Accelerating Evidence Generation and Time to Insight With Clinical AI

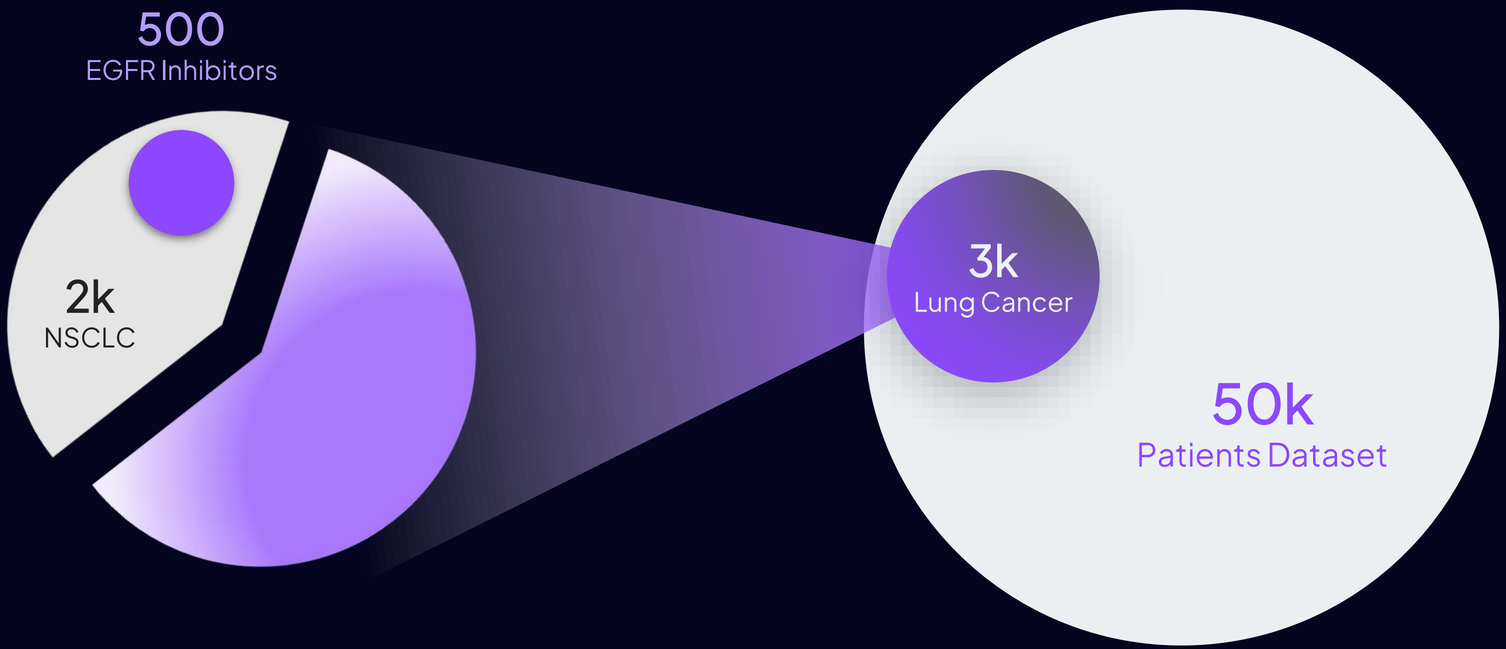
Jose Mena, Sr. Director of Data and Bioinformatics, Mendel AI
Vivek Rudrapatna, MD, PhD, Co-director, Center for Real-World Evidence, UCSF Health
Wael Salloum, PhD, Co-Founder and Chief Scientific Officer, Mendel AI

HEOR Needs Clinical AI

Jose Mena

Success hinges on isolating a narrow cohort & understanding it in-depth

"Find patients who have advanced NSCLC with more than one EGFR TKI"



With Clinical AI, not just any AI

01 CLINICAL REASONING

02 PATIENT-LEVEL, NOT DOCUMENT-LEVEL

03 NO HALLUCINATIONS

04 EXPLAINABLE

Clinical AI Market Landscape

Innovators

AGI
- OpenAI
- cohere
- Google DeepMind

Clinical AI
- MENDEL
- UNLEARN
- PaLM
- Isomorphic Labs

Tools

Domain Agnostic
- Hugging Face
- databricks

Clinical domain specific
- John Snow LABS
- PRIVATE AI

Wrappers

Provider - facing
- ConcertAI
- PathAI
- iz.ai
- AUGMEDIX

Payer - facing
- Hippocratic AI — Do No Harm —
- DataRobot

Pharma - facing
- ATROPOS HEALTH
- HAYSTACK ONCOLOGY
- ConcertAI
- komodo HEALTH
- PathAI
- Benchling
- iz.ai

REASONING

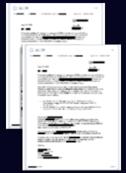
LLM Trained on
medical records

Symbolic AI
Coded by
physicians

Reasoning over longitudinal data is challenging

"Find patients who have advanced NSCLC with more than one EGFR TKI"

Pathologist



"Mass, left lung, needle biopsy: adenocarcinoma, moderately differentiated"

"Left lobe lower biopsy revealed a pulmonary adenocarcinoma. She was staged with 3A (T2, N2, M0) lung cancer of the left lower lobe."

