



VECTOR
INSTITUTE

INSTITUT
VECTEUR

Strategic Information Gathering in Legal AI

A Dempster-Shafer Framework for
Underspecified Queries

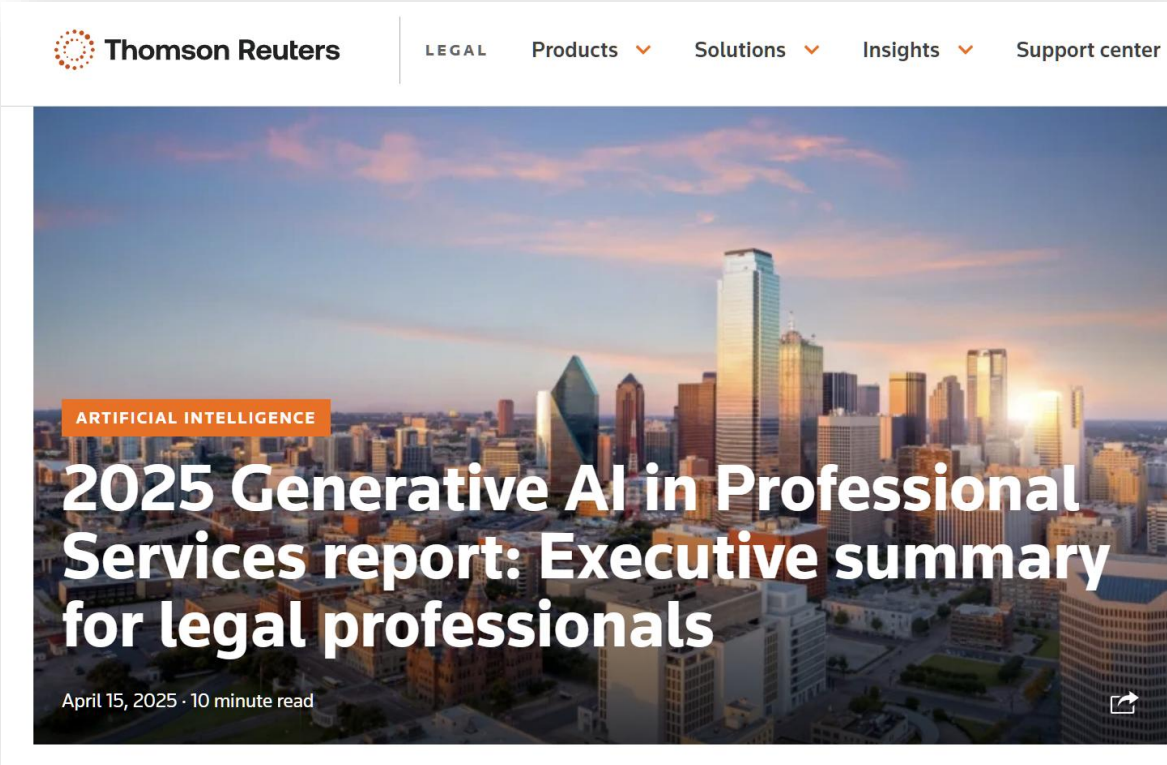
Maksym Taranukhin
Bloomberg Law Symposium

June 9, 2026

AI has the potential to make legal research more efficient

26% already using GenAI

But GenAI hallucinates




<https://legal.thomsonreuters.com/blog/genai-report-executive-summary-for-legal-professionals-tri/>


<https://www.theguardian.com/world/2024/feb/29/canada-lawyer-chatgpt-fake-cases-ai>

Underspecified Queries

RAG System



My flight was delayed and the airline refused to compensate me. What should I do?




If an airline refuses to compensate for a delay, formally submit a written complaint to the airline demanding payment, citing regulations like the Canadian APPR (if applicable). If denied again, file a complaint with the appropriate regulatory body (e.g., [CTA](#) in Canada or relevant European consumer protection). Gather all evidence: boarding passes, delay notifications, and expenses.

