

**“AI For Society Group”**

***Objective Approaches in a Subjective Medical World:***  
**Investigating AI Integrations in Healthcare**  
**Technologies**

**Yuxing Hao**

**IvyPlus Exchange Scholar @ MIT**

**Ph.D. Research Fellow @ Cornell, @ Mayo Clinic**

# Objective Approaches in a Subjective Medical World

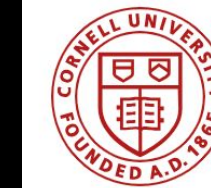
## *Subjective <> Objective Interaction* **Human-AI Interaction**

- i-SDM Framework: Clinical Decision Science [CSCW 23', CSCW 24']
- Patient-centered Shared Decision-Making Tool [CHI 24']
- Explainable AI [CSCW 24']

## *Improve Subjective w/ Objective Approaches* **Healthcare Intelligence**

- Clinical Decision Support Tool (DST) [Bioinformatics 20', CHI 23']
- Retrospective Comparison In-Basket Message Study: Human Care Team VS. GPT4o-based [Mayo Proceeding: Digital Med 25']
- Veterinary Precision Health [AAAI 24', IAAI 25' DAI Workshop]

# CHI 24: i-SDM Frameworks (AI-Based Shared Decision Making Tool)



## CORNELL CHRONICLE

Topics

Campus & Community

All Stories

In the News

Expert Quotes

Cornellians

### AI may improve doctor-patient interactions for older adults with cancer

By [Emily Groff](#), [College of Human Ecology](#)

May 14, 2024



CLINICAL DAILY NEWS

### Report: AI tool could boost communication for older adults with cancer

KRISTEN FISCHER



**Eli Goldring** ★★ • 3rd+

Executive Director of Business Development at Culinary Depot/Founder of the Hope Organization



































































































































**David Lubarsky**, this is a fascinating insight into how AI can bridge the gap for older adults with cancer. It's heartening to see technology being leveraged to empower patients in their treatment decisions, especially for those who might struggle with tech literacy. This could truly transform patient-clinician communication.

Like | Reply

1. **Y. Hao**, Z. Liu, B. Riter, S. Kalantari. Advancing Patient-Centered Shared Decision-Making with AI Systems for Older Adult Cancer Patients. In Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11--16, 2024, Honolulu, HI, USA. <https://doi.org/10.1145/3613904.3642353>

2. **Y. Hao**, C. Löckenhoff, H. Lee, J. Zwerling, S. Kalantari. The i-SDM Framework: Developing AI-based Tools in Shared Decision-Making for Cancer Treatment with Clinical Professionals. Companion of the 2024 Computer-Supported Cooperative Work and Social Computing (CSCW 24'). <https://doi.org/10.1145/3678884.3681841>

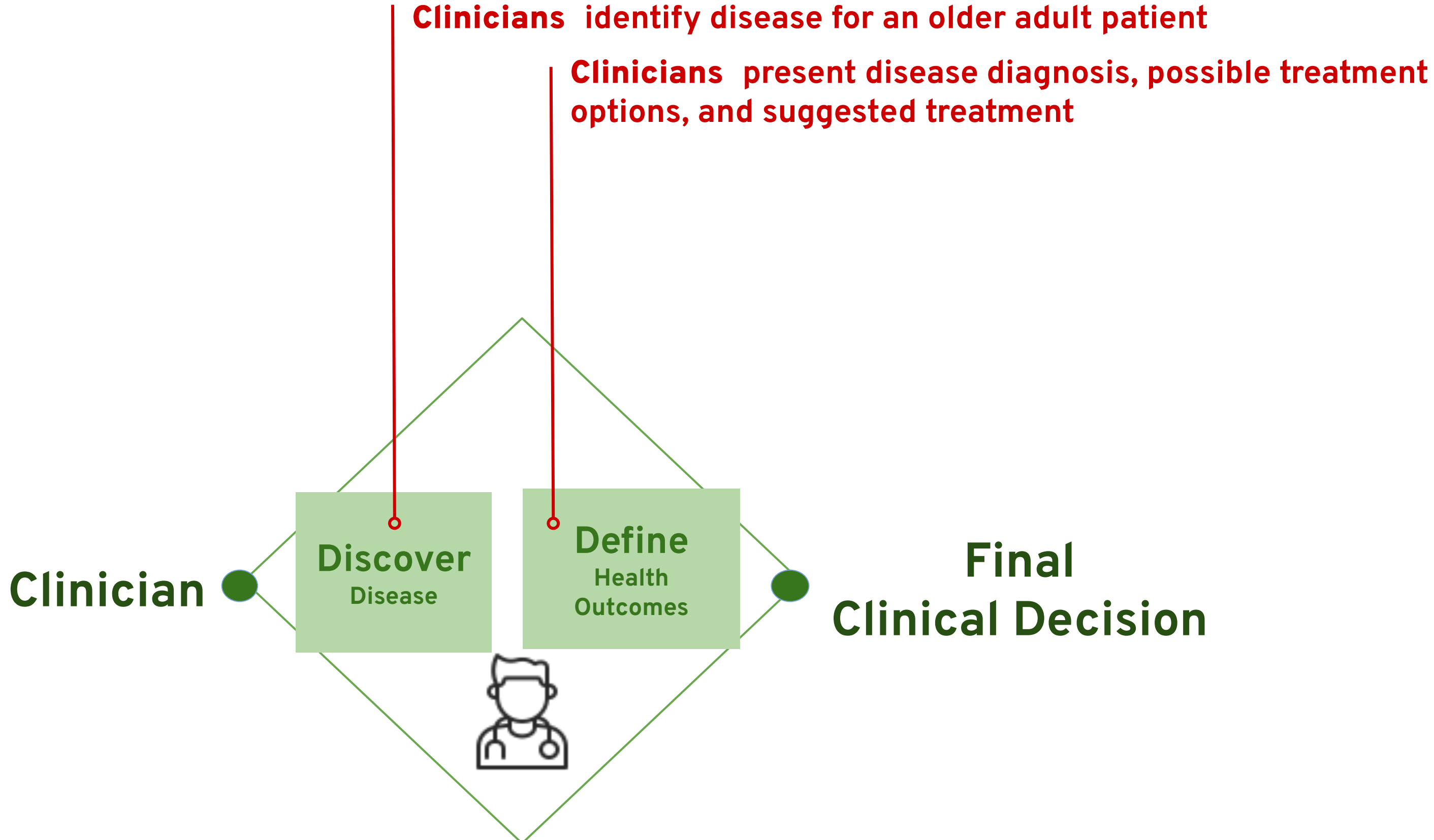
# Shared Decision Making (SDM) Preliminary Ten Important Factors Selections with 12 Participants (7 Patients + 5 Clinicians)