Claims data

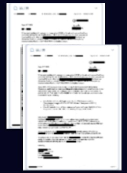
"Neoplasm of lung of uncertain behavior"

History of Tarceva 100mg

Radiologist

"Brain metastasis, for which the patient received palliative radiation therapy"

Sequencing data



T790M mutation

Clinical note

"Developed resistance to TKI"

Standard ontologies vs ontologies for reasoning

	Standard Ontologies	Ontology for reasoning
Use case	Standard vocabulary for coding & billing	Knowledge representation for efficient & effective reasoning
Ease of use	● Too many concepts (3M UMLS) & duplicates ○ Carcinoma appears 4 times in SNOMED CT ● Illogical hierarchy ○ Tagrisso not included under EGFR inhibitors ● Missing concepts ○ "EGFR + lung cancer" but no EGFR	● Reductionism: break down into smallest conceptual units ("concemes") ● Coherence: hierarchy designed logically by clinical experts over years ● Generative: concemes can be combined into any thought natural language can express
Updates	Updated monthly or yearly	Constantly augmented driven by clinical experts and machine learning

Cohort retrieval is a fundamental task in healthcare AI

Closing gaps in clinical practice

Clinical trial recruitment

Descriptive epidemiology studies

Value based care programs

Clinical trial design and optimization

Comparative effectiveness studies

Risk factor identification

Curating data from clinical notes to improve RWE studies

A comparative methods pilot study

Vivek Rudrapatna
Assistant Professor, Department of Medicine
Co-director, Center for Real-World Evidence
University of California, San Francisco

Disclosures

- Research grants to UCSF:
  - Janssen, Merck, Takeda, Genentech, Stryker, Blueprint Medicines, Mitsubishi Tanabe, Alnylam
- Shareholder:
  - ZebraMD
- I am not an employee or consultant of Mendel AI

Background

- Electronic health records data have significant potential to support studies of treatment effectiveness, safety, and value
  - Contain detailed information about patients, treatments, and outcomes
  - Can help identify cohorts, control confounding, estimate causal effects
- But…there's a catch
  - Much of the data needed for real-world evidence studies is unstructured
- Computational methods are needed to accurately structure these data for downstream research

GPT-4 (OpenAI)

- State-of-the art on many clinical language inference tasks
  - Incl. information extraction from clinical notes

- However, there are limitations:
  - GPT-4 is (probably) trained on general-purpose text
    - May not be sufficiently educated on technical domains like medicine
    - Greater risk of inaccuracies (“hallucinations”)
  - GPT-4 is a closed model, hidden behind an API endpoint in the cloud
    - Privacy, security concerns
    - Cost concerns[1]
    - No transparency, interpretability, explainability
      - Potentially prone to algorithmic bias[2]

1. Yim and Rudrapatna, medRxiv, 2024
2. Zack et al, Lancet Dig Health, 2024

Hypercube (Mendel AI)

- A neuro-symbolic reasoning system built for clinical data
  - Utilizes structured medical knowledge (ontologies)
    - Modularizable
    - Auditable
    - Potentially faster run-time and lower inference cost
  - Can identify cohorts, ascertain outcomes in a no-code format

- Not as well known as a clinical data science tool outside of oncology

Objective

- To compare GPT-4 and Mendel on cohort retrieval tasks to support clinical studies
  - Use cases: Inflammatory Bowel Disease, Pancreatic Cancer
  - Target variables:
    - Confounders and outcomes
      - Incl. patient reported outcomes (PROs)
    - Treatment exposures

Methods

- We provided Mendel with annotated notes + annotation protocols
  - IBD PROs (300 labels each, binary)
    - Abdominal Pain
    - Diarrhea
    - Rectal Bleeding
  - IBD Medication History (150 labels each, binary)
    - Previously completed/discontinued treatments
    - Current treatments
    - Planned future treatment
  - IBD disease complications (20 notes, binary)
    - History of a fistula, stricture, bowel resection, ostomy
- Inter-annotator agreement > 90% for all tasks

Methods

- Pancreatic Cancer (20 labels for each)
  - Stage (TNM)
  - Genetic/Genomic testing, results
  - Prior treatments
- Developed a question bank for cohort retrieval designed to test set consistency of retrieved cohorts
  - Synonymy: Do synonymous questions retrieve identical cohorts?
  - Subsumption: Are specific cohorts subsets of more general ones?
    - For example: patients with lung cancer should be a subset of patients with cancer
- Patient charts underwent automated retrieval using GPT-4, Hypercube
  - GPT-4: Full context prompting for retrieval, theoretic limit of performance
  - Hypercube: Prompting with symbolic query language
- MEA on gold annotations still underway

Preliminary Results: Subsumption

Retrieval question		Hypercube		GPT-4	
Top-level category	**Lower category**	**Answers**	**Answers not in parent**	**Answers**	**Answers not in parent**
Show me patients with IBD that have been treated with a biological therapy	Show me patients with IBD that have been treated with an anti-TNF drug	211	0	226	3
Show me patients with IBD that have been treated with a biological therapy	Show me patients with IBD that have been treated with an anti-integrin drug	67	0	72	1
Show me patients with IBD that have been treated with a biological therapy	Show me patients with IBD that have been treated with anti-interleukin therapy	53	0	81	1

*further MEA needed

Preliminary Results: Synonymy

Retrieval question		Hypercube		GPT-4		
Generalized Question	Expanded Question	Answers for both	Set difference	Answers for both	In general but not expanded	In expanded but not general
Show me patients with IBD that have been treated with a biological therapy	Show me patients with IBD that have been treated with Infliximab or Adalimumab or Certolizumab or Golimumab or Vedolizumab or natalizumab or risankizumab or ustekinumab	234	**0**	237	7	3
Show me patients with IBD that have been treated with a JAK inhibitor	Show me patients with IBD that have been treated with Tofacitinib or Upadacitinib	10	**0**	9	2	1
Show me patients with IBD that have been treated with an anti-TNF drug	Show me patients with IBD that have been treated with Infliximab or Adalimumab or Certolizumab or Golimumab	211	**0**	194	32*	3
Show me patients with IBD that have been treated with an anti-integrin drug	Show me patients with IBD that have been treated with Vedolizumab or natalizumab	67	**0**	60	12*	2
Show me patients with IBD that have been treated with anti-interleukin therapy	Show me patients with IBD that have been treated with risankizumab or ustekinumab	53	**0**	40	41*	0

Preliminary Results: Set composition

Set composition		Hypercube			GPT-4		
Parent category	**Composed child categories**	**Answers in parent**	**Answers in children**	**Difference**	**Answers in parent**	**Answers in children**	**Difference**
Show me patients with IBD that have been treated with a biological therapy	Show me patients with IBD that have been treated with an anti-TNF drug OR an anti-integrin drug OR an anti-interleukin therapy	234	234	0	244	249	5

Discussion

- Both models achieved similar performance across a wide range of clinical domains and target tasks
  - Very simple questions: what happens as they become more complex?

