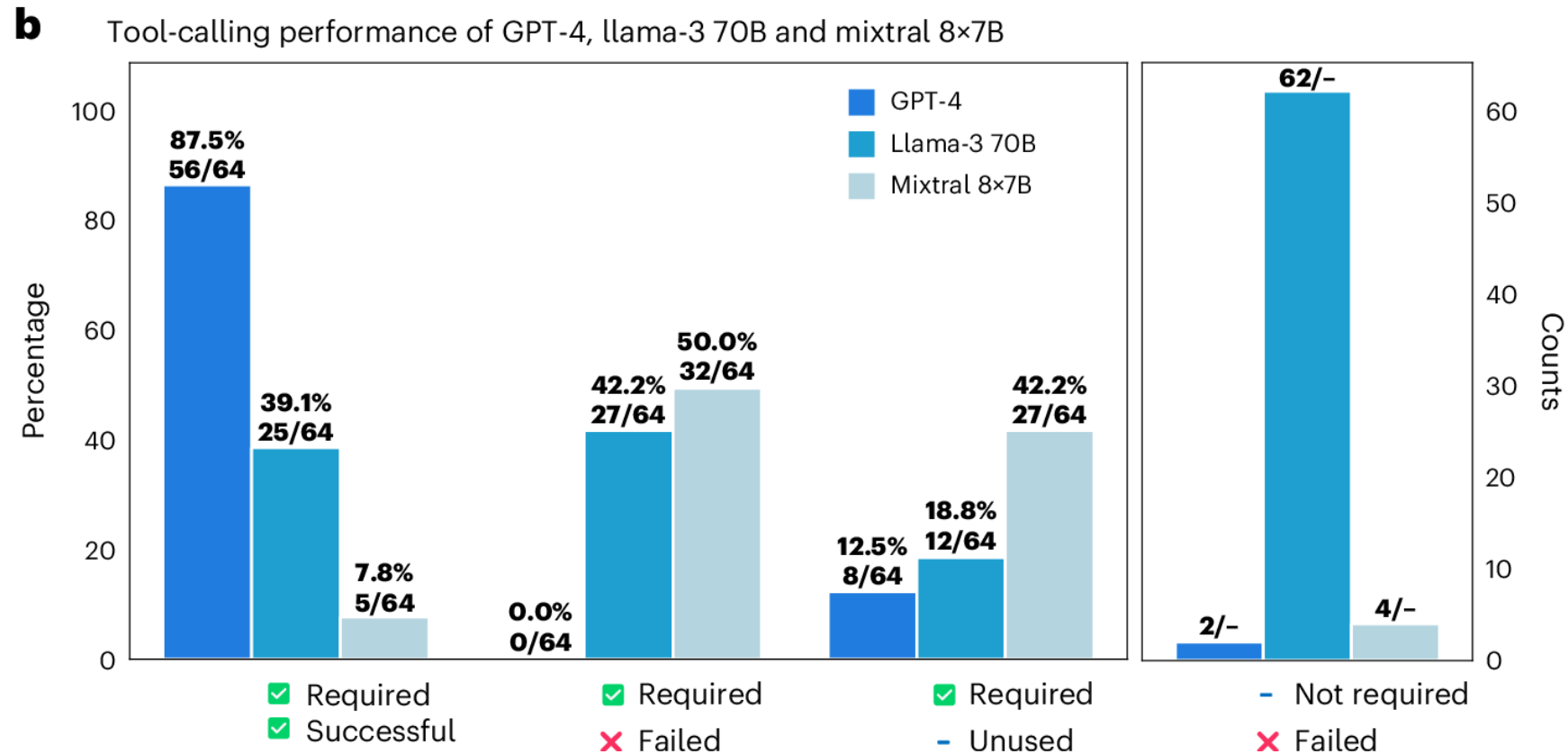
Mixtral

✗ Calling „segment_image“: Error: No such file: ...
 The tool call with id "call_194n" failed due to a file not found error. The specified file path for the image "MRI_February_28_2024.png" is incorrect or the file does not exist in the specified location. Since the tool call failed, there is no new information to summarize.
 It's important to double-check the file path and make sure the image file is available in the specified location before attempting to call the tool again.

GPT-4

✓ 1. Segmentation of MRI images: The MRI images from December 3, 2023, and February 28, 2024, were analyzed to determine the size of the hepatocellular carcinoma (HCC) lesion. The segmentation tool provided the sizes of the regions of interest as follows:
 - December 3, 2023, MRI: The area was 2077 square units.
 - February 28, 2024, MRI: The area was 2057 square units.
 ✓ 2. Calculation of tumor volume change: The change in the tumor's size was calculated to determine if the disease had progressed according to RECIST criteria. The calculations revealed:
 - The difference in tumor size between the two dates was -20 square units, indicating a slight decrease in size.
 - The percentage change in tumor size was calculated as approximately 0.96%.

Comparing models: close and open source



Discussion

- LLM (GPT-4) alone only generates very generic or wrong responses
- Using LLM as reasoning engine + knowledge database and tools
 - Improves accuracy and reliability
 - Increases efficacy in very context-specific and specialized medical queries
 - Predicting rare mutations
 - Measuring disease progression
 - Allows rapid update of medical knowledge by searching Google or PubMed without the need to update the base model itself
- Using RAG addresses hallucination by supporting citation allowing quick fact-checking

Limitation

- Small sample size for evaluation
- Core focus lies in the tool-using abilities
 - Each tools require independent optimization and validation
- Experimental stage limiting clinical applicability. For examples:
 - Using one slice of radiology images
 - More advanced medical image analysis models to replace tool in this pipeline
 - Using lesion size to evaluate disease progression
 - More input data to better evaluate current state of disease
- Limited utilization of tools compared to clinicians
- Multiturn chat agent but evaluate as single interaction
- Oncological use cases only

Future work

- Locally deployed open-source models
 - Llama-3 405B (newer and better)
 - Hermes-3-Llama-3.1 (medical fine-tuned)
- Potential improvements to the retrieval processes
 - Domain-specific models
 - Hybrid search
- Larger context windows (such as Gemini 1.5)
- How LLM handles temporal dependencies in treatment recommendations
- Training of model specialized in tool selection