Total Participants' Approved Factors (out of 12)	12	12	11	10	7	7	6	6	4	3
Clinician Participants (C1 - C5)										
										
										
										
										
Patient Participants (P1 - P7)										
										
										
										
										
										
										
										
Ten Important factors for SDM (From high to low)	Survival Rates in 5 years	Potential Risk	Alternative Treatment Options	Average Patient Age for the Treatment	Distance (Treatment Location)	Detailed Treatment Understanding	Total Treatment Duration	Pain Degree	Nursing Service	Treatment Fee

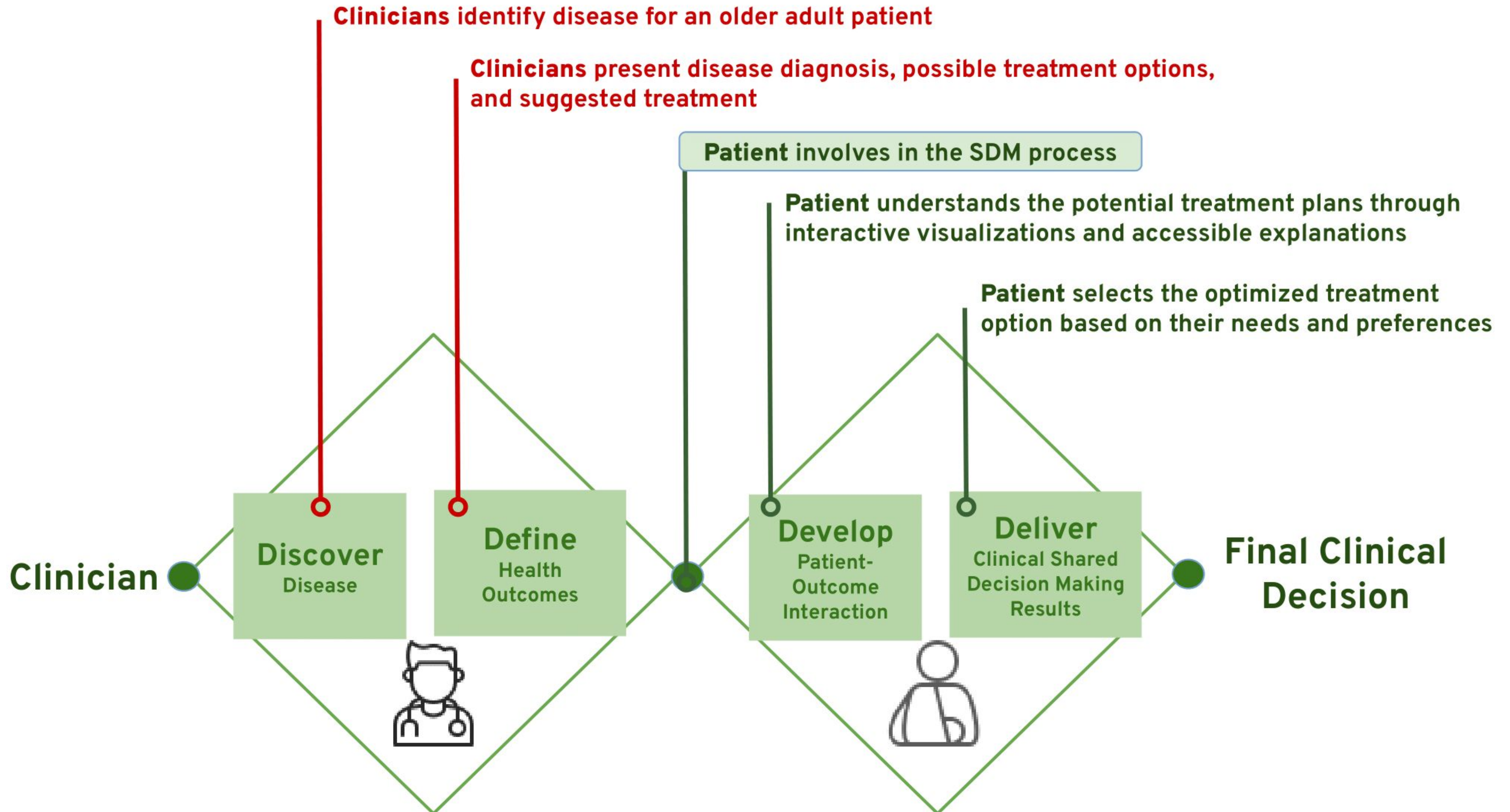
 Patient Participant Approved  
  Patient Participant Felt This Option Indifference  
  Clinician Participant Approved  
  Clinician Participant Felt This Option Indifference



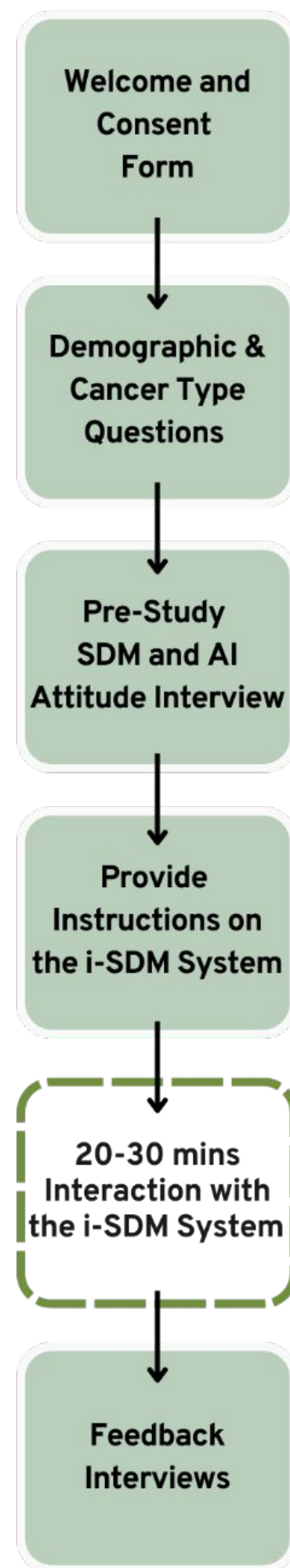
# Single-Diamond Framework for Patient-Centered SDM process



