



Universität Augsburg
Fakultät für Angewandte
Informatik

Neurosymbolic Models of Uncertainty and Logical Reasoning

Prof. Dr. Annemarie Friedrich

CLASP Seminar – University of Gothenburg

December 3, 2025

Ongoing Work: Modeling Discourse Modes

With a tiny fur-piece wrapped around her shoulders, she **wiggled** her satin-covered buttocks down the street before him and **didn't stop**.

In one hand she **clutched** a hundred dollar bill and in the other a straw suitcase. The way she **strutted** down the street, the Old Man would have been blind not to have noticed both. Without looking at him, without looking at anything except Drexel Street directly in front of her, she **climbed up** into one of those orange streetcars, **rode** away in it, and never came back.

NARRATIVE MODE

– Octavia Waldo, A Cup of the Sun

ARGUMENT MODE

INFORMATION MODE

Carlota Smith. Modes of Discourse.
The Local Structure of Texts. 2010.
Cambridge University Press.

A week after Ethiopia started an offensive that it says is aimed at ending the two-year-old war it is now clear that the whole of Eritrea could become a battlefield.

REPORT MODE

With hundreds of civilians fleeing the region, Colonel Kidane said Ethiopian soldiers continue to skirmish with Eritrean soldiers on the run here in western Eritrea. Tonight, Ethiopian officials said planes bombed the main Eritrean military training center at Sawa (...).

– ...; as sampled in Smith (2003, p. 16)

Ongoing Work: Modeling Discourse Modes

DESCRIPTION MODE

With a tiny fur-piece wrapped around her shoulders, she covered buttocks down the street before him. In one hand she clutched a hundred dollar bill.

Even the stable-garage, which housed nothing now but the scent of rot, had a lawn before it. And the coffee shop on Drexel Street, where the men spent their evenings and Sundays playing cards, had a rose hedge beneath its window. The hedge reeked of coffee dregs thrown against it.

- Octavia Waldo, A Cup of the Sun

ARGUMENT MODE

INFORMATION MODE

Carlota Smith. Modes of Discourse. The Local Structure of Texts. 2010. Cambridge University Press.

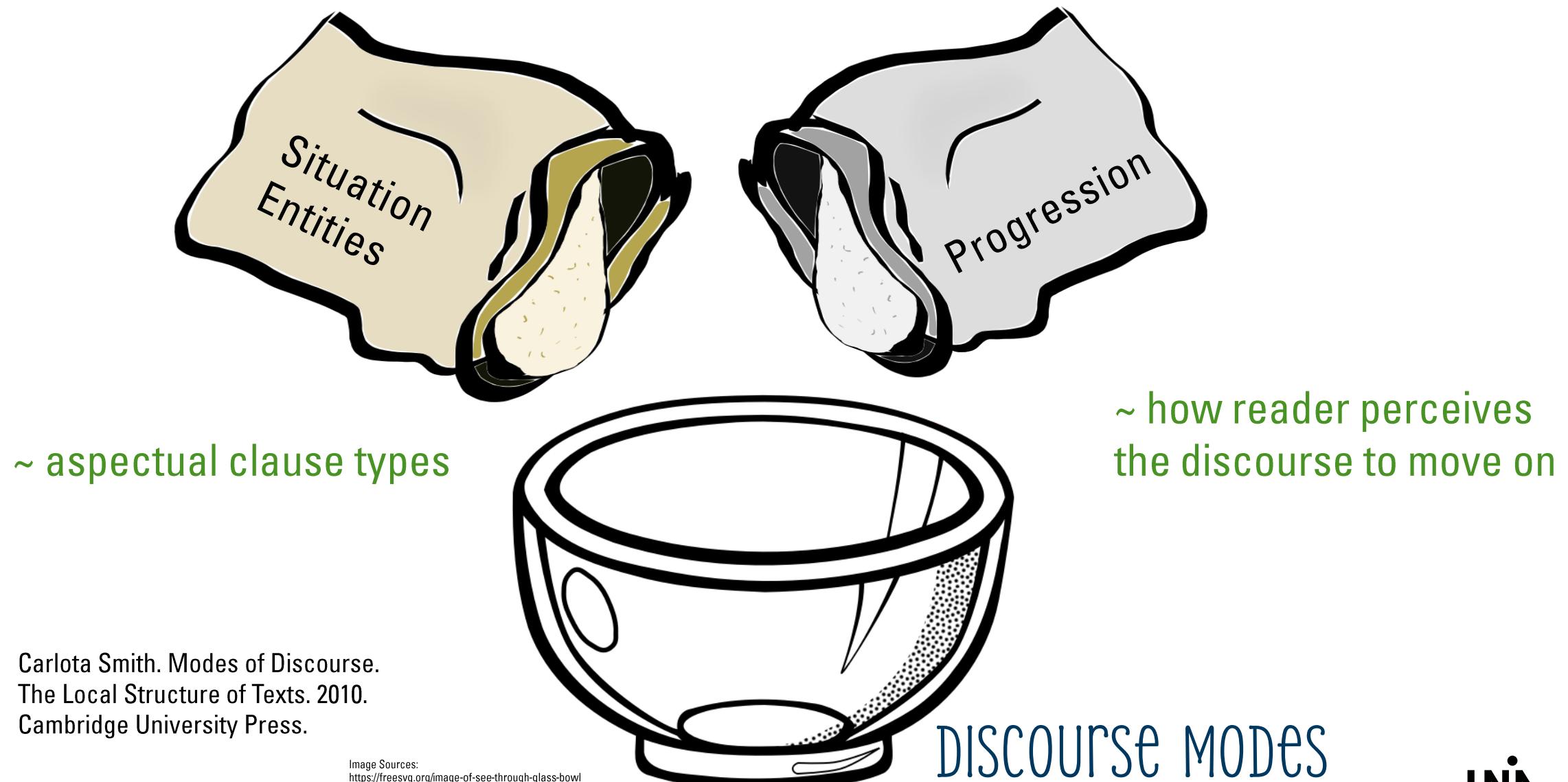
A week earlier, he had aimed a speech at the army, in which he said that the war was aimed at the military training center at Sawa (...).



- ...; as sampled in Smith (2003, p. 16)

ODE

Ongoing Work: Modeling Discourse Modes



Ongoing Work: Modeling Discourse Modes

Situation Entity Types

she **climbed up** into one of those orange streetcars (EVENT)

+ automatic modeling
+ extension to historic language variants

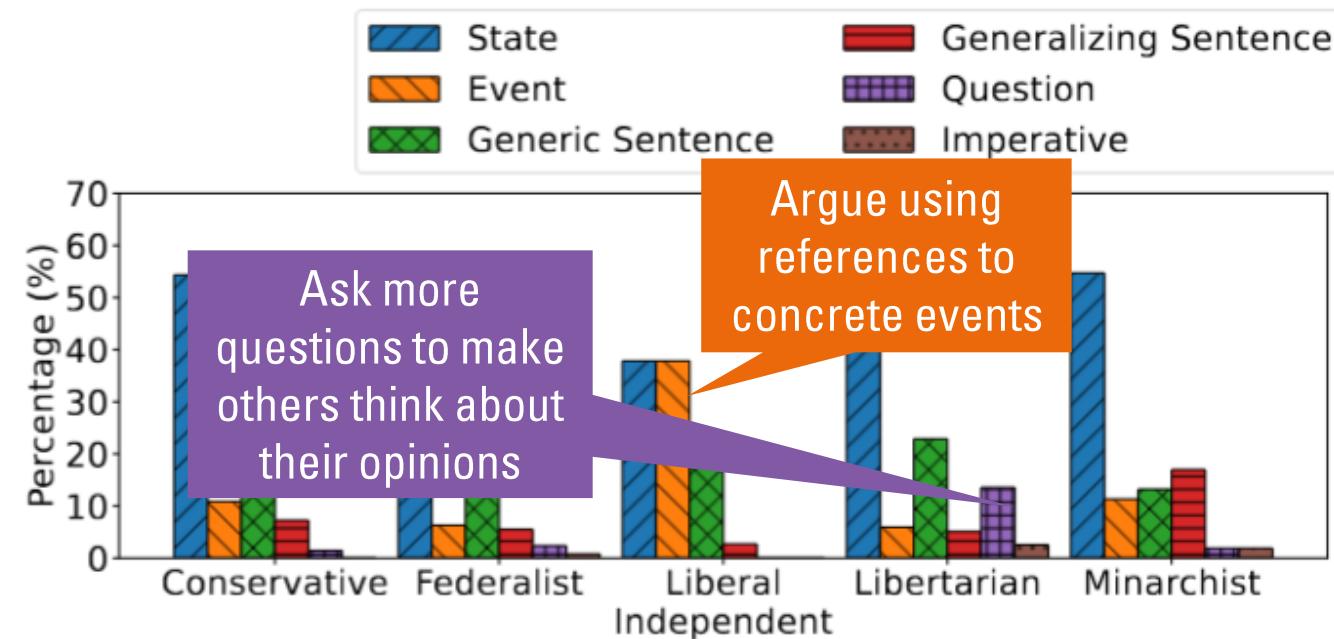
[the coffee shop] **had** a rose hedge beneath its window. (STATE)

where the men spent their **evenings and Sundays** playing cards (GENERALIZING SENTENCE)

Penguins cannot fly.
(GENERIC SENTENCE)

Situation entity types: automatic classification of clause-level aspect (Friedrich et al., ACL 2016)

(Very) Explorative Case Study on Reddit Posts on Democracy



Hanna Schmück, Michael Reder, Katrin Paula, and Annemarie Friedrich. [A Case Study on Annotating and Analysing Situation Entity Types in Reddit Discussions on Democracy](#). CMC Corpora. September 2025. Bayreuth, Germany.