Steps to Take Following a Refused Compensation Claim:


- **Submit a Formal Written Claim:** Do not rely on verbal refusals at the airport. Send a formal letter or email to the airline's customer service detailing your flight, the delay length, and why the delay was within their control.
- **Understand Your Rights:** Compensation is generally required if the delay was within the airline's control (e.g., crew scheduling, maintenance) and you arrived at your final destination 3+ hours late.
- **File a Regulatory Complaint:** If the airline fails to respond or denies the claim unjustly, file a complaint with the proper authority.
 - **Canada:** File a complaint with the Canadian Transportation Agency (CTA).
 - **EU:** Contact the local [National Enforcement Body](#) if traveling within or to the EU.

[...]


InfoGatherer




My flight was delayed and the airline refused to compensate me. What should I do?




How long was the delay?



5 hours




What was the reason the airline provided for the delay?



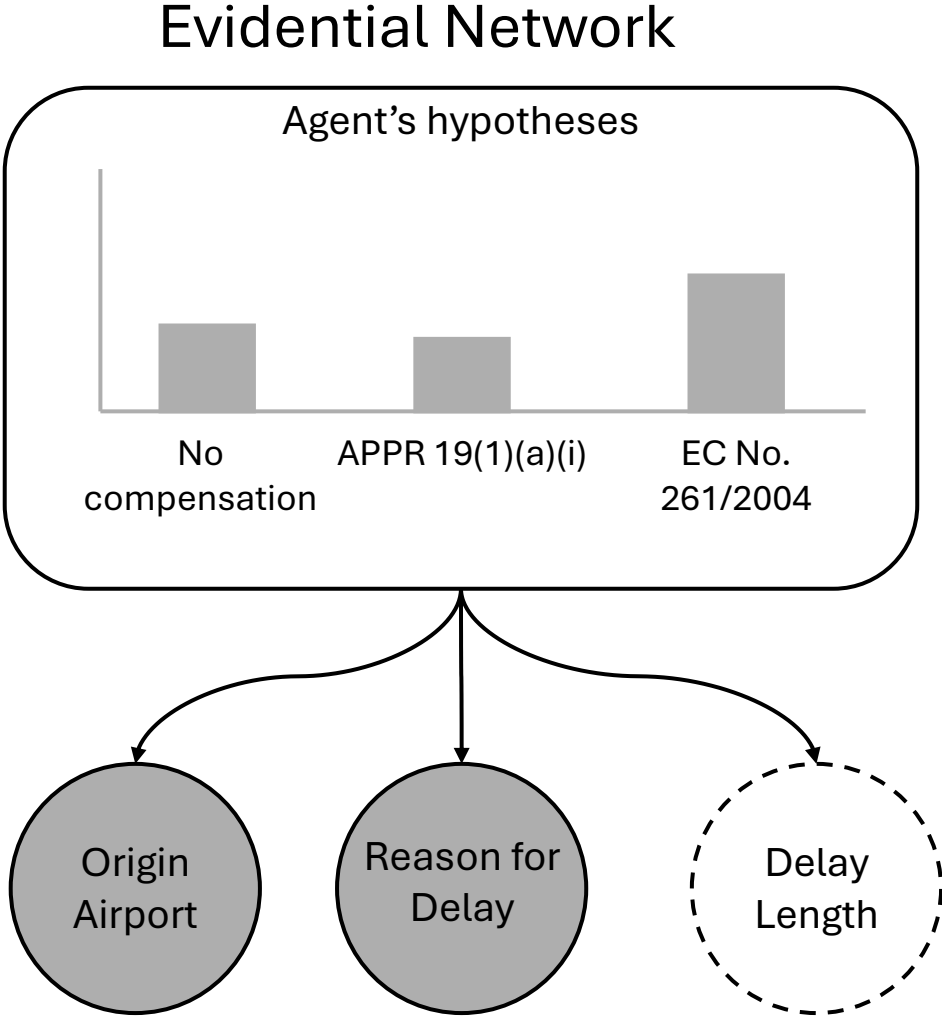
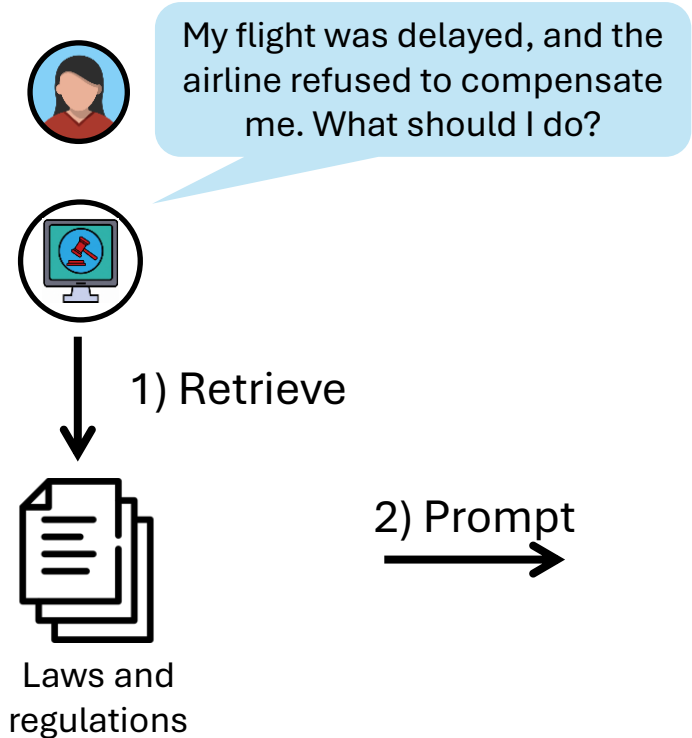
It was aircraft maintenance