# Double-Diamond Framework for Patient-Centered SDM process



# i-SDM Step-wise Process



### 1 Stage 1: Patient Assessment

#### Step 1 Factor Selections

You selected:

Urinary Incontinence	Erectile Dysfunction	Survival Rates	Select
Location of Treatment	Duration of Treatment	Service of Nursing	

- Location of Treatment
- Duration of Treatment
- Service of Nursing
- Urinary Incontinence
- Erectile Dysfunction
- Pain Degree
- Average Treatment Age
- Facility Quality
- Survival Rates
- Medical Cost

#### Step 2 Basic Information

**John Doe**

Date of Presentation: 12 / 15 / 2018  
 Age: 62 Gender: M  
 Ethnicity: Caucasian  
 Employment Statute: Retired  
 Cancer: Prostate Cancer  
 Gleason Score: 7

Full Description

Summary

### 2 Stage 2: Risk Evaluation with AI

#### Step 3 Survival Rates

Your Predicted Survivals in Five Years Prediction Length: 5 Years

Treatment	Survival Rate
Surgery	94-98%
Radiation	93-98%
Active Surveillance	78-96%
No Treatment	33-50%

#### Step 4 Side Effects

You may have Erectile Dysfunction

You may have Urinary Incontinence

Others

### 3 Stage 3: Patient End Decision

#### Step 5 Other Factors

- Location of Treatment
- Time of Treatment
- Service of Nursing
- Others

Time of Treatment (Days)