Ongoing Work: Modeling Discourse Modes

Annotation of Progression Type(s)

NARRATIVE MODE

temporal progression (continuity):
events and states relate to another

temporal progression (deictic):
events and states relate to reference time

REPORT MODE

ARGUMENT MODE

INFORMATION MODE

DESCRIPTION MODE

spatial: movement through visual scene



Anaphoric/metaphorical movement through „domain“



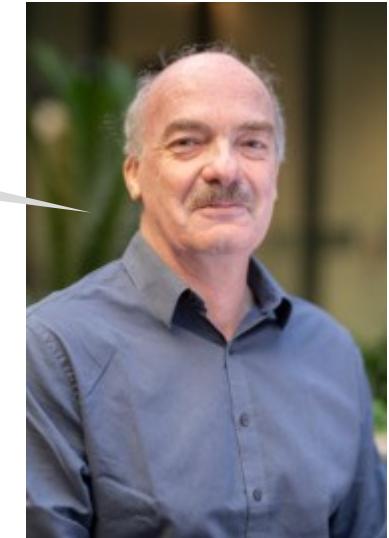
Image: generated with bing.com/create (DALL-E)

Question Mark Image Course:
<https://pixabay.com/illustrations/illustrator-graphic-design-8721204/>

Popcorn Papers



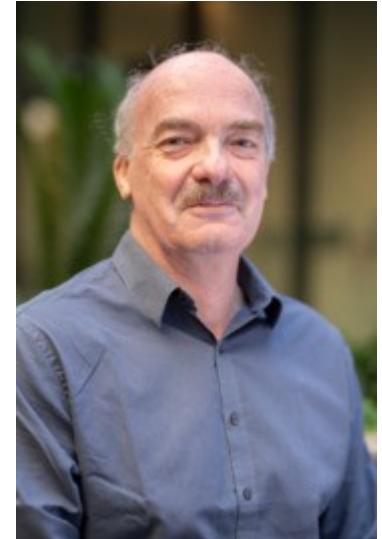
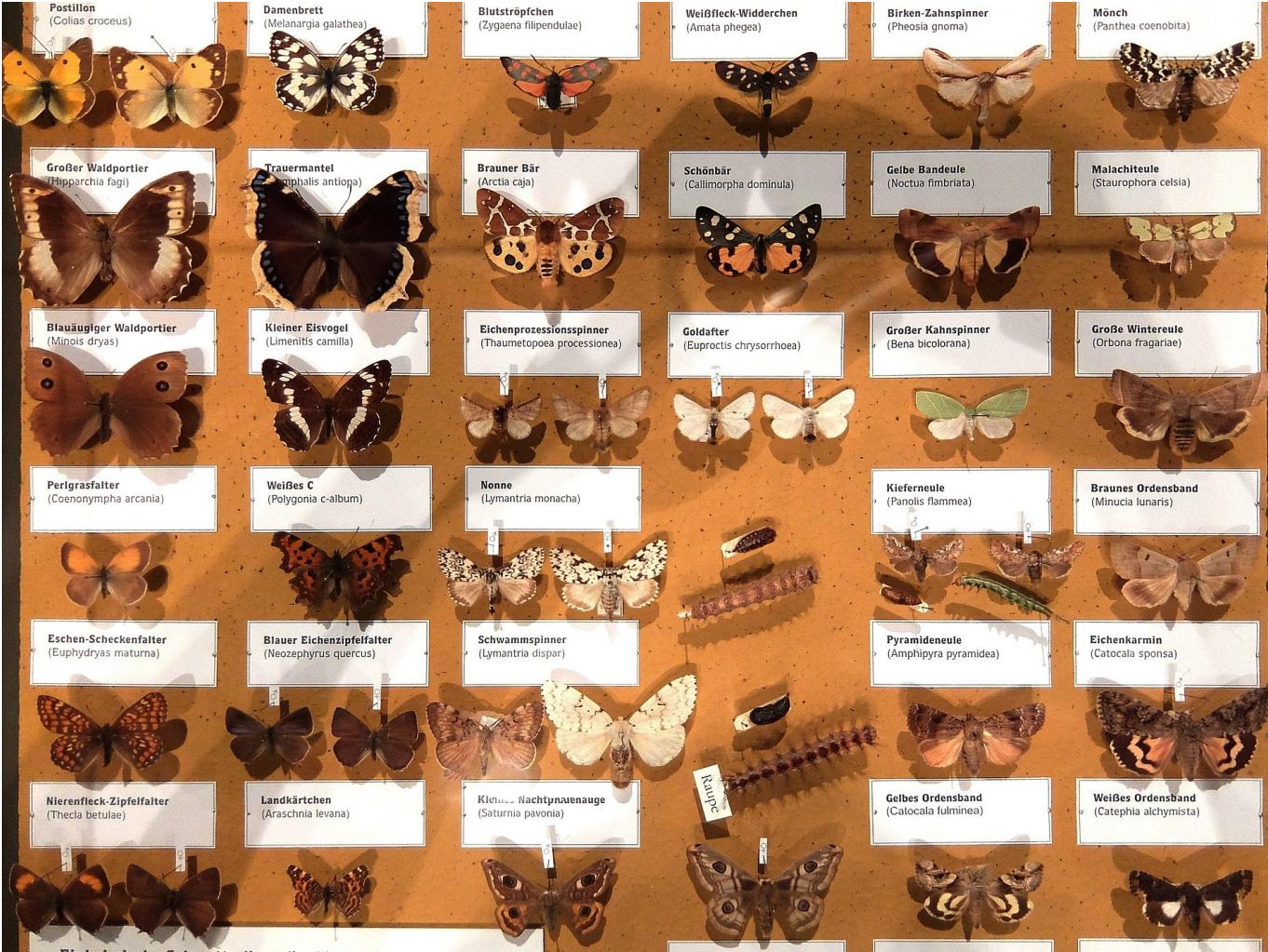
LLMs are like beautiful cars
that occasionally kill
someone.



Ed Hovy, ACL 2025

Image Sources:
<https://www.pexels.com/de-de/foto/roter-alfa-romeo-c4-auf-strasse-nahe-baumen-210019/>
<https://pixabay.com/photos/popcorn-movie-theater-theatre-food-143327/>

Butterfly Collections



Some Butterflies



LLMs cannot always generalize procedural knowledge (but sometimes they can).



Current LLMs can apply zero-shot task generalization, but they do not generalize as we humans do.



Neurosymbolic methods should not be necessary to solve NLU because our brain is a neural network, but something is still missing from LLMs.