- Potential advantages to the Mendel system:
  - Lower cost/computational burden
  - Greater transparency
  - Set closure
  - Use of ontologies -> can incorporate new medical knowledge

- Potential advantages to GPT, other GenAI models:
  - Extremely well-tested across domains
  - Potentially more versatile
    - Question answering, summarization
    - Potential for multimodal inference
      - E.g. obtain clinical endpoints from images, waveforms, etc.

Conclusion

- GPT-4 and Hypercube appear to have similar accuracy on information extraction for at least two use cases

- Future work is needed to
  - More rigorously compare these models + other open source comparators
  - Enhance their accuracy to meet (or exceed) human annotators
  - Apply them to enable higher quality RWE studies

Automatic Cohort Retrieval (ACR)

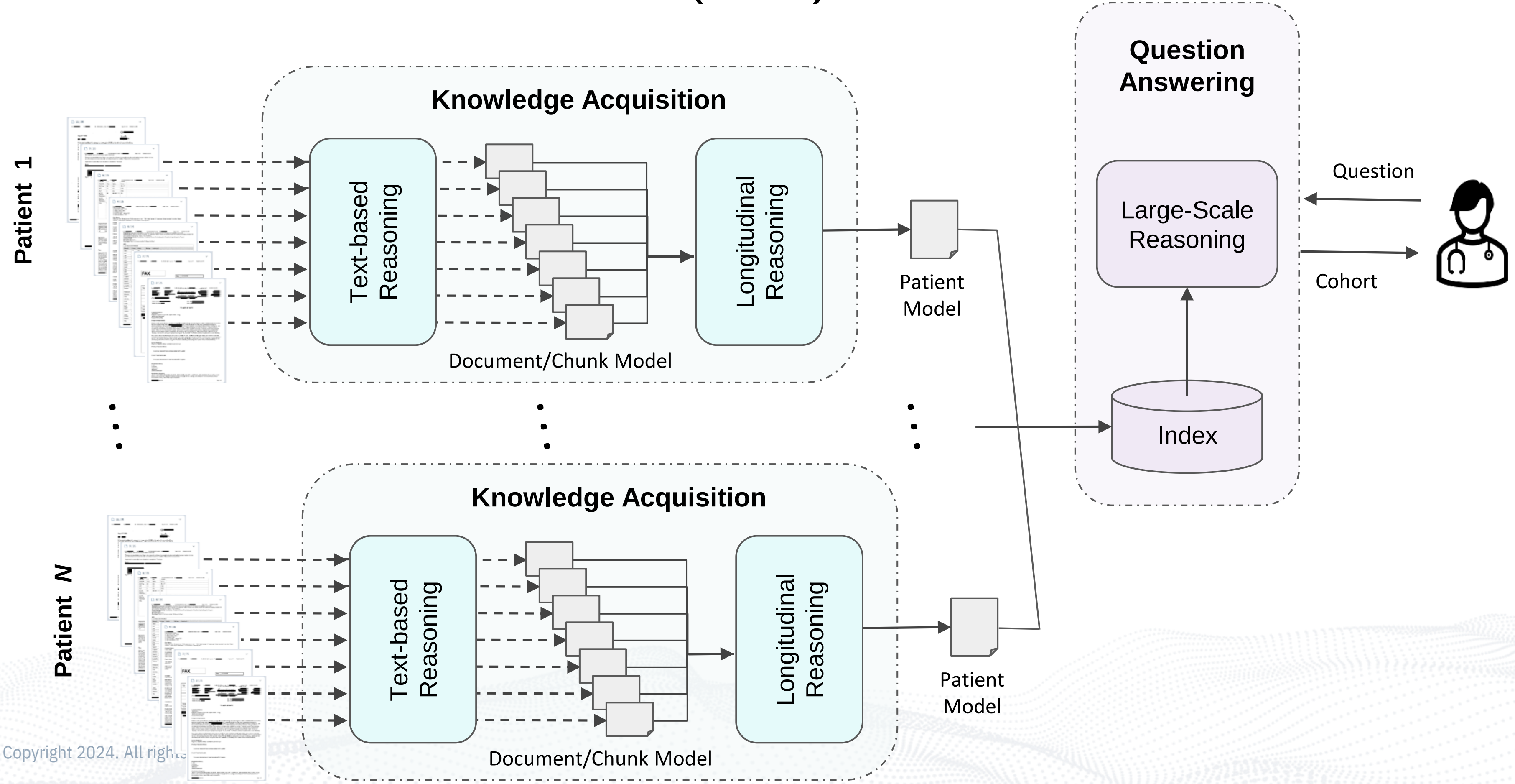
Wael Salloum, PhD

Topics

1. **Read:** Text-based Reasoning
    - NLP/NLU on free text; produces normalized facts mapped to standard ontologies
2. **Resolve:** Longitudinal Reasoning
    - Reasoning across facts over time to build a consolidated patient's journey
3. **Hypercube:** Large-Scale Reasoning
    - Cohort selection & question answering over those patient journeys