Surgery	3	305
Radiation	110	305
Active Surveillance		305

#### Step 6 Finalize Decisions

SIDE EFFECTS

Your Predicted Survivals in Five Years

Your Choice

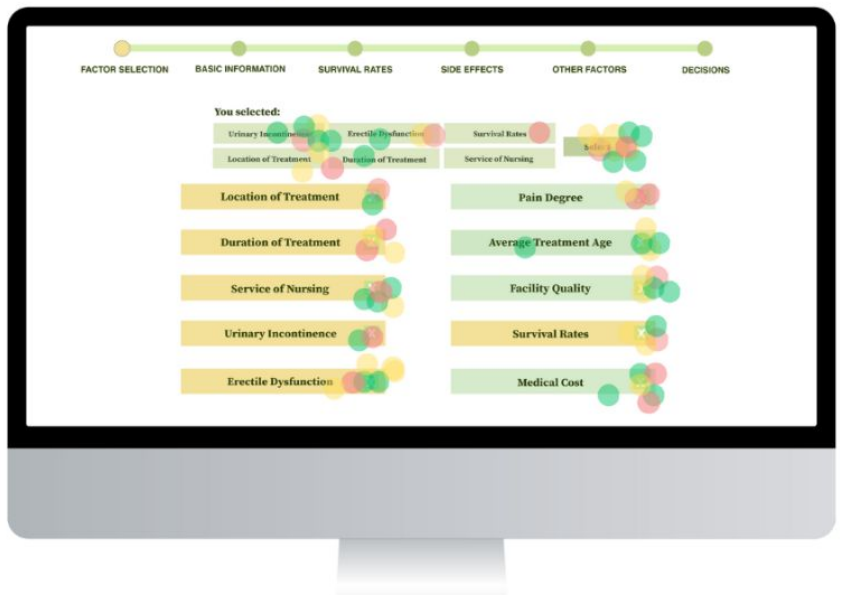
Submit



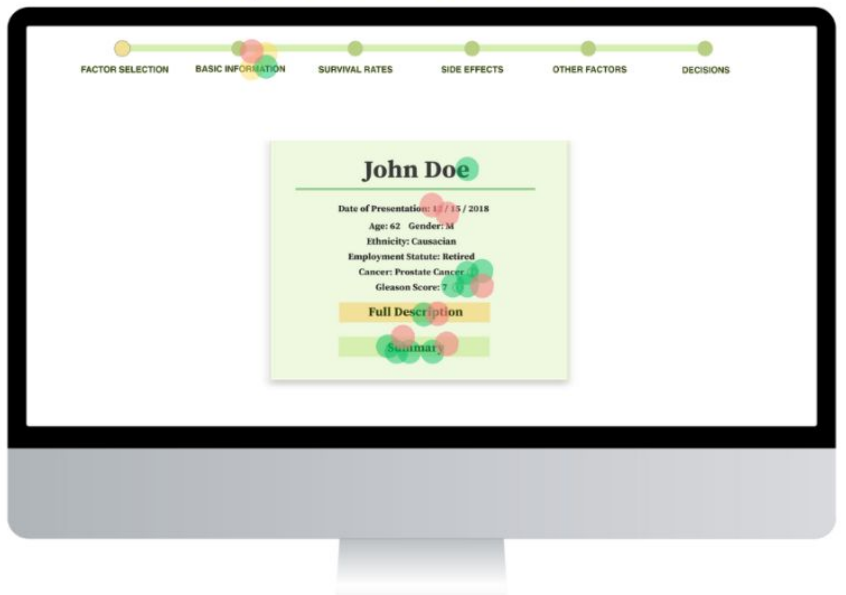
# i-SDM Prototype's Heatmap

- Negative Verbal Cues
- Positive Verbal Cues
- Neutral Verbal Cues

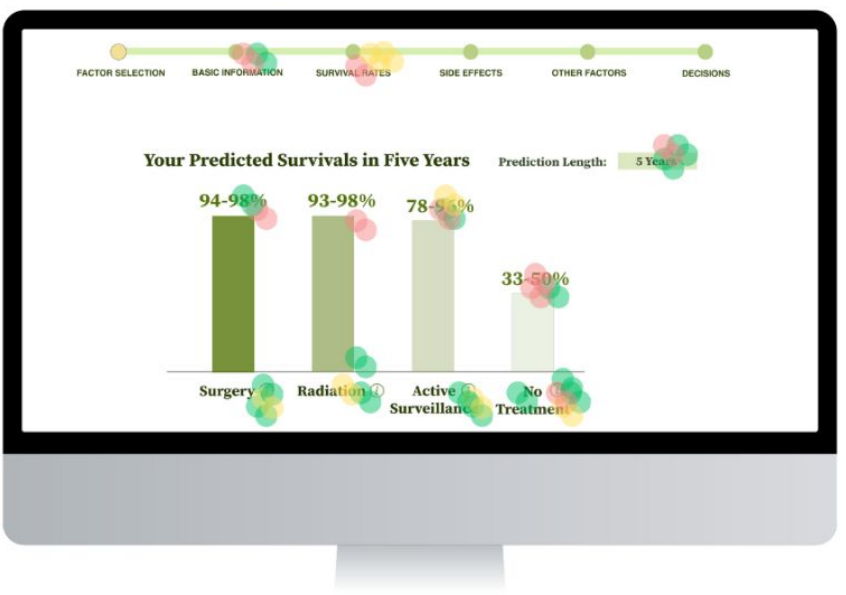
Step 1 Factor Selections



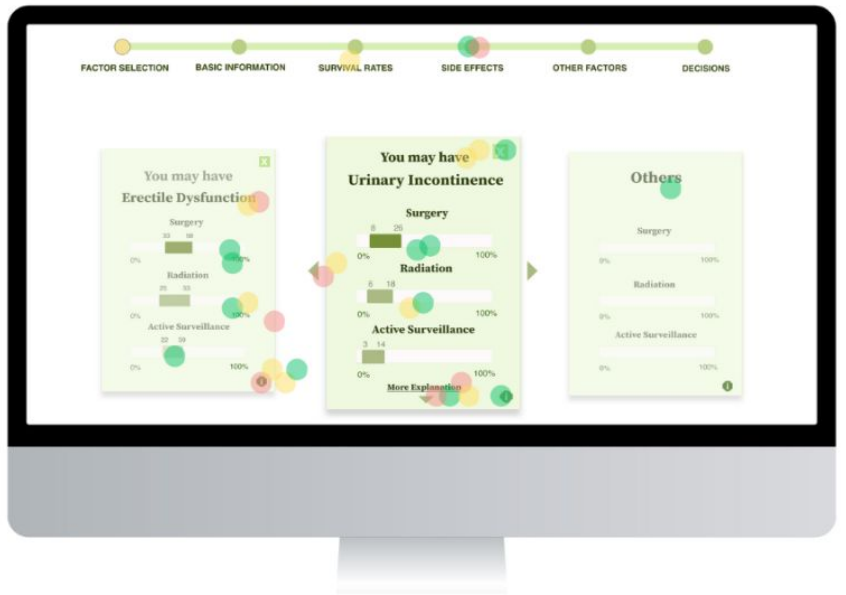
Step 2 Basic Information



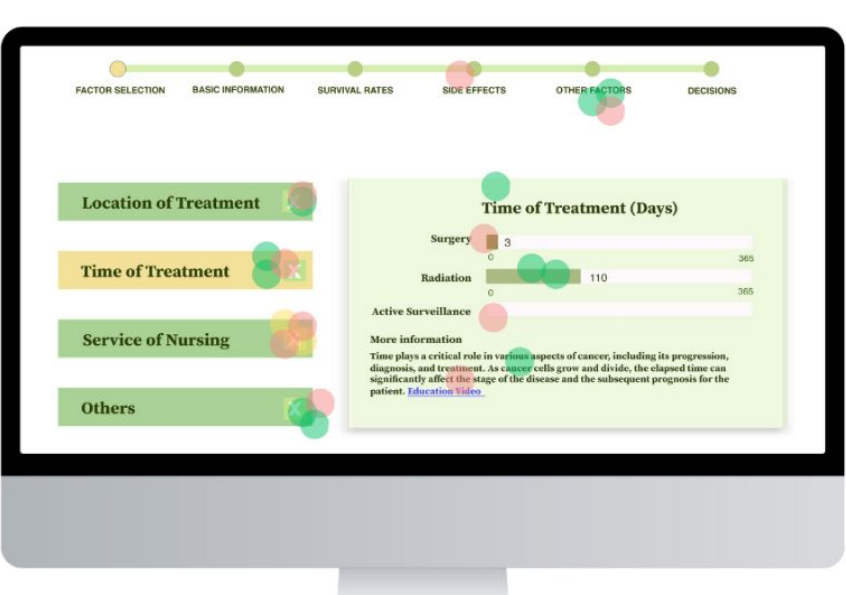
Step 3 Survival Rates



Step 4 Side Effects



Step 5 Other Factors



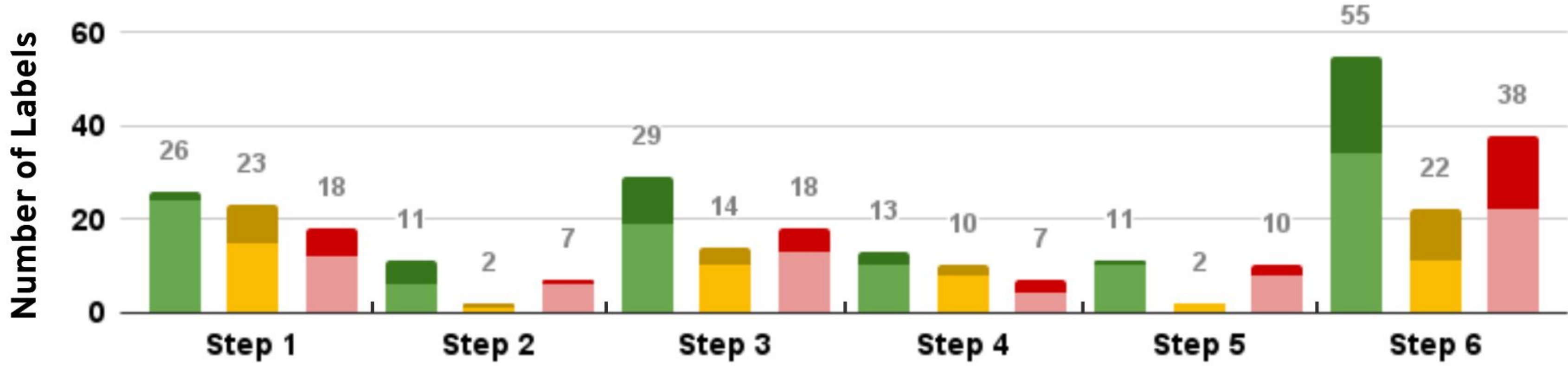
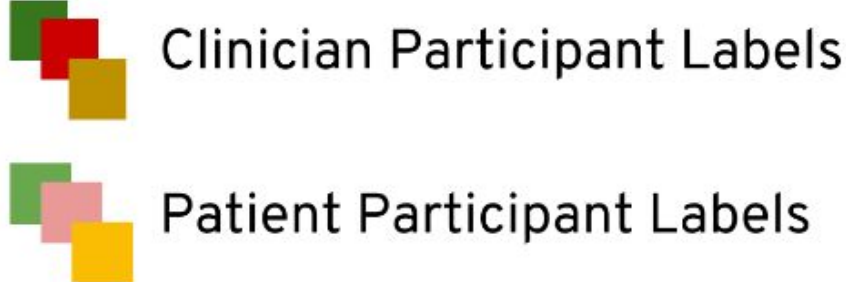
Step 6 Finalize Decisions





# Cue Labels Analysis across Steps

316 labels, comprising 145 positive cues labels (45.89%), 73 neutral cues labels (23.10%), and 98 negative cues labels (31.01%).