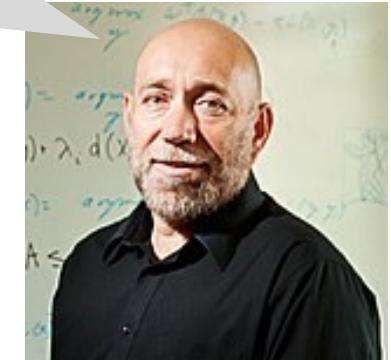
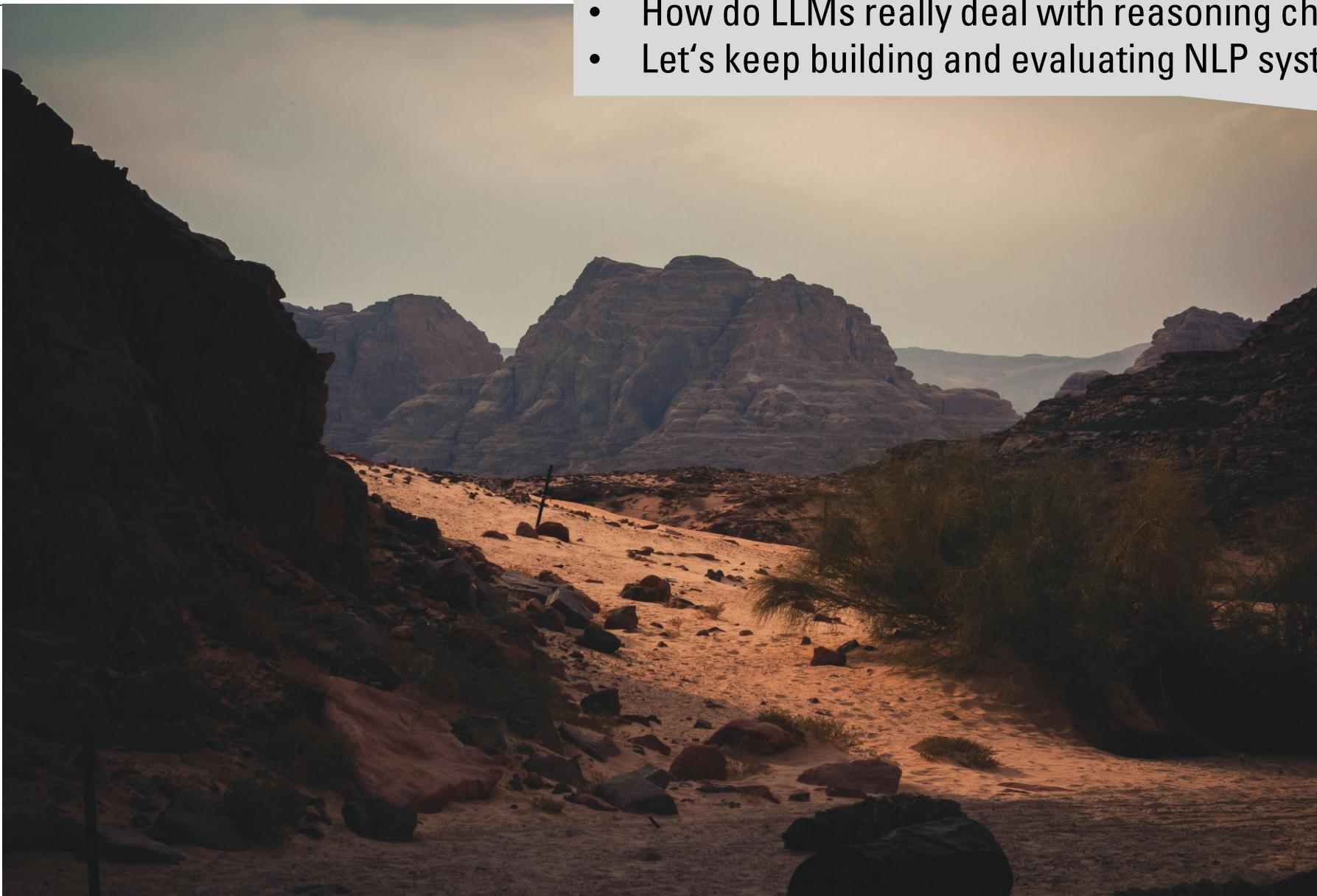
LLMs may be okay for lookups, but do we really want to trust them with decision making? Maybe we need more neurosymbolic methods.



BUT we do not know the extent of generalization for sure because of **pre-training data contamination!**

Reasoning

- LLMs generalize, but sometimes not in the way they should.
- How do LLMs really deal with reasoning chains?
- Let's keep building and evaluating NLP systems!



Dan Roth,
ACL 2025



Image Sources:
<https://www.pexels.com/de-de/foto/roter-alfa-romeo-c4-auf-strasse-nahe-baumen-210019/>
<https://www.pexels.com/photo/scenic-landscape-of-mountains-and-desert-19311674>

Neurosymbolic Approaches



ProbLog

Bayesian reasoning

Timo Pierre Schrader, Lukas Lange, Simon Razniewski, and Annemarie Friedrich. QUITE: Quantifying Uncertainty in Natural Language Text in Bayesian Reasoning Scenarios. In *EMNLP 2024*.



clingo

Answer Set
Programming (ASP)

Timo Pierre Schrader, Lukas Lange, Tobias Kaminski, Simon Razniewski, Annemarie Friedrich. A Solver-in-the-Loop Framework for Improving LLMs on Answer Set Programming for Logic Puzzle Solving. *AAAI 2026*. (to appear)

Question Answering in Complex Technical Domains

Question Answering



Task

Answer questions posed in natural language
Extract textual evidence



Challenges

Correct text interpretation and logical reasoning

Mathematical reasoning

(Data) Uncertainty, modality, hedging, negation

Conditions

Model Uncertainty: Confidence Estimation

We produced mutations in homologous recombination, loss of the most abundantly death in the larval or pupal stage. Under optimal conditions (20–30 isolated homozygous larvae per vial), about 60% of larvae homozygous for the *adgf-a* mutation reach the third instar. Development during the third larval instar is [...] at about 5 d of development. Some homozygous third-instar larvae can be found alive in the vial even after 10 d of development. Mutant third-instar larvae show fat body disintegration (Figure 1A and 1B) and multiple melanotic tumors (Figure 1C), predominantly in the caudal part of the body and accompanied by disintegration of the fat body. Melanization of the lymph glands was never observed in these larvae, and the imaginal discs and brain appear normal. Less than 30% of homozygotes eventually pupate. Homozygous pupae usually die soon after pupariation; [...]. Less than 2% of mutant pupae develop normally and eventually emerge as adults without any obvious abnormalities besides the abdominal curvature; some of them are sterile.

What is the likelihood of larvae homozygous developing normally?



A B

Adenosine Deaminase in *Drosophila*

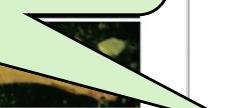


Figure 1. *adgf-a* Mutant Phenotype
(A and B) Fat body disintegration caused by GFP expression driven by heat shock promoter in *adgf-a* homozygous third instar larvae. While *adgf-a* homozygous third instar larvae have normal flat layers of fat body (A), *adgf-a* mutant showed extensive fat body disintegration into small pieces of fat body (B).
(C) Multiple melanotic tumors present in *adgf-a* mutant third-instar larvae.
(D) An *adgf-a* mutant pupa with typical abdominal curvature.
DOI: 10.1371/journal.pbio.0030301.g001

while non-heat shocked animals still produced many melanotic tumors, only 22% of animals that were heat shocked at late embryonically first instar developed these tumors (Figure 2B). This result confirmed that the phenotype was caused by overexpression in the *ADGF-A* gene. This conclusion is further supported by the even more efficient rescue achieved by expression of transgenically provided ADGF-A in the lymph glands using the Gal4/UAS system (see below).

The *adgf-a* Mutant Phenotype Is Associated with Elevated Levels of Adenosine and/or Deoxyadenosine
Using liquid chromatography and mass spectrometry of deoxynucleosides in hemolymph samples, we measured adenosine concentrations in hemolymph of mutant and wild-type third-instar larvae. The adenosine concentration in the *adgf-a* mutant was $1.11 \pm 0.29 \mu\text{M}$ compared to less than $0.08 \mu\text{M}$ in wild-type larvae. The deoxyadenosine concentration in mutants was $1.66 \pm 0.69 \mu\text{M}$ compared to an undetectable level in the wild type.

The Catalytic Activity of ADGF-A Is Required for Its Function

To test whether the function of ADGF-A in vivo is also dependent on its catalytic activity, we produced two versions

1214 July 2005 | Volume 3 | Issue 7 | e201

Mathematical Reasoning, Explainability & LLMs ??

Correct interpretation requires translation into Bayesian probabilities and mathematical calculations:

$$P(\text{Pupation}) = 0.3$$

$$P(\text{Normal}|\text{Pupation}) = 0.02$$

$$\begin{aligned} P(\text{Normal}) &= P(\text{Pupation}) \cdot P(\text{Normal}|\text{Pupation}) \\ &= 0.3 \cdot 0.02 = 0.006 = 0.6\% \end{aligned}$$



Answer is not very informative.



WHY!?

All larvae in study carry the mutation!

We produced mutations in homologous recombination, loss of the most abundantly death in the larval or pupal stage. Under optimal conditions (20–30 isolated homozygous larvae per vial), about 60% of

What is the likelihood of larvae homozygous developing normally?



The likelihood of homozygous larvae developing normally and emerging as adults is less than 2%. Although less than 30% of the larvae reach the pupal stage, only a small fraction of the pupae (less than 2%) go on to become adults.

What is the likelihood of an adult in the study carrying the mutation?



In the study, it is reported that less than 2% of mutant pupae develop normally and eventually emerge as adults. This indicates that the percentage of adults carrying the mutation is less than 2%.

Mathematical Reasoning, Explainability & LLMs ??

Freiesleben & König, 2024: „We overfit explanation techniques to human intuitions.“ → trustwashing



Chain-of-
Thought
Prompting

Think step by step.

Generating Post-
Hoc Explanations

Explain your
calculations /
reasoning.



Thank you for clarifying that all larvae in the study were mutants. Given this context, we can now determine the percentage of adults carrying the mutation based on the information provided in the excerpt:

1. **Total Mutant Larvae:** All larvae mentioned are mutants (homozygous for the adgf-a mutation).
2. **Development to Adulthood:**
 - Less than 2% of the homozygous pupae (mutants) develop into adults.