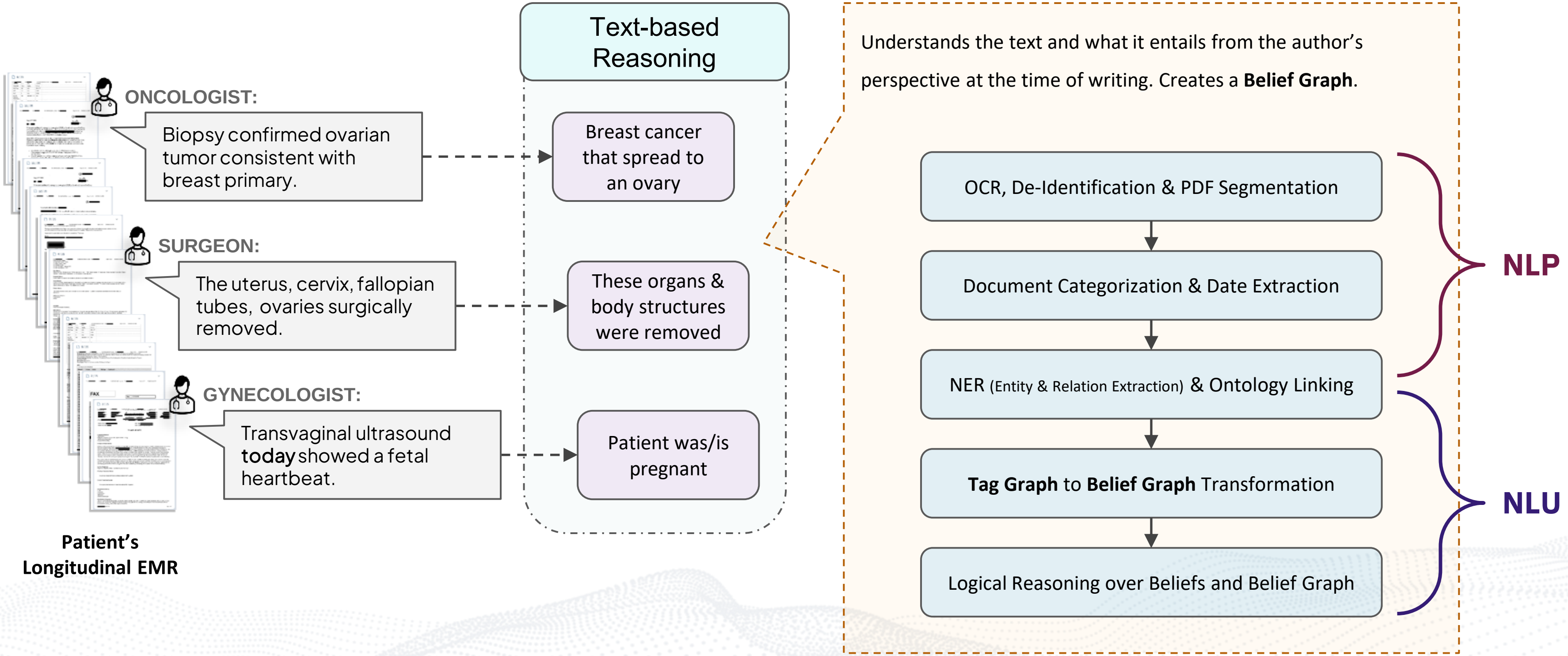
MENDEL

Automatic Cohort Retrieval (ACR)



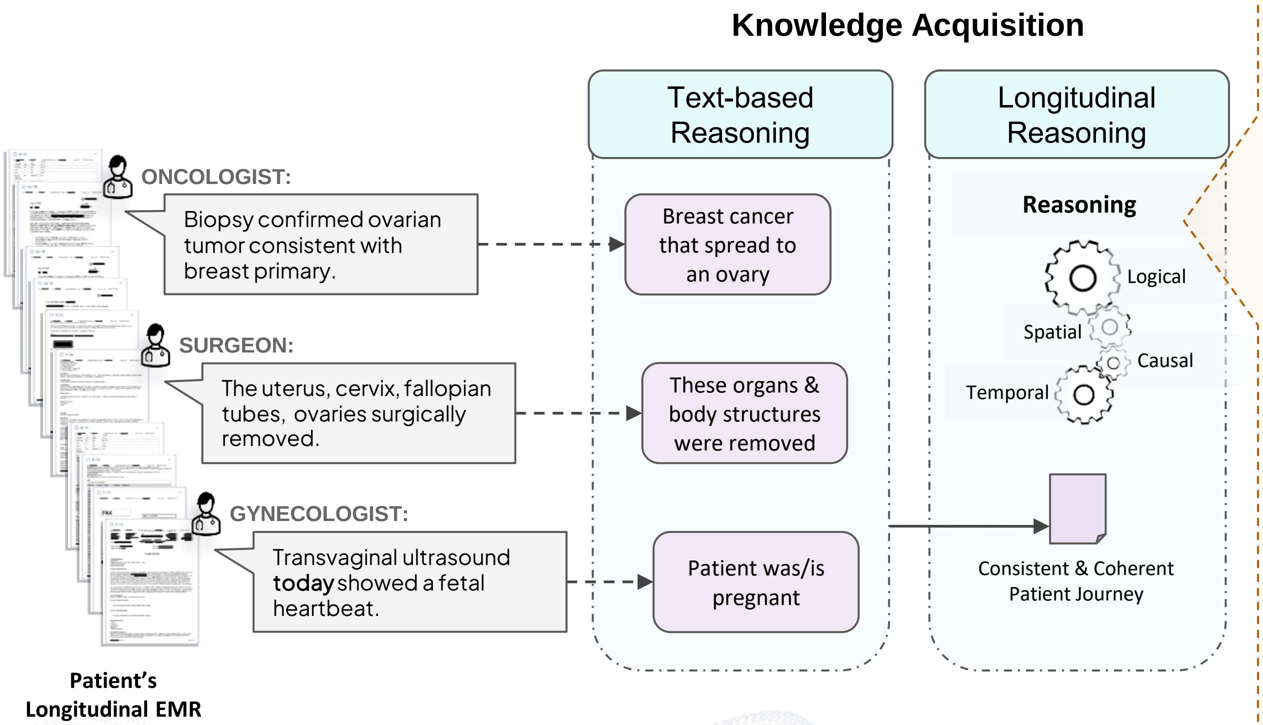
Example: Text-based Reasoning

Knowledge Acquisition



Example: Longitudinal Reasoning

MENDEL



Detects & resolves a logical paradox in event order.