Hypotheses



Contextual Documents

Evidential Network Construction

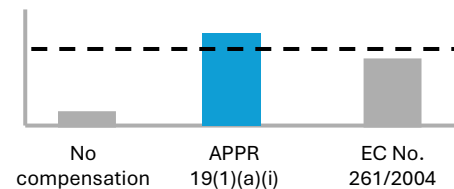
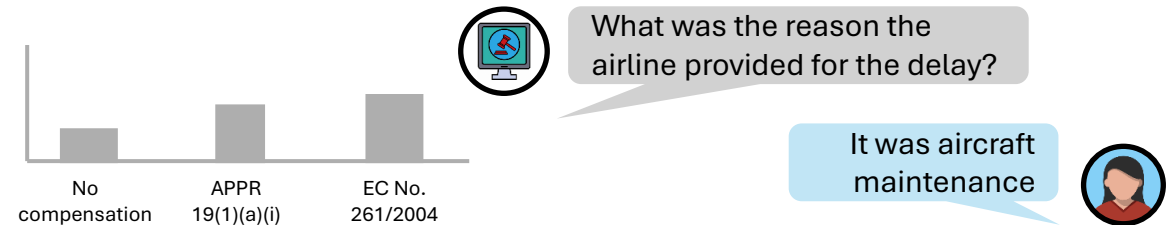
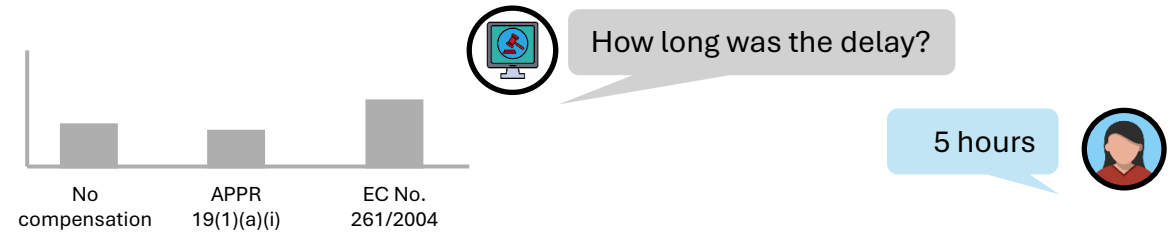