Calculation of Percentage:

Since all larvae are mutants and less than 2% of them ultimately emerge as adults, we can conclude that:

- **Percentage of Adults Carrying the Mutation: Less than 2%**

This percentage reflects the proportion of mutant larvae that develop into adults. If, for instance, there were 100 homozygous larvae, fewer than 2 of them would survive to become adults.

So the answer is: **less than 2% of the adults in the study are carrying the mutation (since all larvae were mutants).**

Mathematical Reasoning, Explainability & LLMs ??



Image Source: <https://pixabay.com/photos/tree-face-forest-nature-smile-4721608/>

Mathematical Reasoning, Explainability & LLMs ??

Freiesleben & König, 2024: „We overfit explanation techniques to human intuitions.“ → trustwashing



Chain-of-
Thought
Prompting

Think step by step.

Generating Post-
Hoc Explanations

Explain your
calculations /
reasoning.



Thank you for clarifying that all larvae in the study were mutants. Given this context, we can now determine the percentage of adults carrying the mutation based on the information provided in the excerpt:

1. Total Mutant Larvae: All larvae mentioned are mutants (homozygous for the adgf-a mutation).
2. Development into Adults:
 - Less than 2% of the

Calculation of Percent:

Since all larvae are mutants and less than 2% of them ultimately emerge as adults, we can conclude that:

- Percentage of Adults Carrying the Mutation: Less than 2%

This percentage reflects the proportion of mutant larvae that develop into adults. If, for instance, there were 100 homozygous larvae, fewer than 2 of them would survive to become adults.

So the answer is: **less than 2% of the adults in the study are carrying the mutation (since all larvae were mutants).**

Popularity of these methods due to Eliza Effect, the linguistic correlate of pareidolia (Steven Bird, ACL 2024)

Quantifying Uncertainty In natural Text (QUITE)

Better calculations that make intermediate steps explicit? → Neurosymbolic models

For research & development, we need texts like this ...

I am an insurance agent supposed to decide how much an insurance for a customer should cost.

If the insurance holder is a senior with a cautious behaviour in terms of **avoiding risk**, then it is **highly likely** that they have received **senior training**, and it is **improbable** that they have not received **such training**.
[...]

The driver has attended a senior driver's training and the car's mileage is 20,000. What is the likelihood of being involved in a moderate car accident?

+ gold standard answers

P = 34.5%

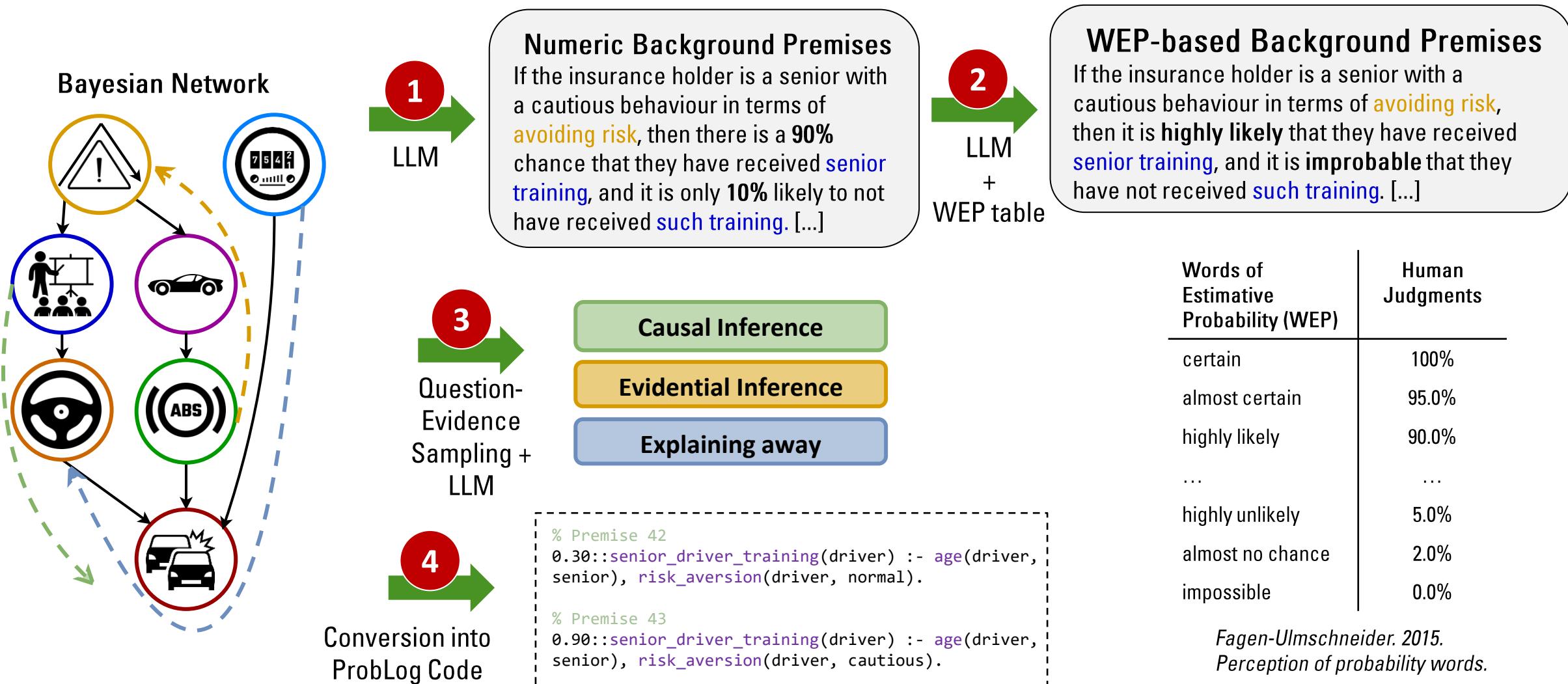
+ gold standard symbolic representations

```
% Premise 0
0.9::attends_training(driver) :- cautious(driver);
0.1::(not attends_training(driver)) :- cautious(driver).

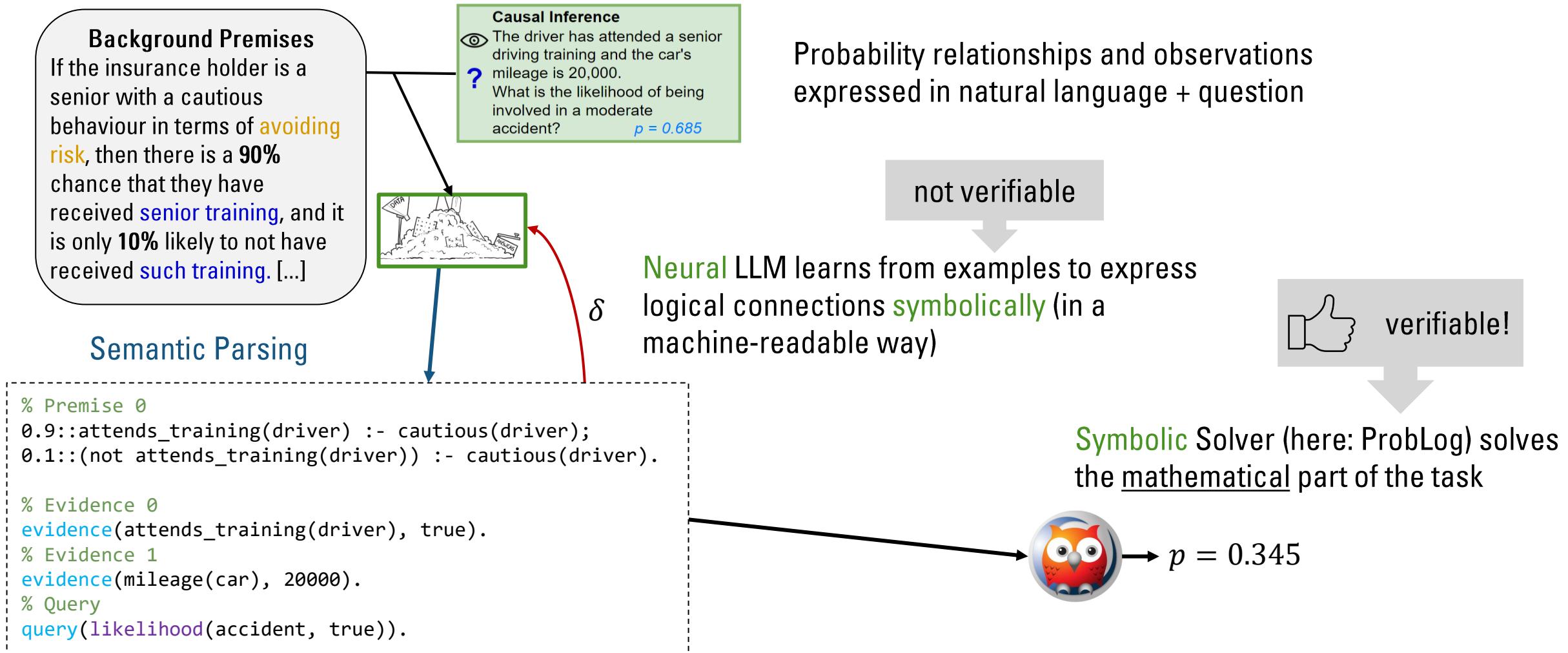
% Evidence 0
evidence(attends_training(driver), true).
% Evidence 1
evidence(mileage(car), 20000).
% Query
query(likelihood(accident, true)).
```

Timo Pierre Schrader, Lukas Lange, Simon Razniewski, and Annemarie Friedrich. [QUITE: Quantifying Uncertainty in Natural Language Text in Bayesian Reasoning Scenarios](#). In *EMNLP 2024*.