# i-SDM Prototype's Heatmap for Evaluating i-SDM Prototype

- Negative Verbal Cues
- Positive Verbal Cues
- Neutral Verbal Cues



# Reason More != Perform Better?

Investigating the *Trade-Off* Between

**Depth of Reasoning** and **Performance**  
in **LLMs' Medical AI Reasoning**



E.g. Iterations, Inference time,  
# of step-wise process

E.g. Accuracy, Confidence,  
Error propagation (overfitting)



# PubMedQA

A Dataset for Biomedical Research Question Answering

## Question:

Do preoperative statins reduce atrial fibrillation after coronary artery bypass grafting?

## Context:

**(Objective)** Recent studies have demonstrated that statins have pleiotropic effects, including anti-inflammatory effects and atrial fibrillation (AF) preventive effects [...]

**(Methods)** 221 patients underwent CABG in our hospital from 2004 to 2007. 14 patients with preoperative AF and 4 patients with concomitant valve surgery [...]


**(Results)** The overall incidence of postoperative AF was 26%. **Postoperative AF was significantly lower in the Statin group compared with the Non-statin group (16% versus 33%,  $p=0.005$ ).** Multivariate analysis demonstrated that independent predictors of AF [...]

## Long Answer:

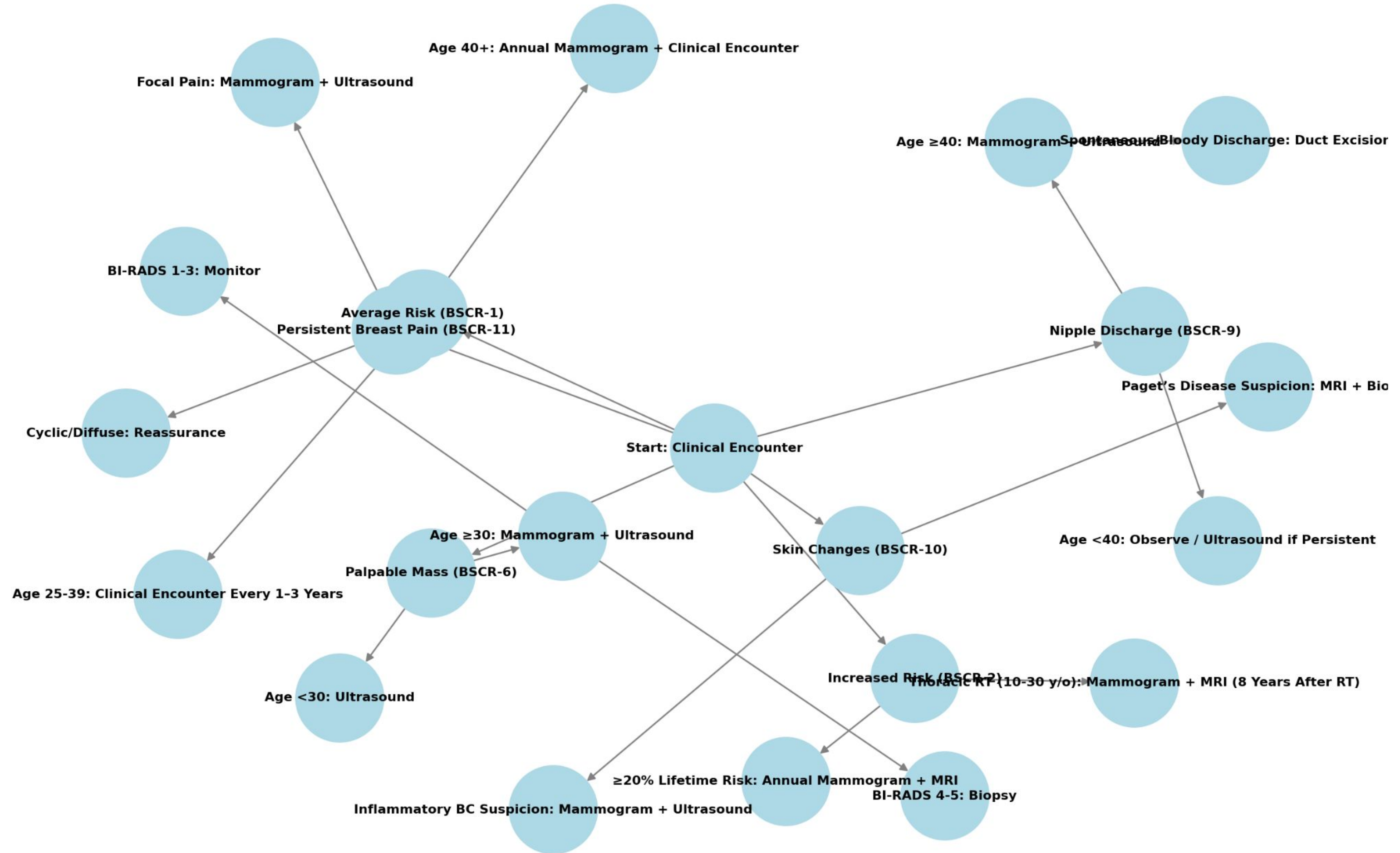
**(Conclusion)** Our study indicated that preoperative statin therapy seems to reduce AF development after CABG.