Belief Update via Strategic Questioning

Next question selection

$$q^* = \arg \max_q \mathbb{E}[\Delta E_d(q)]$$

Deng entropy = Non-specificity - Discord

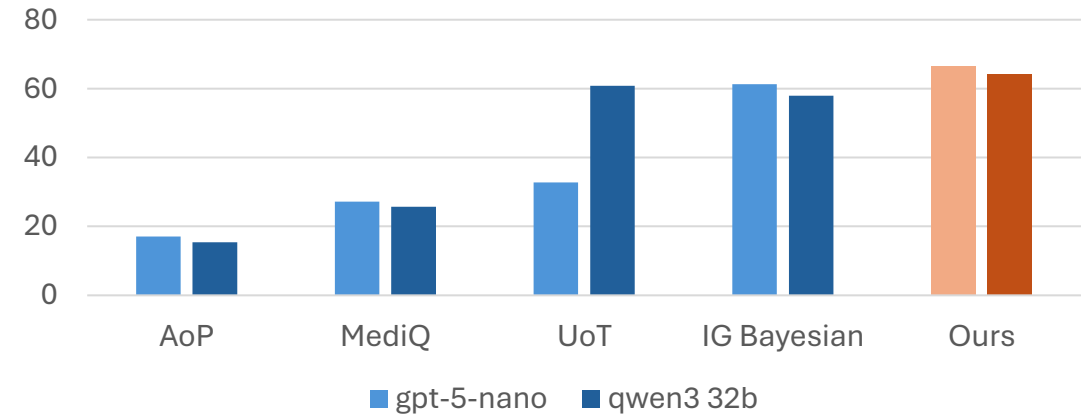



**Stopping
Condition Met**

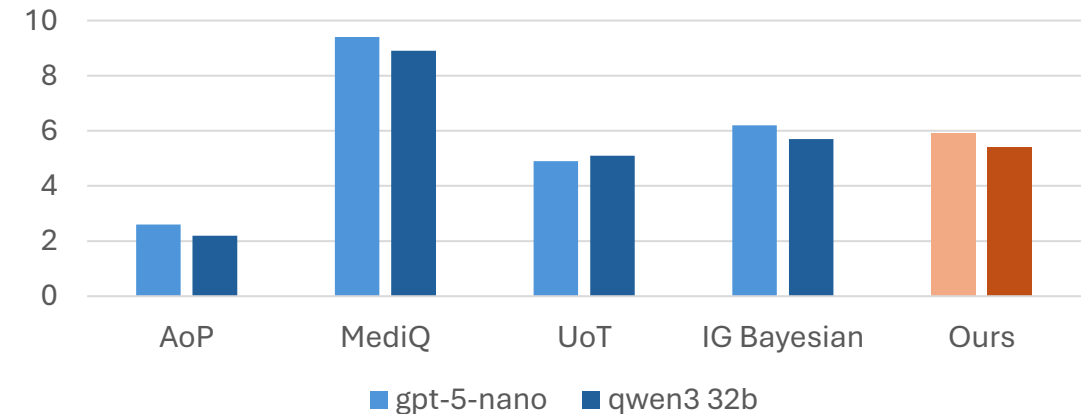
InfoGatherer improves accuracy and efficiency

- Evaluated on BarExamQA, where the model must ask questions before choosing an answer
- Baselines:
 - **AoP** (Zero-shot LLM-based)
 - **MediQ** (LLM estimates its own uncertainty)
 - **UoT** (selecting the question that maximizes expected reduction in uncertainty)
 - **IG Bayesian** version of InfoGatherer
- InfoGatherer achieves higher success while using short dialogues.
- Document-grounded uncertainty helps the model ask better questions and make more reliable decisions.