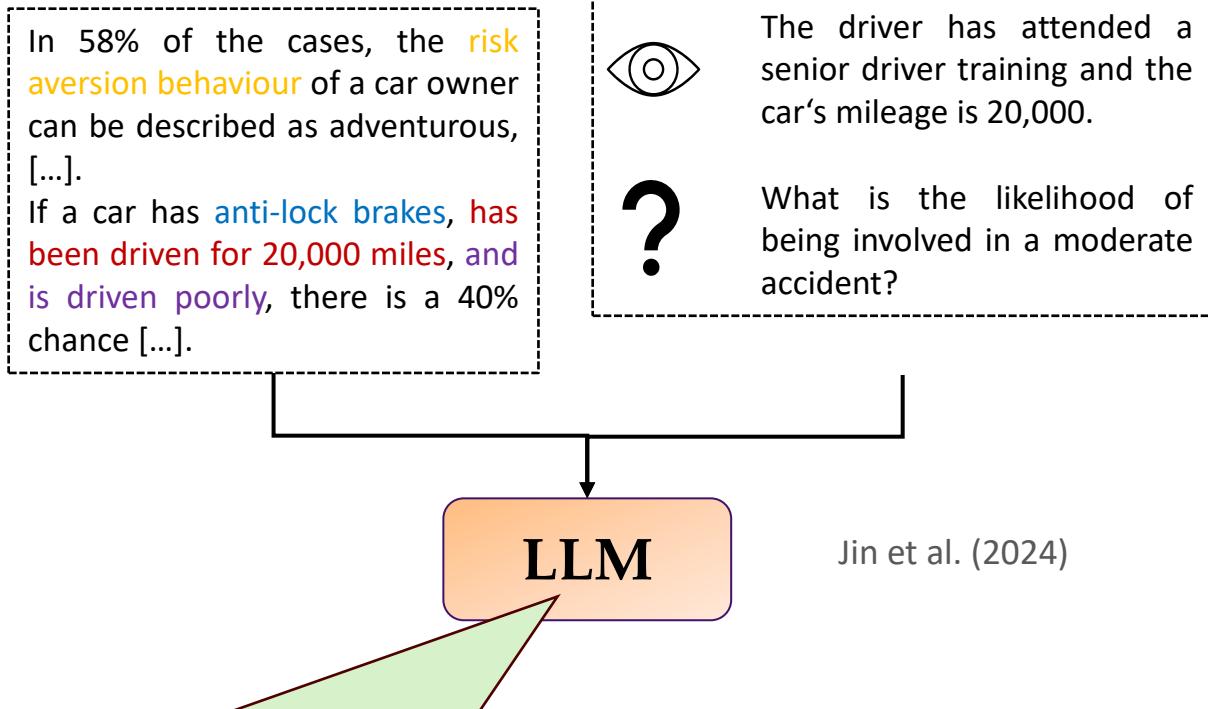
Quite Dataset Creation Process



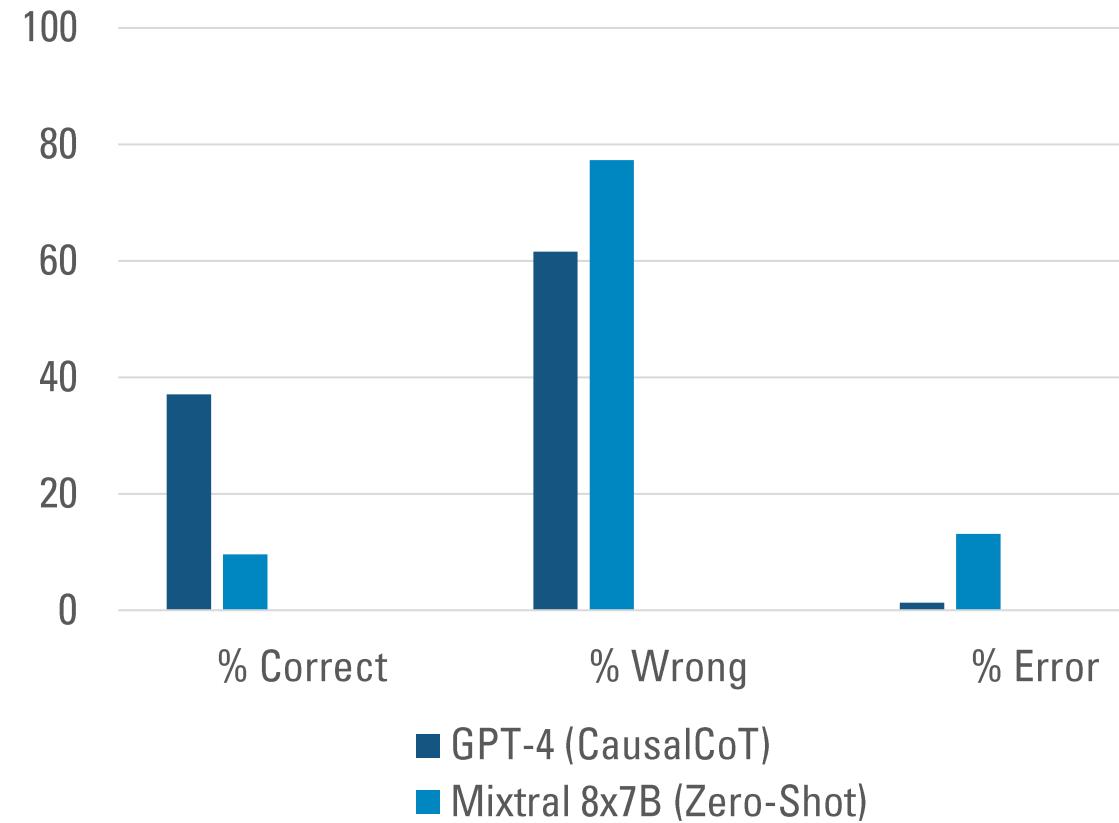
Neurosymbolic Model (here: for Bayesian Reasoning)