**Answer:** yes

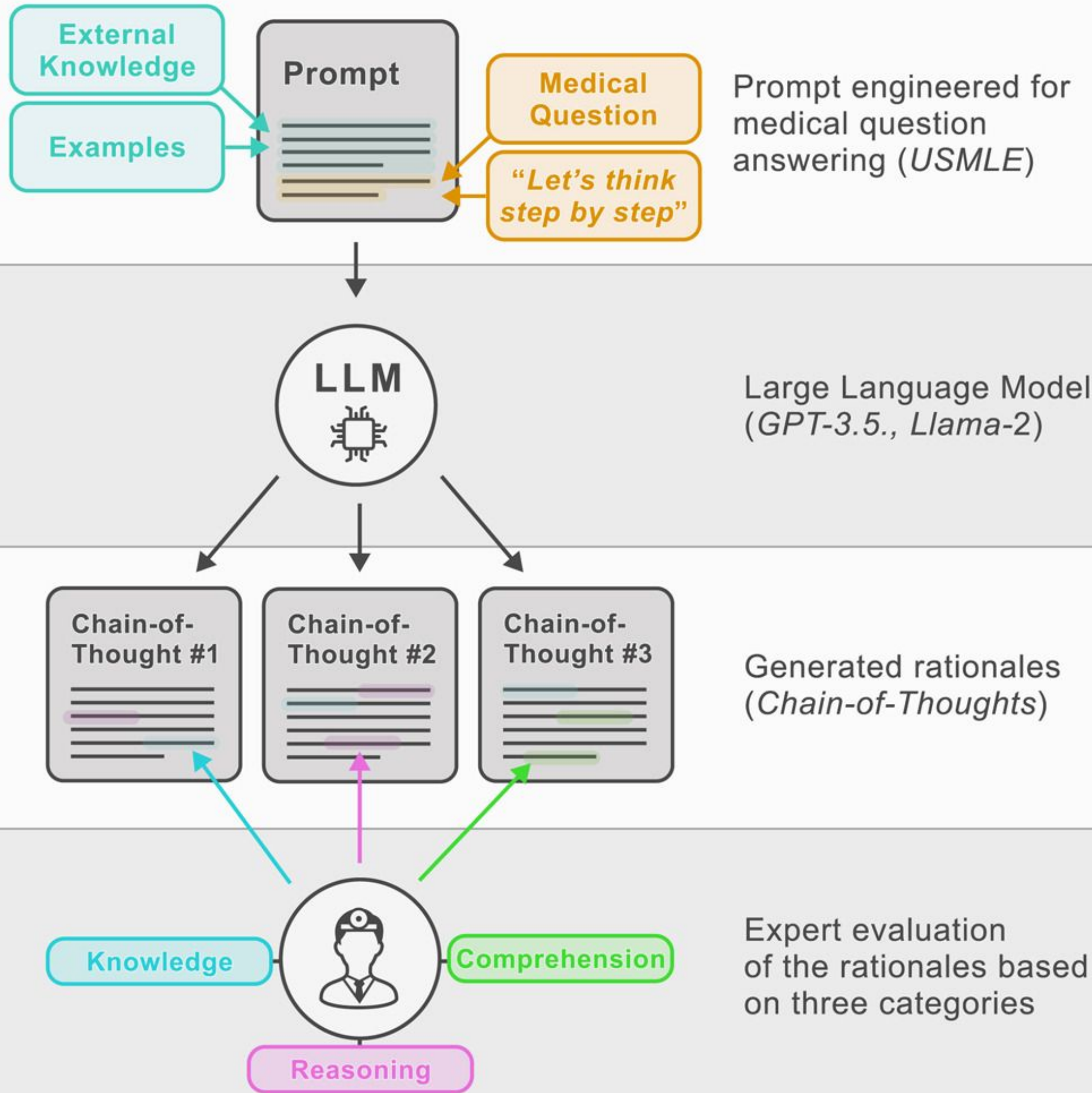
## Leaderboard (reasoning-required setting)

	Model	Code	Size	Accuracy (%)	Macro-F <sub>1</sub> (%)
1 Nov 28, 2023	GPT-4 (Medprompt) Microsoft <a href="#">(Nori et al. 2023)</a>		NA	82.0	NA
2 May 16, 2023	Med-PaLM 2 Google Research & DeepMind <a href="#">(Singhal et al. 2023)</a>		NA	81.8	NA
3 Nov 27, 2023	MEDITRON EPFL <a href="#">(Chen et al. 2023)</a>		70B	81.6	NA
4 Jul 6, 2023	Palmyra-Med Writer Inc. <a href="#">(Kamble et al. 2023)</a>		40B	81.1	NA

# Breast Cancer Screening Decision Tree







# Using Chain-of-Thoughts

Liévin, Valentin, et al. "Can large language models reason about medical questions?." *Patterns* 5.3 (2024).

*"Based on detailed step-by-step answers generated by LLMs, we show that LLMs can often answer challenging medical questions by mobilizing expert knowledge and advanced reasoning capabilities."*



USMLE [test:731] Three days after induction chemotherapy, a 9-year-old boy is admitted to the hospital for fever and neutropenia. He was initially diagnosed with t(9;22)-positive acute T cell lymphoblastic leukemia following workup for fatigue and pancytopenia. He is fully conscious. His blood pressure is 110/65 mm Hg, temperature is 39.5°C (103.1°F), pulse is 98/min, and respiratory rate is 14/min. Physical examination including examination of the central venous catheter site shows no source of infection. Broad-spectrum antibiotics are initiated. Which of the following factors about this patient most strongly indicates a poor prognosis?

A) Age **B) Cytogenetics** C) Fever and neutropenia D) Initial presentation

✓ Let's use step by step inductive reasoning, given the medical nature of the question.

The first thing that stands out is that the patient is 9 years old. In general, children have a better prognosis than adults when it comes to leukemia.

The next thing to consider is the patient's cytogenetics. The fact that the patient has t(9;22)-positive acute T cell lymphoblastic leukemia indicates that the patient has a specific chromosomal abnormality that is associated with a poor prognosis.

The next thing to consider is the patient's fever and neutropenia. Fever and neutropenia are both signs of infection, and infection is a serious complication in leukemia patients, particularly in those who are immunocompromised. Finally, we need to consider the patient's initial presentation. The fact that the patient was initially diagnosed with leukemia following workup for fatigue and pancytopenia indicates that the patient's leukemia was advanced at the time of diagnosis.

All of these factors indicate a poor prognosis for this patient.

Therefore, among A through D, the answer is B, cytogenetics.

✗ Let's think step by step like a medical expert. The first step is to gather information about the patient. In this case, we are given the following information: -The patient is 9 years old -The patient has t(9;22)-positive acute T cell lymphoblastic leukemia -The patient is currently experiencing fever and neutropenia

The next step is to analyze the information and look for any patterns or clues that could help us answer the question. In this case, the most important factor that indicates a poor prognosis is the patient's age. Children under the age of 10 have a much higher chance of relapsing after treatment and a lower chance of achieving remission.

Therefore, among A through D, the answer is A.

✗ Let's derive the differential diagnosis step by step:

- 1) The patient has fever and neutropenia.
- 2) The patient is 9 years old.
- 3) The patient has t(9;22)-positive acute T cell lymphoblastic leukemia.
- 4) The patient's initial presentation was fatigue and pancytopenia.