Success Rate



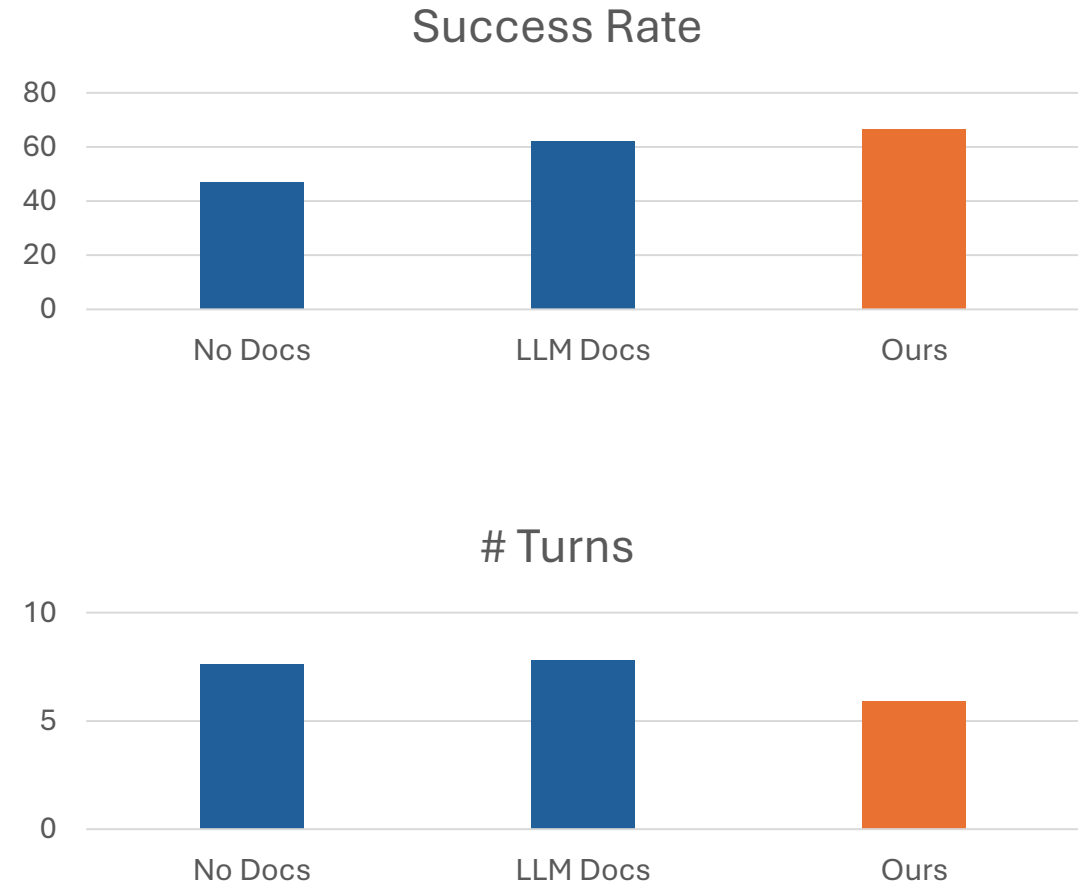
Turns



Impact of Contextual Documents

External documents help the model ask better questions and stop sooner

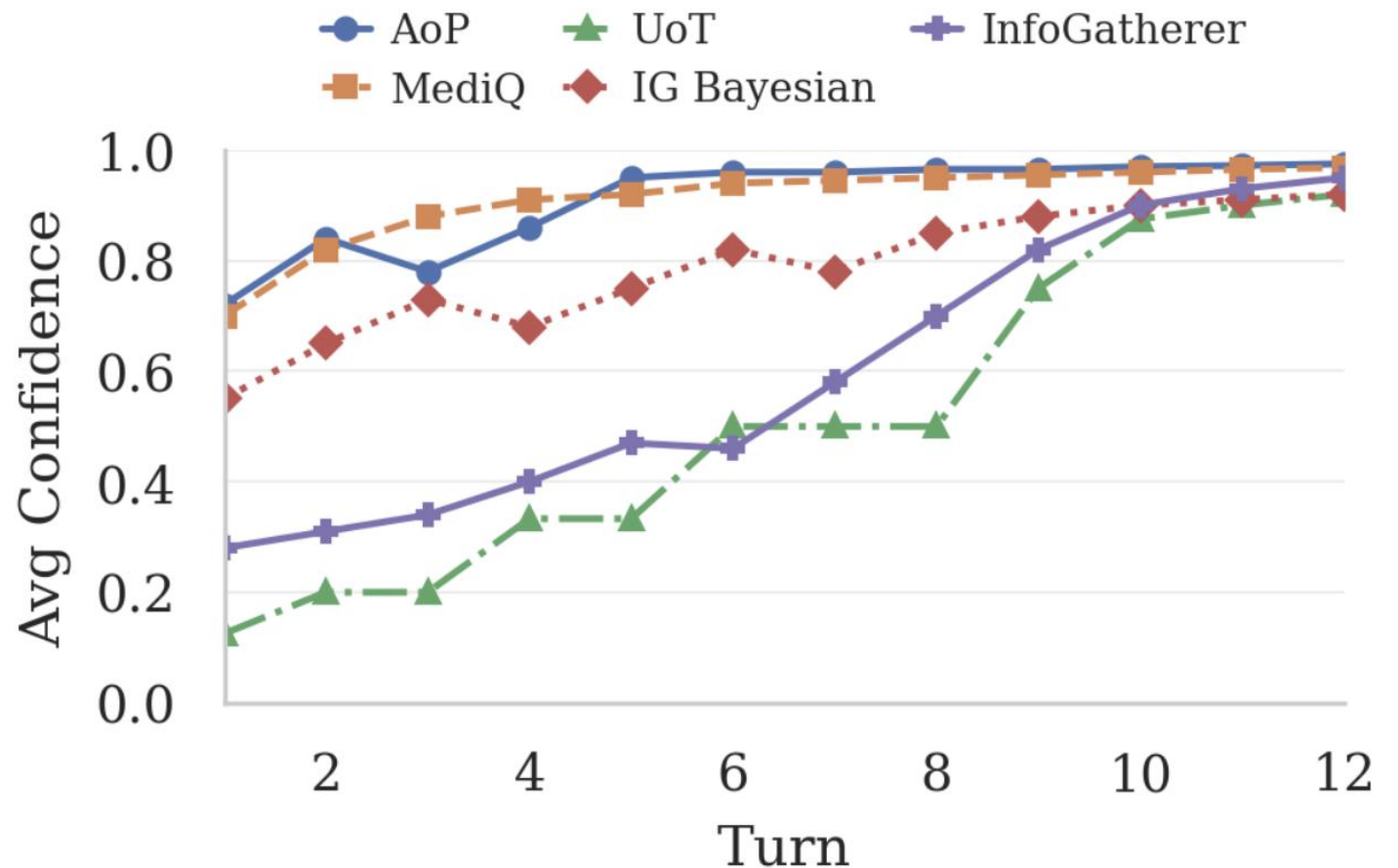
- Model-generated references can help, but their value is less consistent.