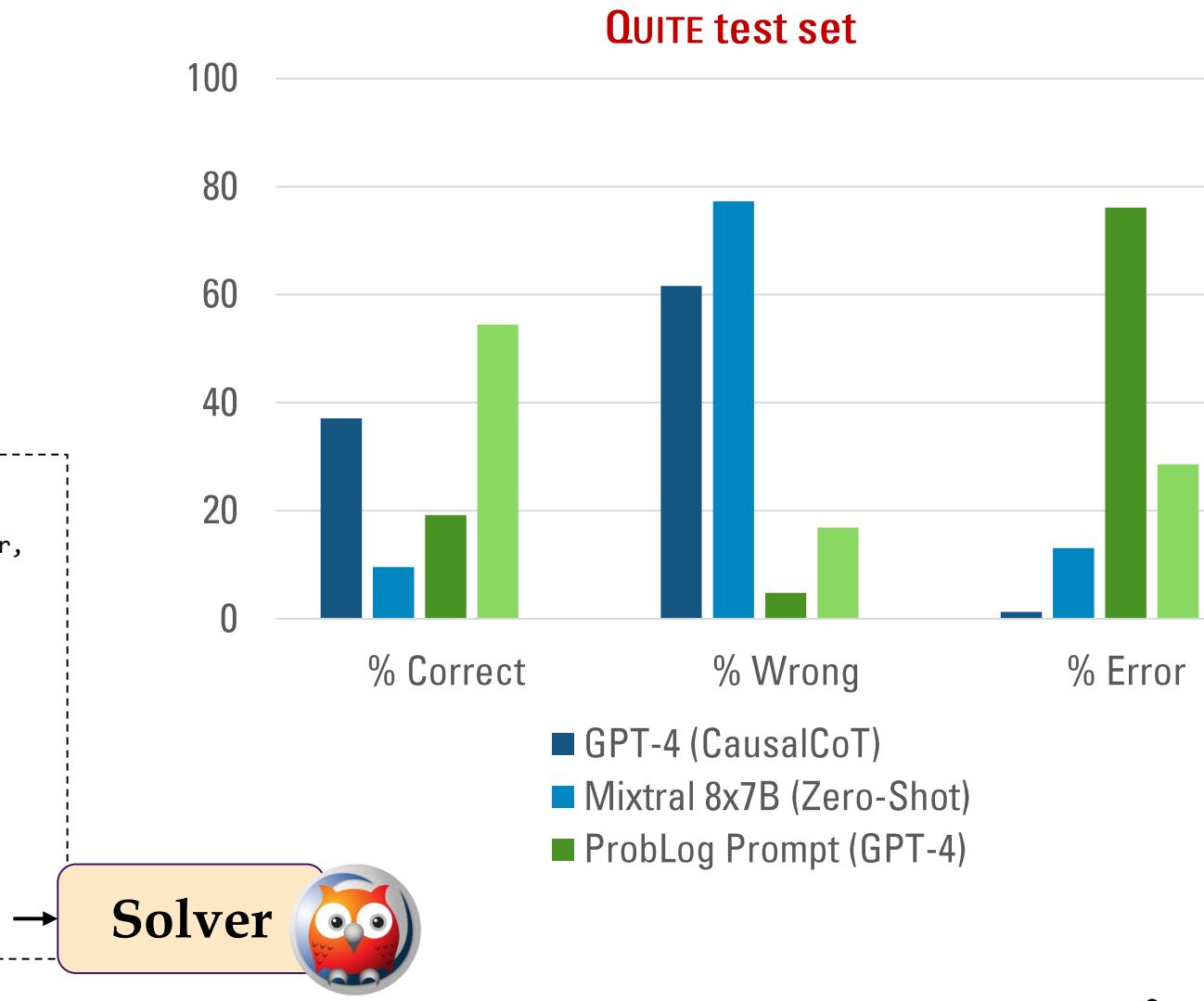
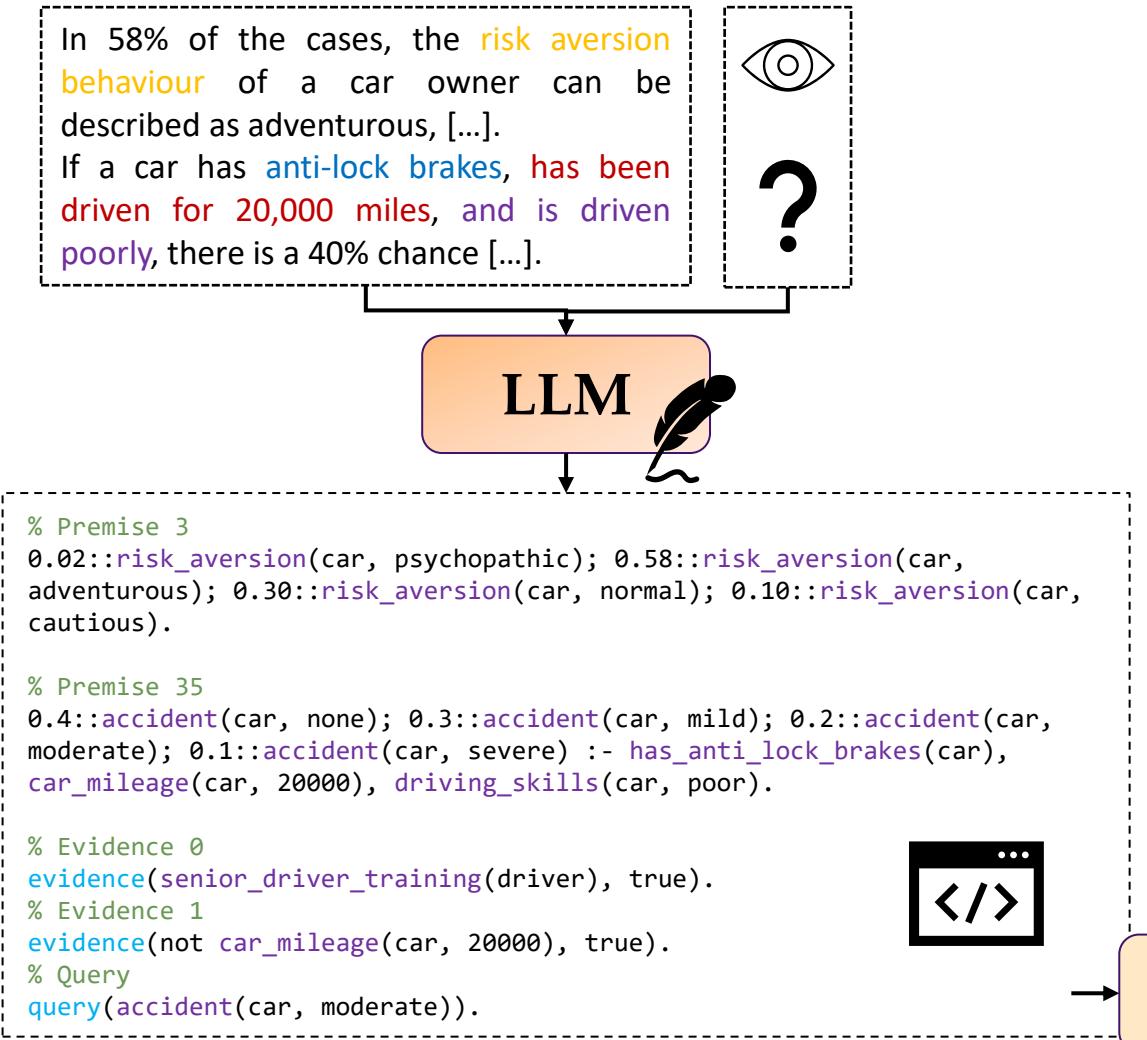
Experiments on QUITE: LLM Baseline