The most likely cause of the patient's fever and neutropenia is infection. The patient's age, cytogenetics, and initial presentation are all risk factors for infection.

Therefore, the factor that most strongly indicates a poor prognosis is the patient's age. Therefore, among A through D, the answer is A.

Liévin, Valentin, et al. "Can large language models reason about medical questions?." *Patterns* 5.3 (2024).

Pattern A	Correct reasoning step	Pattern D	Incorrect reasoning step
Pattern B	Correct recall of knowledge	Pattern E	Incorrect or insufficient knowledge
Pattern C	Correct reading comprehension	Pattern F	Incorrect reading comprehension



# Samples

## Question

Do LLM generate correct answer without reasoning?

A 55-year-old male patient, with a smoking history of 60 pack-years, has had a chronic cough for over 10 years. He reports that about three months ago, he noticed the presence of blood in the sputum. He also mentions a weight loss of about 15% of his usual weight during the same period, anorexia, weakness, and night sweats. The chest X-ray taken at the time of the consultation is shown below. What is the most likely diagnostic hypothesis in this case?

- A) Pulmonary aspergillosis.
- B) Lung carcinoma.**
- C) Cavitory tuberculosis.
- D) Bronchiectasis with infection.
- E) Chronic obstructive pulmonary disease.

Answer is right, but reasoning is (partially) wrong.

Hemoptysis (blood in sputum) is commonly seen in tuberculosis or fungal infections rather than cancer. ✘ (Incorrect: While TB can cause hemoptysis, lung cancer is a leading cause as well.)

Despite the radiologic and symptom findings suggesting infection, the most likely diagnosis is lung carcinoma due to the patient's smoking history and progressive systemic symptoms. ✔ (Correct final answer, but flawed reasoning.)

**Final Answer: B - Lung Carcinoma**

Answer is wrong, as well as reasoning.

The long history of chronic cough indicates a latent tuberculosis infection that has reactivated. ✘ (Incorrect: While TB can present with a chronic cough, the patient's history suggests a different etiology.)

Weight loss, anorexia, and night sweats are classic B-symptoms of TB, making it the most likely diagnosis. ✘ (Incorrect: These symptoms overlap with lung cancer and must be evaluated in context.)

**Final Answer: C - Cavitory Tuberculosis**

# Clinicians in Medical Decision-Making

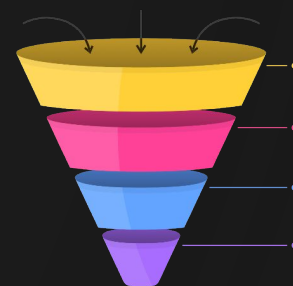
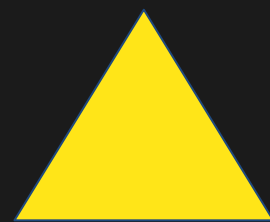
## Current Situation: Input → Output

Electronic Health Records  
Clinical Notes  
Exam Results



Clinical Decision

## Better Solution: Input → *Step-Wise Reasoning Process* → Output



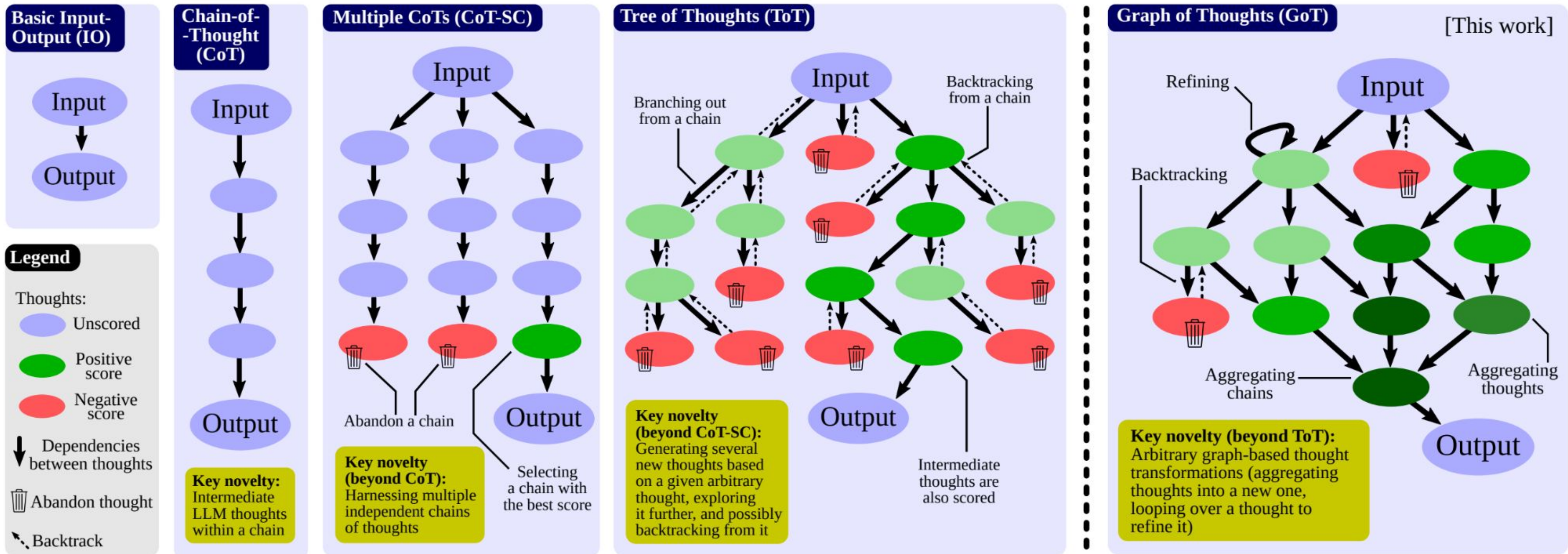




$$\text{Logit}(P) = \beta_0 + \beta_1(\text{Smoking history}) + \beta_2(\text{Hemoptysis}) + \beta_3(\text{Weight loss}) + \beta_4(\text{Age}) + \dots$$

Where each **β coefficient** represents the contribution of that factor to the likelihood of lung cancer.