- Retrieval-grounded evidence improves both accuracy and efficiency.
- Overall, grounding in retrieved evidence supports stronger and more reliable dialogue



Dynamics of questioning objective

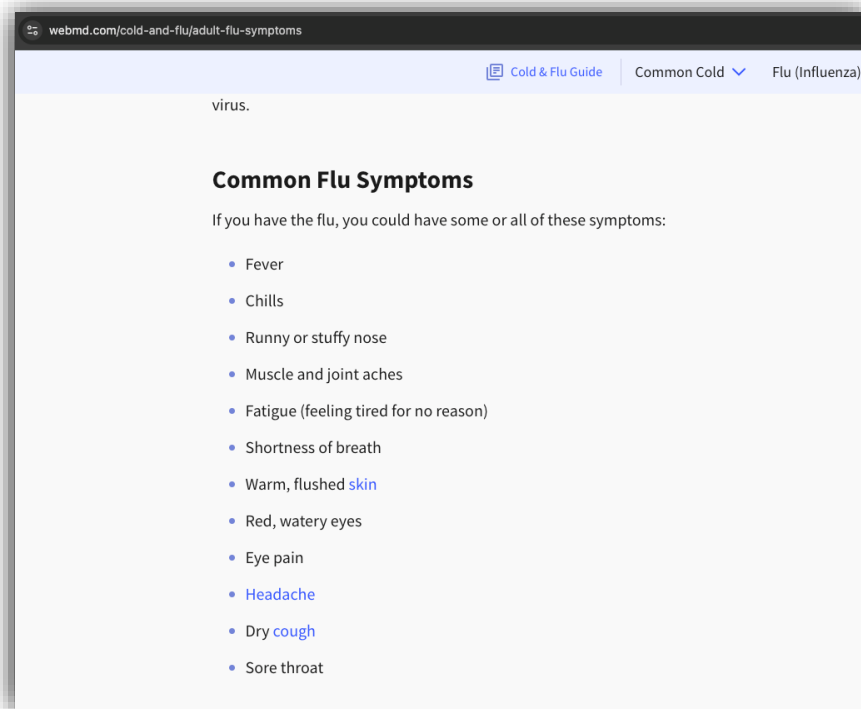
Questioning strategies impact how confidence evolves over a dialogue