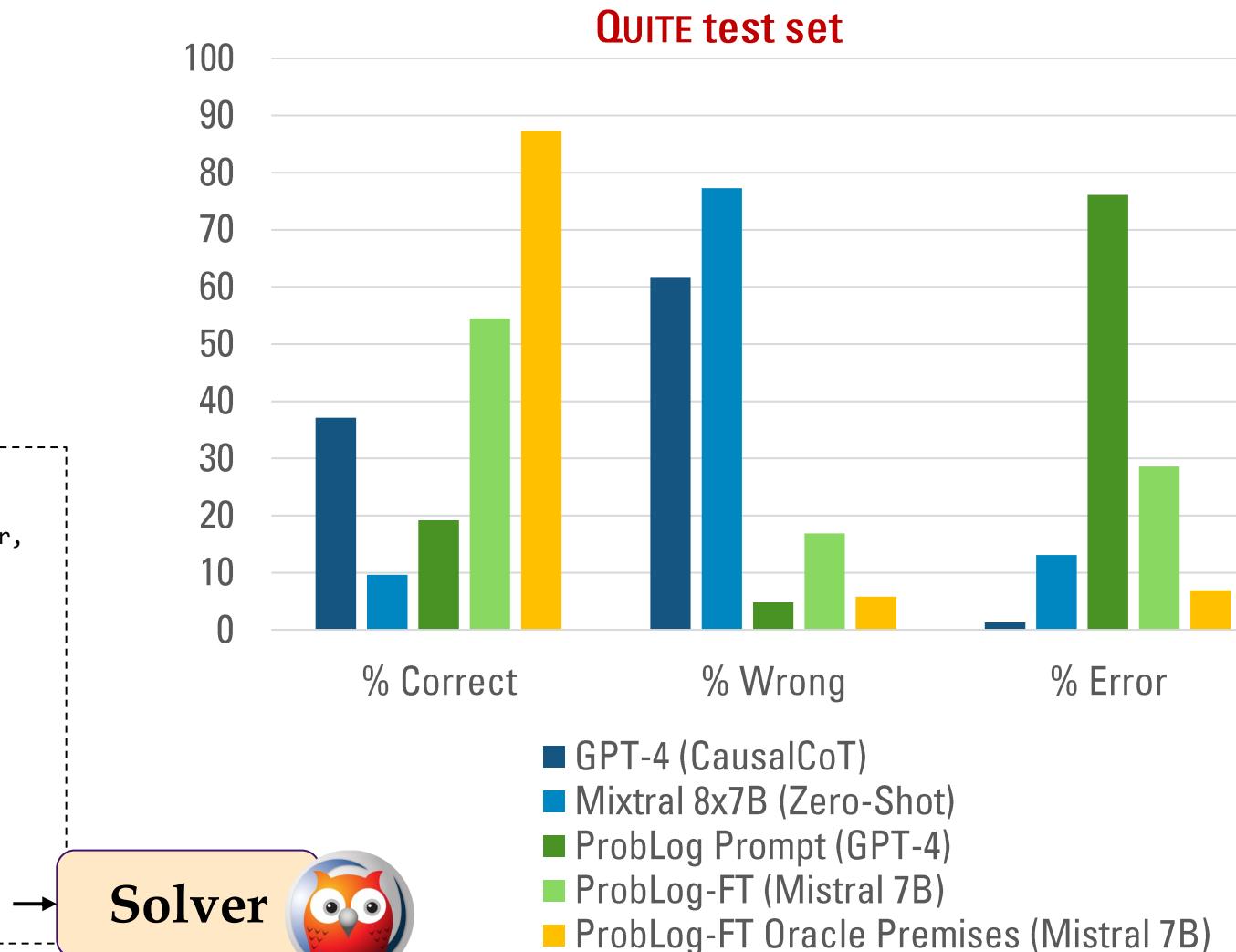
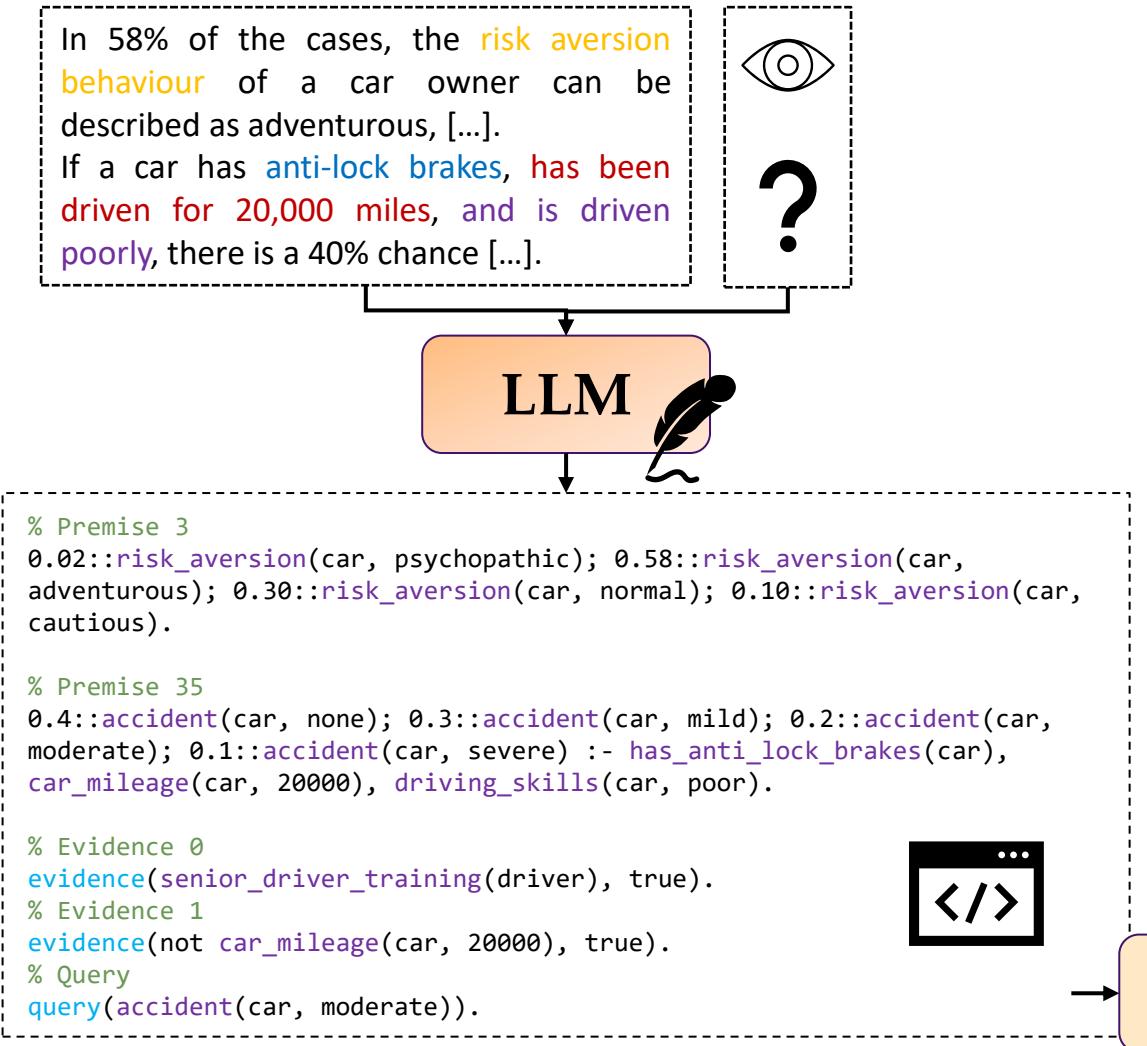
"To calculate the likelihood of the driver being involved in a moderate accident, we follow a similar approach to the example provided. Let's analyze the premises and evidences ..."



Experiments on QUITE: Neuro-Symbolic Approach



Experiments on QUITE: Neuro-Symbolic Approach



Experiments on QUITE: Neuro-Symbolic Approach

WEP-based premises

Fallback for errors = 50%

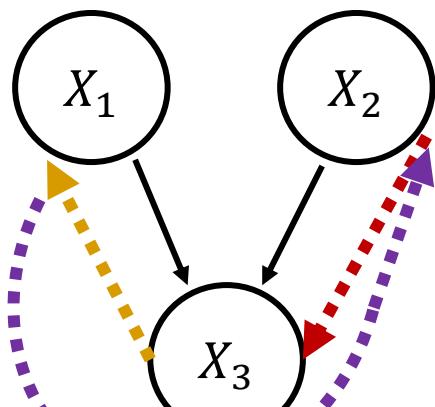
Zero-shot and CoT models almost always predict a number, which is often wrong.

ProbLog models run into parsing errors a lot (probably too little training data)

ZERO-SHOT		T4-Turbo	82.1	12.2
	Llama-3-8B	2.2	83.4	14.4
	Mixtral-8x7B	3.5	50.7	45.9
CAUSALCoT		GPT4-Turbo	89.1	2.2
	Llama-3-8B	3.5	91.7	4.8
	Mixtral-8x7B	2.6	59.4	38.0
Regression-FT		Llama-2-7B	0.0 ± 0.0	100.0 ± 0.0
LLM-FT		Mistral-7B	3.6 ± 0.9	96.4 ± 0.9
ProbLog-Prompt		GPT4-Turbo	0.4	8.7
ProbLog-FT		Mistral-7B	1.3 ± 0.6	32.8 ± 4.6

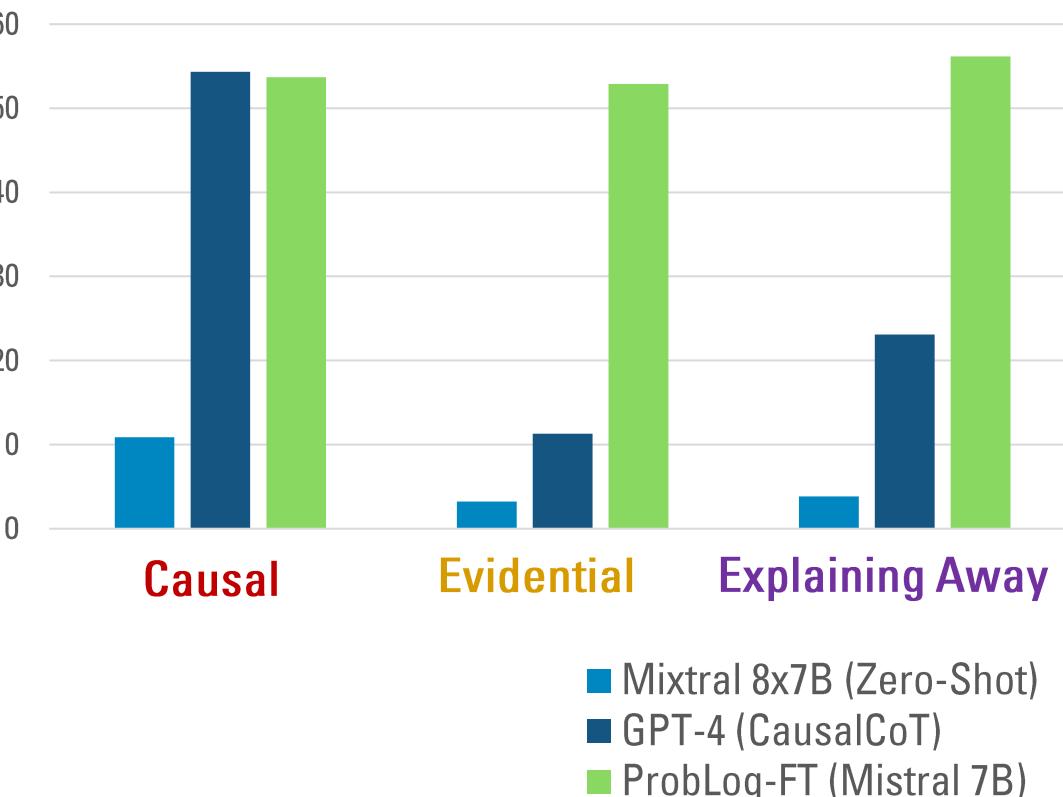
But the predicted outcomes are closest to what they should be.

Quite: Results by Reasoning Type



% Correct

LLMs mostly assume causal relations.
Neurosymbolic model is the only one that
is robust across reasoning types!



Logic Puzzles

Tom, Anna, and Eva are walking their dogs: a Saint Bernard, a Golden Retriever, and a Beagle. Who is walking which dog and at what time?

