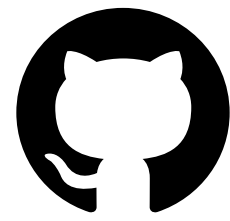


Abductive Reasoning in Deep Learning

choucisan



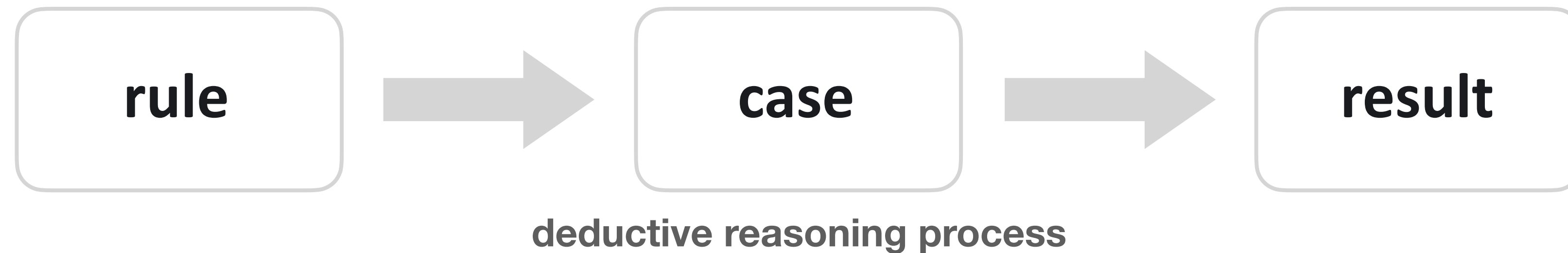
2025 / 6 / 13

Reasoning

- Three major types of **reasoning** in philosophy. [Analytica Priora, Aristotle]
 - deductive reasoning
 - inductive reasoning
 - abductive reasoning

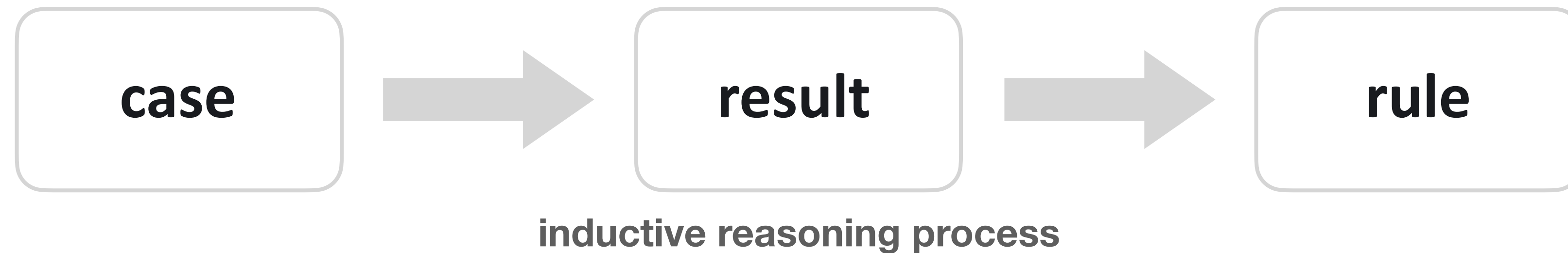
Deductive Reasoning

- Deduce results from existing rules combined with the current situations.
 - **Knowledge Graph** and Reasoning
 - **Information Retrieval** and **Retrieval-Augmented Generation**



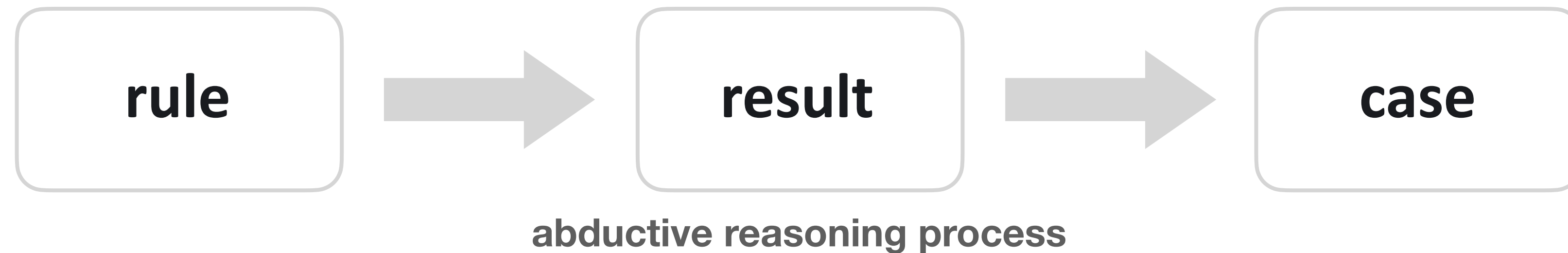
Inductive Reasoning

- Induce rules from a large number of situations (data) and results (labels).
 - **Image classification** and **segmentation**
 - **Speech recognition** and **generation**