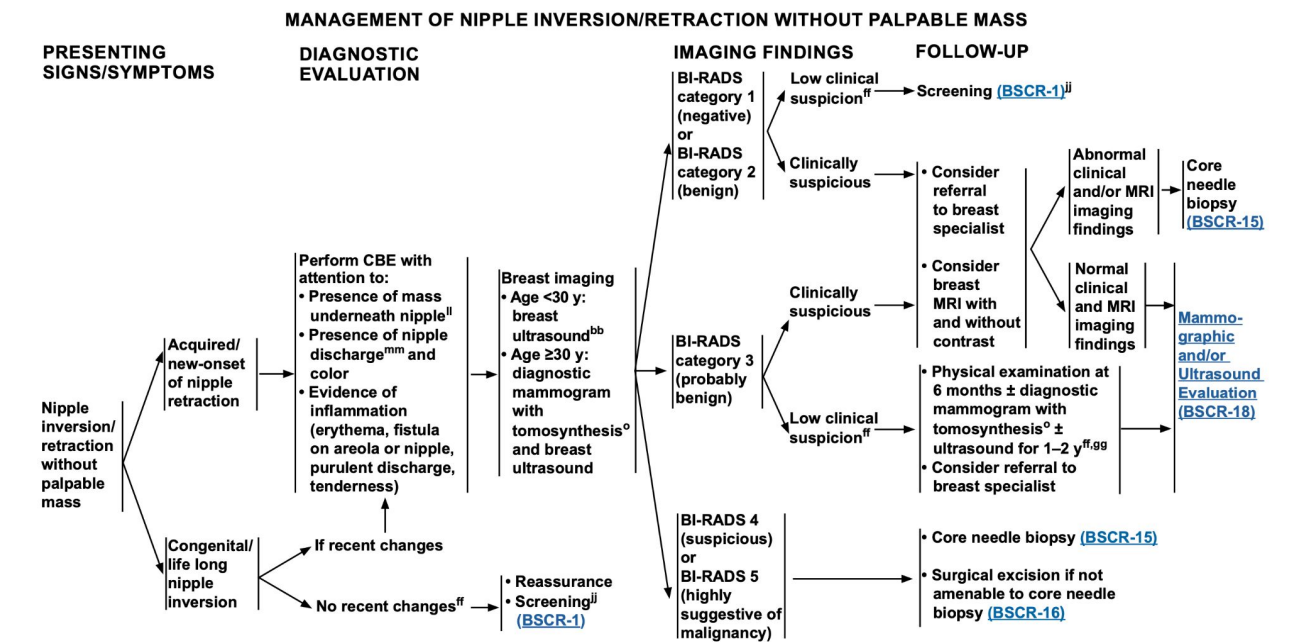
Feature	Condition Present?	LR <sup>+</sup> (Estimate)	Contribution to Diagnosis
Smoking (60 pack-years)	Yes	10	Strong
Chronic cough (10+ years)	Yes	3	Moderate
Hemoptysis (Blood in sputum)	Yes	6	Strong



1. Besta, Maciej, et al. "Graph of thoughts: Solving elaborate problems with large language models." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 38. No. 16. 2024.
2. Yao, Shunyu, et al. "Tree of thoughts: Deliberate problem solving with large language models." *Advances in neural information processing systems* 36 (2023): 11809-11822.



# Based on NCCN guidelines, generate hypothetical case. For example:



## Case 1: Benign Low-Risk Presentation

A 32-year-old female presents with new-onset nipple retraction in her right breast. She denies any associated symptoms such as pain, discharge, or inflammation. On clinical breast examination, no palpable mass is detected, and there are no visible signs of erythema or tenderness. Given her age, she undergoes breast imaging with a diagnostic mammogram using tomosynthesis along with an ultrasound. The imaging results are classified as BI-RADS category 2, indicating a benign finding. As there are no concerning features, she is reassured and advised to continue with routine breast cancer screening as per guidelines.

**What is the most appropriate next step in the management of this patient?**

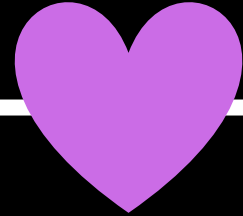
- A) Immediate biopsy to rule out malignancy
- B) Reassurance and continued routine breast screening
- C) MRI with contrast to further assess the nipple inversion
- D) Referral for surgical excision of the affected nipple

**Correct Answer:** B) Reassurance and continued routine breast screening



No LLM

With LLM



"LLM generates lots of errors"

**Errors**

"LLM identifies more details and mitigates errors"

"GPT4 is great but may not apply to prostate cancer domain"

**Efficacy**

"LLM accelerates clinical workflow"

"LLM creates human-like responses but not fact-checked"

**User-Centered**

"LLM is very informative, from a patient-centered manner"

"LLM cannot afford liability"

**Trust**

"LLM provides all the details in a timely fashion, which makes it reliable"

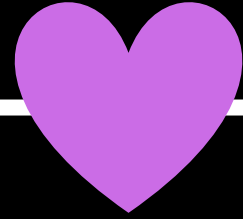
"LLM has no standardized evaluation metrics now"

**Evaluation**

"LLM is emerging and the evaluations should be adapted at this stage"

No LLM

With LLM



Ben Shneiderman's Book  
*Designing the User Interface* (1986):

Human  
Control

Computer  
Automation





**Thanks to all of my collaborators!**

**\* Cornell affiliations unless otherwise specified**



**Saleh Kalantari,**



**Aditya Vashistha,**



**Corinna Loeckenhoff,**



**Ken Birman,**



**Parminder Basran,**



**Matthias Wieland**



**Drew Margolin,**



**Lichao Sun,  
(Lehigh Univ)**



**Liu Wei,  
(Mayo Clinic)**



**Ryan Ries,  
(Cornell Tech)**



**Monika Safford,  
(Weill Cornell Medicine)**



**Qian Yang**



**Robert Riter,**



**Yuting Yang,**



**Tiancheng Yuan,**



**Aarushi Gupta,**



**Yiran Zhao,**



**Zhiwen Qiu,**



**Zeyu Liu,**



**Kexin Quan  
(UIUC)**



How might objective approaches **reshape the future** of personalized medicine?

What are the **limitations** of relying solely on objective data in healthcare decisions?

How can objective data **improve patient outcomes** in a field often driven by subjective judgment?

## **Objective** Approaches in a **Subjective** Medical World

[yuexing@mit.edu](mailto:yuexing@mit.edu) / [yh727@cornell.edu](mailto:yh727@cornell.edu) / [hao.yuexing@mayo.edu](mailto:hao.yuexing@mayo.edu)

IvyPlus Exchange Scholar @ MIT

Ph.D. Research Fellow @ Cornell, @ Mayo Clinic