1. Tom is walking the Beagle one hour before the Golden Retriever.
2. ...

	Tom	Anna	Eva	8	9	10
	X					
		O	X	X		X
	X				X	
8						
9						
10	X					

Answer Set Programming (ASP)

Finding optimal assignments to combinatorial problems is time-intensive, complex, and hence costly.

Project Management

Requirements:

1. In every project, there must be at least one senior developer (exp. > 5 yrs.)
2. Every developer must [...]
3. No two SW testers [...]
4. Every skill must be [...]



Scheduling

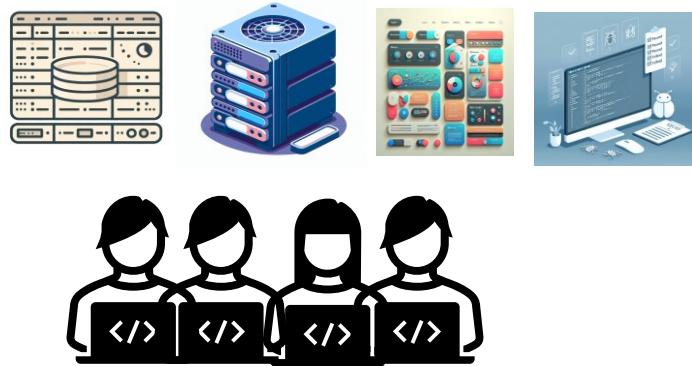
Requirements:

1. No worker must work longer than 10 hours.
2. One can only work [...]
3. Nobody can work 7 days per week.
4. A break after at most [...]



Solver-in-the-Loop Method for Creating Preference Data for ASP Coding

Datasets with problems / logic puzzles
in natural language + solution.



Requirements:

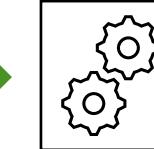
1. In every project, [...]
2. Every developer must [...]
3. No two SW testers may be part of the same project.
4. Every skill must be [...]

Timo Pierre Schrader, Lukas Lange, Tobias Kaminski, Simon Razniewski, Annemarie Friedrich. [A Solver-in-the-Loop Framework for Improving LLMs on Answer Set Programming for Logic Puzzle Solving. AAAI 2026. \(to appear\)](#)

Each added ASP statement further constrains set of possible assignments.

```
:- assigned(D1, P), assigned(D2, P), has_skill(D1, test),  
has_skill(D2, test).  
  
:- assigned(D1, P), assigned(D2, P), has_skill(D1, test), not  
has_skill(D2, test).  
  
:- assigned(D1, one), assigned(D2, two), has_skill(D1, test),  
has_skill(D2, two).  
  
:- assigned(D1, one), assigned(D2, one).
```

ASP Solver
in the Loop



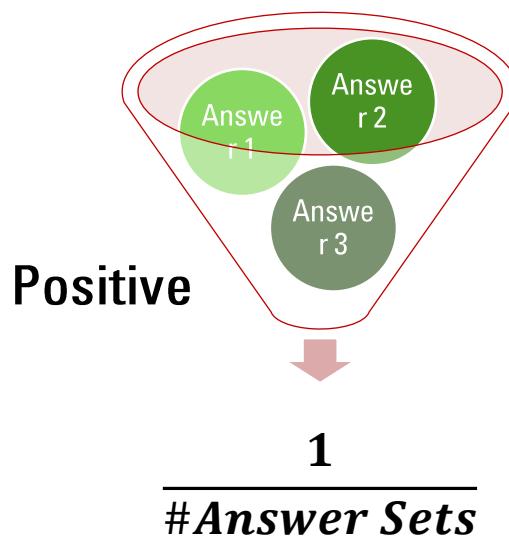
Checklist	
✓	Produces Correct Partial Answer
✗	Produces Errors
✗	Is Unsatisfiable
✗	Produces Wrong Answer

✓ chosen
✗ rejected
✗ rejected
✗ rejected

LLM

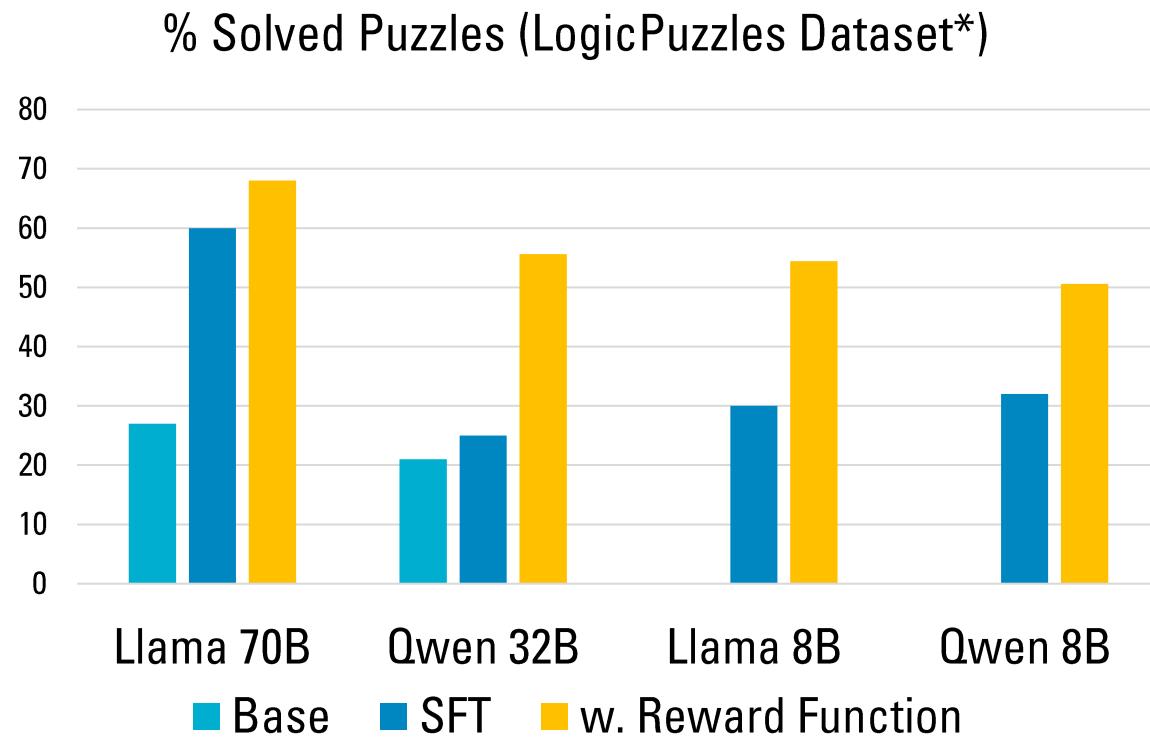
Experimental Results

Test Time Decision Support



	Negative
✗	Produces Errors
✗	Is Unsatisfiable
✗	Produces Wrong Answer

Timo Pierre Schrader, Lukas Lange, Tobias Kaminski, Simon Razniewski, Annemarie Friedrich. [A Solver-in-the-Loop Framework for Improving LLMs on Answer Set Programming for Logic Puzzle Solving](#). AAAI 2026. *(to appear)*



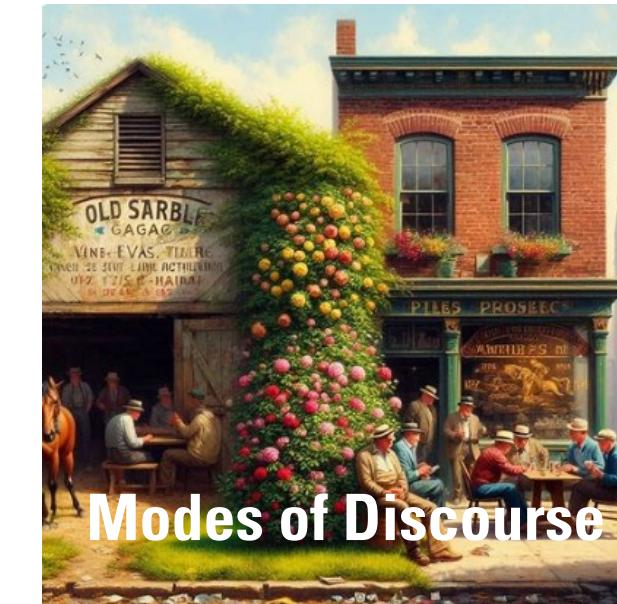
*[Learning to Automatically Solve Logic Grid Puzzles](#)
(Mitra & Baral, EMNLP 2015)

Summary



Image Source: <https://www.pexels.com/de-de/foto/foto-des-nebenbaum-geparkten-fahrzeugs-2527923/>

- Build systems that are much better at **complex reasoning tasks**
- Help to gain **insights** in reasoning strengths and weaknesses of pure LLMs
- Modeling **linguistic uncertainty** (+ interaction with model calibration) needs more attention
- Implement **linguistic frameworks** to gain a better understanding of **intra-text variation**



Modes of Discourse



boschresearch/quite-emnlp24



datasets/timo-pierre-schrader/QUITE