Options:

1. **Reorder events:** Surgery last to treat the cancer ⇒ pregnancy first:
   ⇒ Coherent Journey
1. **Discard events:** One or more events are wrong, delete:
   ⇒ Consistent Journey

Reordering options:

- ❑ Pregnancy → Cancer → Surgery(treating)
- ❑ Cancer → Pregnancy → Surgery(unrelated)

Discard 1 event options:

- ❑ Cancer & Surgery ; ~~NO Pregnancy~~
- ❑ Pregnancy & Surgery ; ~~NO Cancer~~
- ❑ Pregnancy & Cancer ; ~~NO Surgery~~

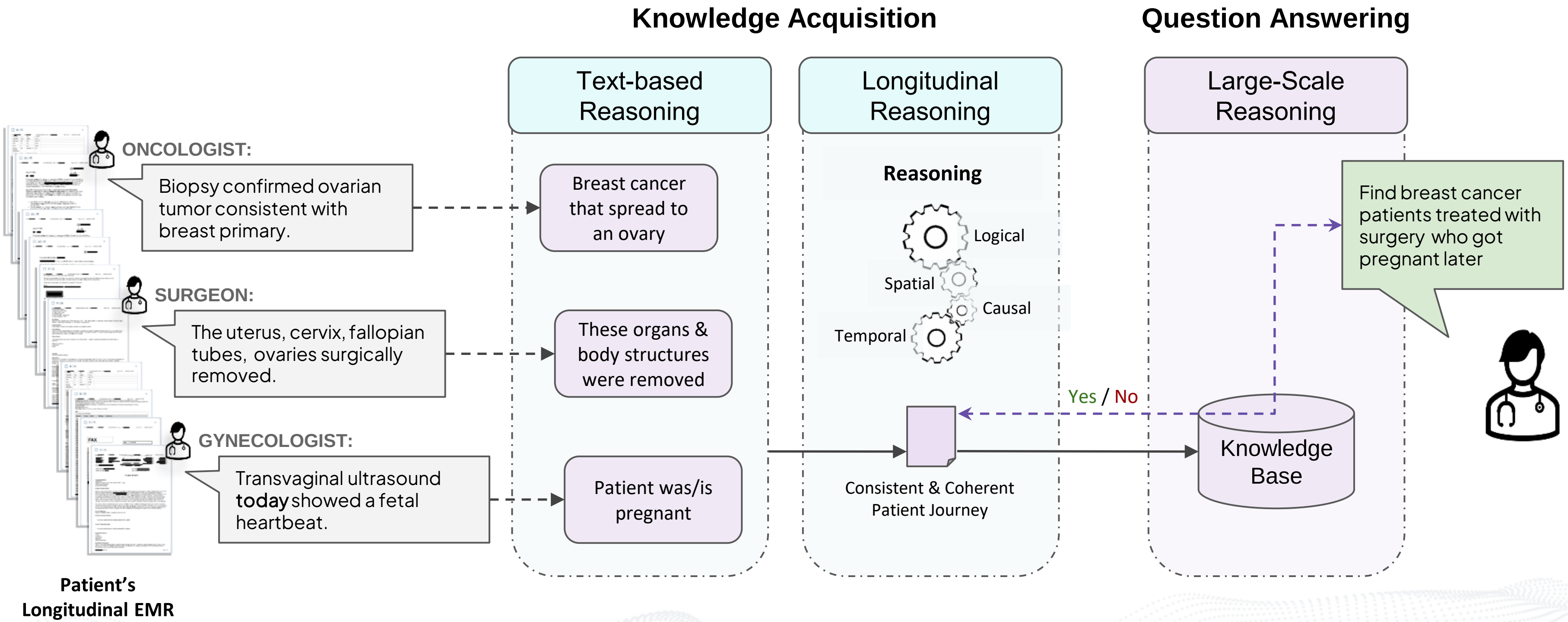
Discard 3 events options:

- ❑ Cancer ; ~~NO Pregnancy & Surgery~~
- ❑ Pregnancy ; ~~NO Cancer & Surgery~~
- ❑ Pregnancy ; ~~NO Surgery & Cancer~~

Discard all events options:

- ❑ {}

Example: Large-Scale Reasoning



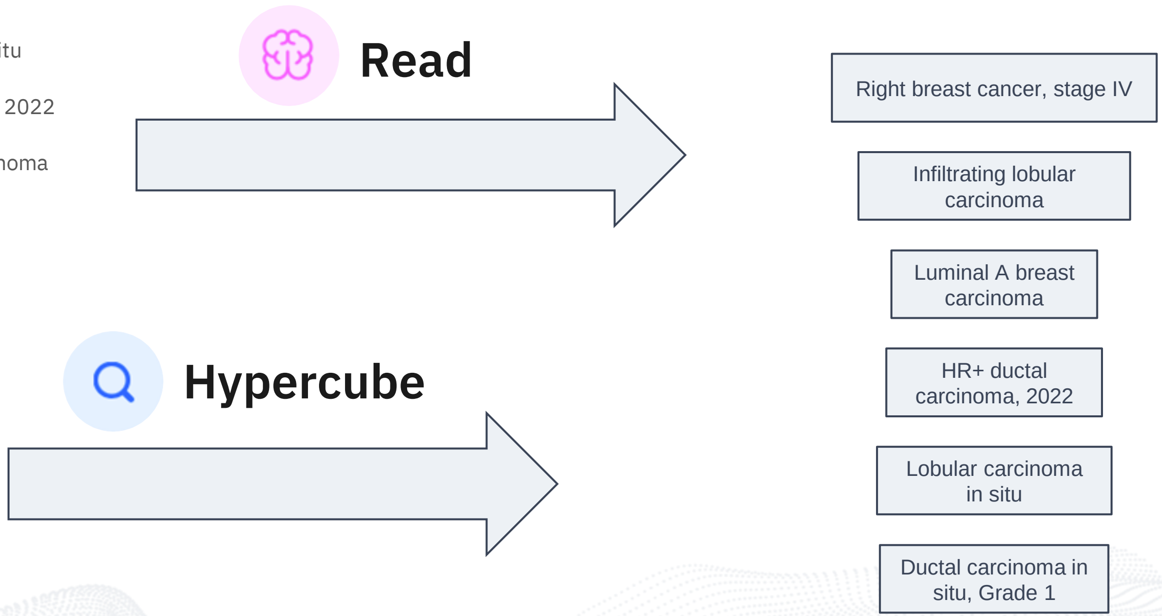
Example: Knowledge Consolidation from individual facts