- Methods based on internal agreement or point probabilities can become confident too early
- InfoGatherer asks questions that reduce uncertainty, first narrowing broad possibilities, then resolving conflicts between specific hypotheses.
- This leads to steadier progress, more informative questions, and more reliable stopping decisions.

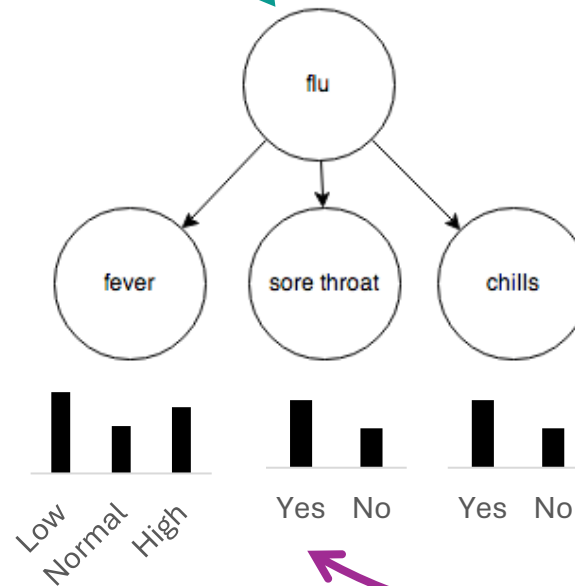


Ongoing Work: Improving Uncertainty Estimation

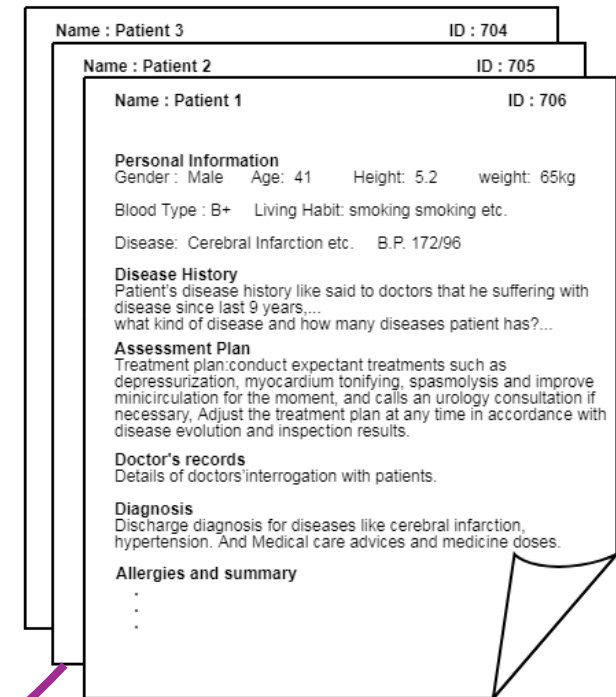
Semantic Knowledge: Medical Guidelines



Factors



Episodic Knowledge: Patient Records



Probabilities

Takeaways

- InfoGatherer helps AI agents identify what they do not know before answering
- It combines retrieved evidence, user input, and formal uncertainty reasoning to guide follow-up questions
- This leads to higher success with efficient dialogues
- Reliable AI in high-stakes domains needs both strong language models and principled decision-making