Facts about this patient scattered across different documents:

- Right breast cancer, stage IV
- Infiltrating lobular carcinoma
- Ductal carcinoma in situ, Grade 1
- Lobular carcinoma in situ
- HR+ ductal carcinoma, 2022
- Luminal A breast carcinoma

Questions that cannot be answered without consolidation (No single fact has all the required information):

- Metastatic ductal carcinoma ✖
- Her2-negative lobular carcinoma ✖
- ER/PR+, early grade, breast cancer ✖
- Unilateral Luminal A breast cancer ✖



Example: Knowledge Consolidation from individual facts

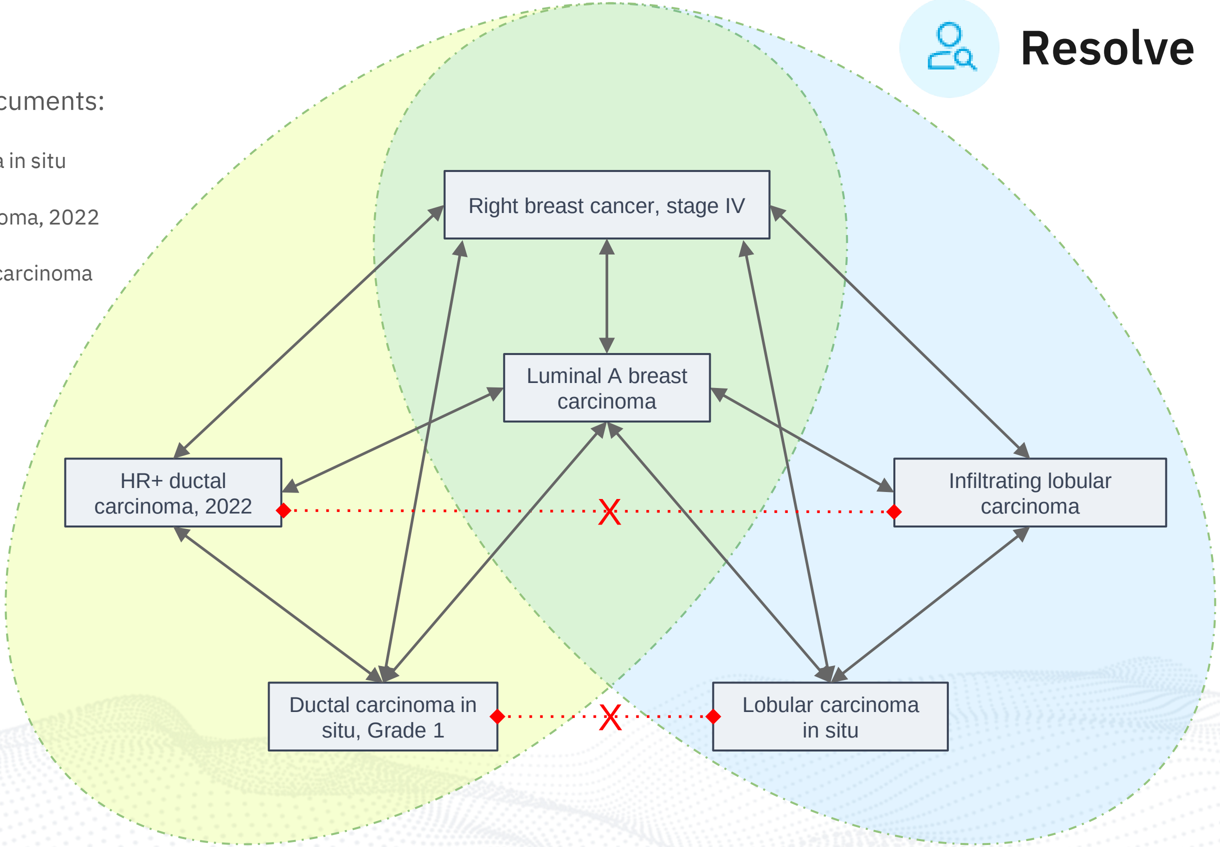
Resolve

Facts about this patient scattered across different documents:

- Right breast cancer, stage IV
- Infiltrating lobular carcinoma
- Ductal carcinoma in situ, Grade 1
- Lobular carcinoma in situ
- HR+ ductal carcinoma, 2022
- Luminal A breast carcinoma

Questions that cannot be answered without consolidation (No single fact has all the required information):

- Metastatic ductal carcinoma ✖
- Her2-negative lobular carcinoma ✖
- ER/PR+, early grade, breast cancer ✖
- Unilateral Luminal A breast cancer ✖



Example: Knowledge Consolidation from individual facts

MENDEL

Facts about this patient scattered across different documents:

- Right breast cancer, stage IV
- Infiltrating lobular carcinoma
- Ductal carcinoma in situ, Grade 1
- Lobular carcinoma in situ
- HR+ ductal carcinoma, 2022
- Luminal A breast carcinoma

Questions that cannot be answered without consolidation (No single fact has all the required information):

- Metastatic ductal carcinoma ✗ ✓
- Her2-negative lobular carcinoma ✗ ✗
- ER/PR+, early grade, breast cancer ✗ ✓
- Unilateral Luminal A breast cancer ✗ ✓

Resolve

Biomarkers: ER/PR+ HER2- (Luminal A)
Primary Site: Breast
Primary Site Laterality: Right
Morphology: Ductal Carcinoma
Differentiation: Grade 1 (Well Differentiated)
Stage Group: Stage IV
Date of Diagnosis: 2022

Primary Site: Breast
Morphology: Infiltrating lobular carcinoma

Hypercube

Evaluation set

MENDEL

Human-abstracted Gold Set:

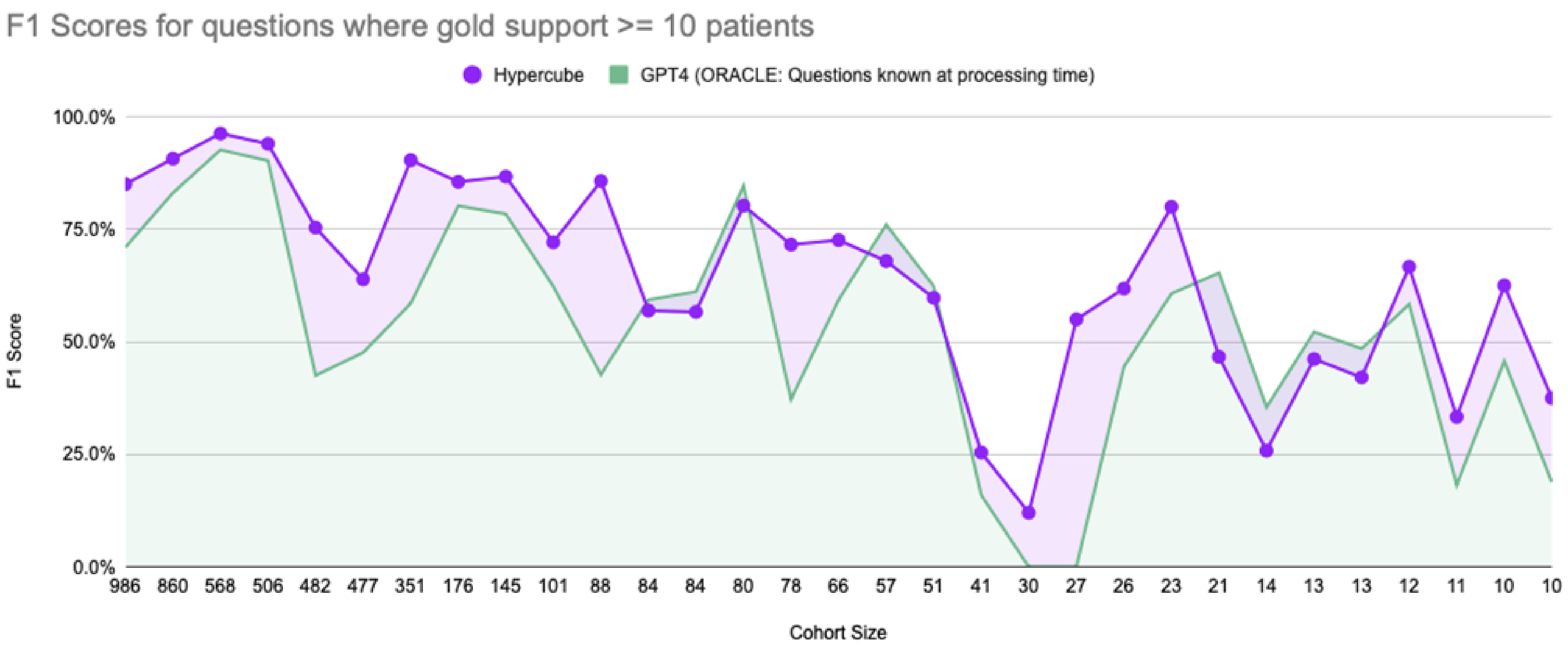
- **1,436 patients** from 4 sites
- 91K documents
- 10 of the top 14 cancer types included
- Average age at diagnosis, 64 years
- Random sample from contributing sites biased toward breast cancers (40% of patients with malignancy) and toward solid tumors

Large-Scale Reasoning: Evaluation

Evaluate on four dimensions:

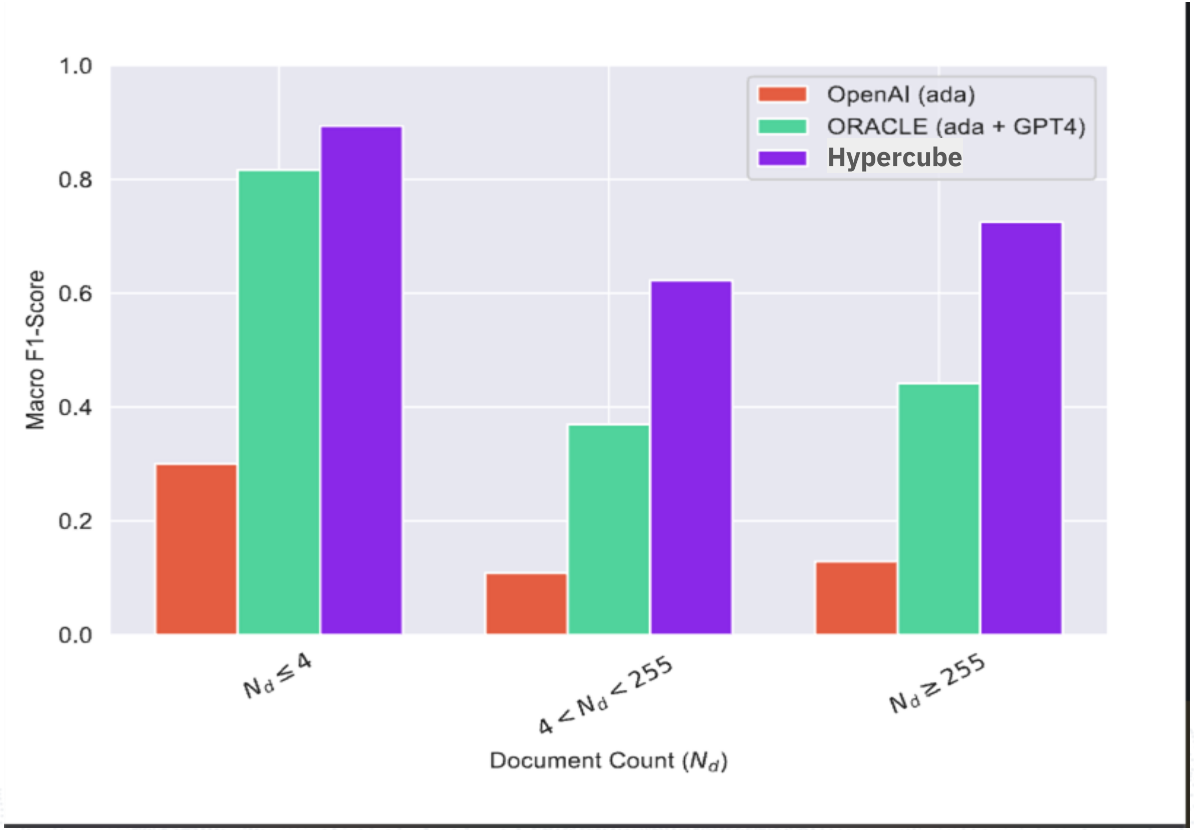
1. System's Quality
2. Effect of Longitudinality on Quality
3. System's Hallucinations Tendencies
4. System's Set Theoretic Consistency

Large-Scale Reasoning: Quality

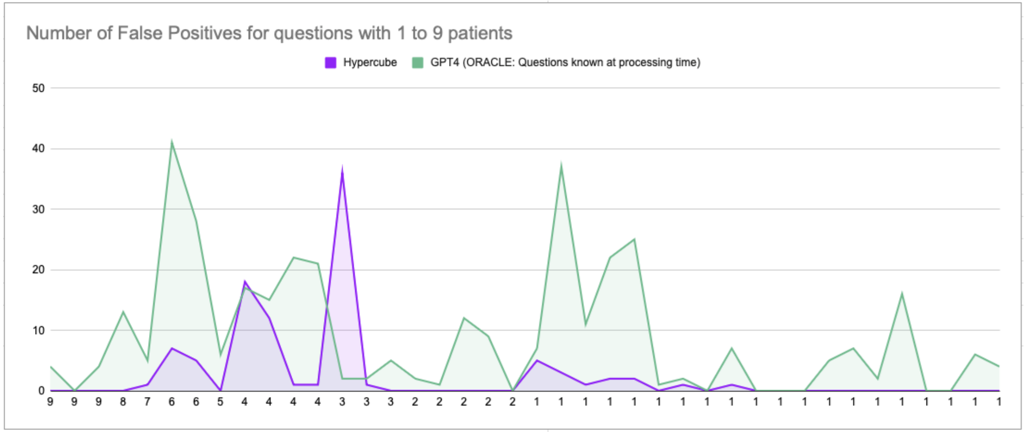


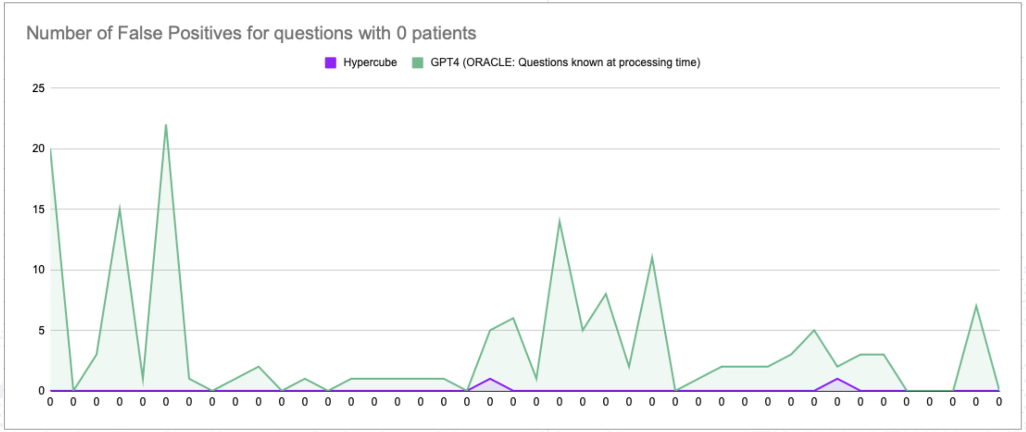
Large-Scale Reasoning: Longitudinality

MENDEL



Large-Scale Reasoning: Hallucinations





Large-Scale Reasoning: Set Theoretic Consistency

MENDEL

Questions	# Gold Answers	# HC Answers	# HC Answers NOT in Parent Q	# GPT Answers	# GPT Answers NOT in Parent Q
Find me patients with breast cancer	568	531		613	
breast cancer patients with carcinoma	506	492	0	547	22
breast cancer patients who received a breast cancer chemo except tamoxifin	351	375	0	196	10
Early stage breast cancer patients treated surgically other than mastectomies.	41	77	0	298	13

Find me patients receiving systemic therapy	986	1324	0	651	
Find me patients receiving targeted therapy	482	776	0	181	36
Find me patients treated with a TKI	57	102	0	64	24
Find me patients treated with an EGFR TKI	23	32	0	33	12
Find me patients treated with osimertinib	9	9	0	12	2
Find me patients treated with RxCUI code 1721560	9	9	0	0	0
Find me patients treated with Tagrisso	9	9	0	13	1
Tagrisso and any other EGFR inhibitor	4	5	0	25	18

Thank you.