Abductive Reasoning

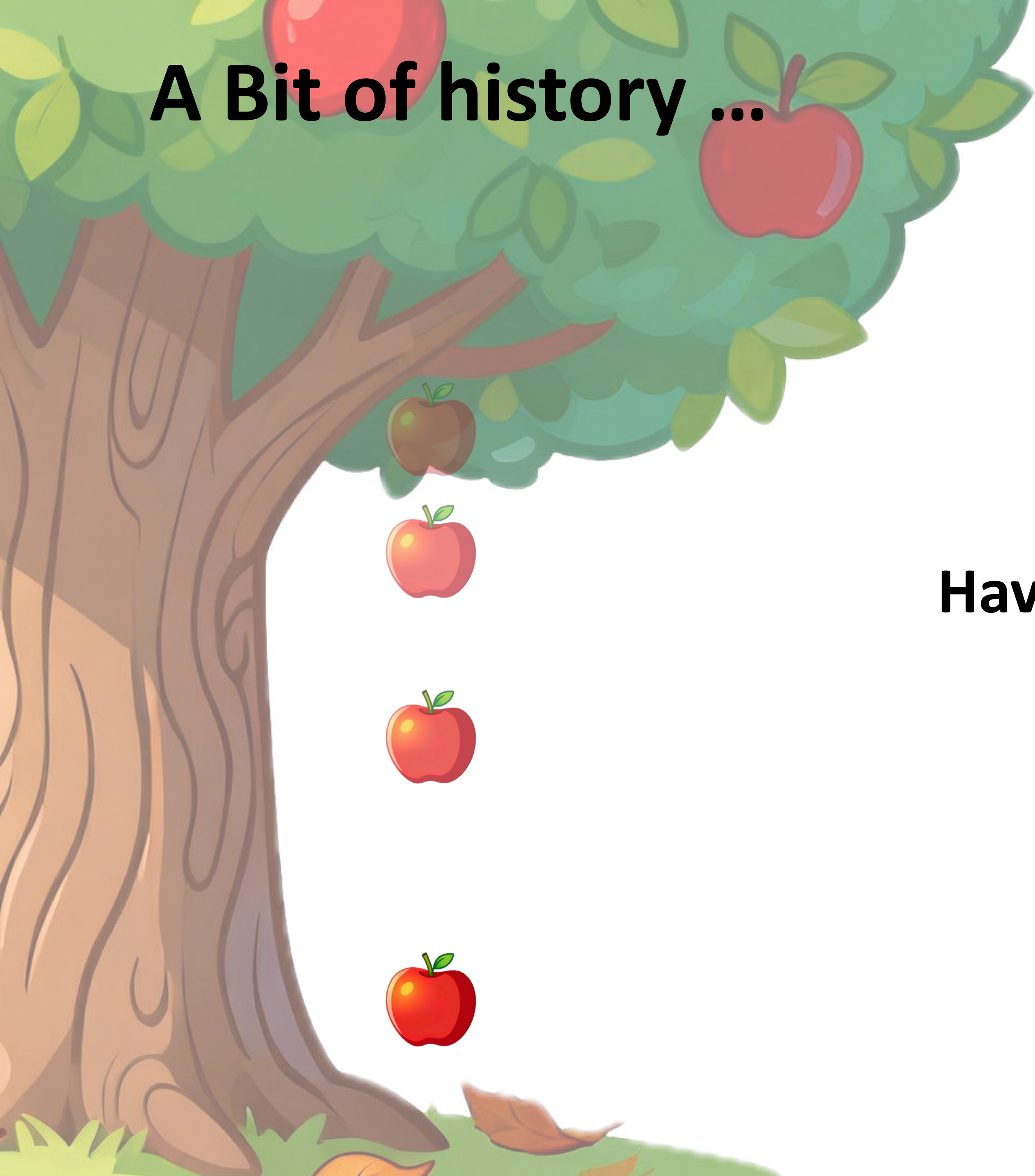
- Abduce a hypothesis from a series of observations and existing rules.
- **Clinical diagnosis** and **Scientific discovery**
- **Subjective experience** and **Memory**



Abductive Reasoning and Pattern Recognition

“Recognizing the underlying unity among phenomena that initially appear unrelated is an experience of great intellectual grandeur.”

– *Albert Einstein, the Knight of the Light Beam*



A Bit of history ...

Have you ever seen an apple tree?

Abductive Reasoning and Galileo's Discoveries

- How did Galileo use **abductive reasoning** to discover the **acceleration**.

Observation

$$t = 0$$



$$t = 1s$$



$$t = 2s$$



$$t = 3s$$



$$v = ?$$



Hypothesis

Speed is **proportional to time**,
Free fall is a kind of **uniformly
accelerated linear motion**.



Equation of free fall motion:

$$v = at$$

Abductive Reasoning and Newton's Discoveries

- How did Newton use **abductive reasoning** to discover the **law of universal gravitation**.

Observation

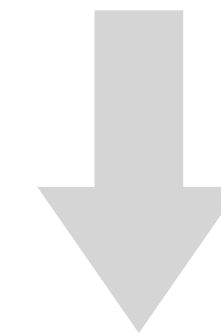


$$F = ?$$

Equation of the law of universal gravitation:

Hypothesis

It might be the **Earth** that **attracts apples**. This force is directed towards **the center of the Earth**.



$$F = G \frac{m_1 m_2}{r^2}$$

Abductive Reasoning and Einstein's Discoveries

- How did Einstein use **abductive reasoning** to discover the **equivalence principle**.

Observation

$$F = m_i a = G \frac{m_g M_g}{r^2}$$

The **inertial** mass is proportional to the **gravitational** mass.



$$m = ?$$

Hypothesis

Inertial mass and **gravitational** mass are **equivalent**.



Field equation:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi G T_{\mu\nu}$$



Curiosity, Imagination, Creativity

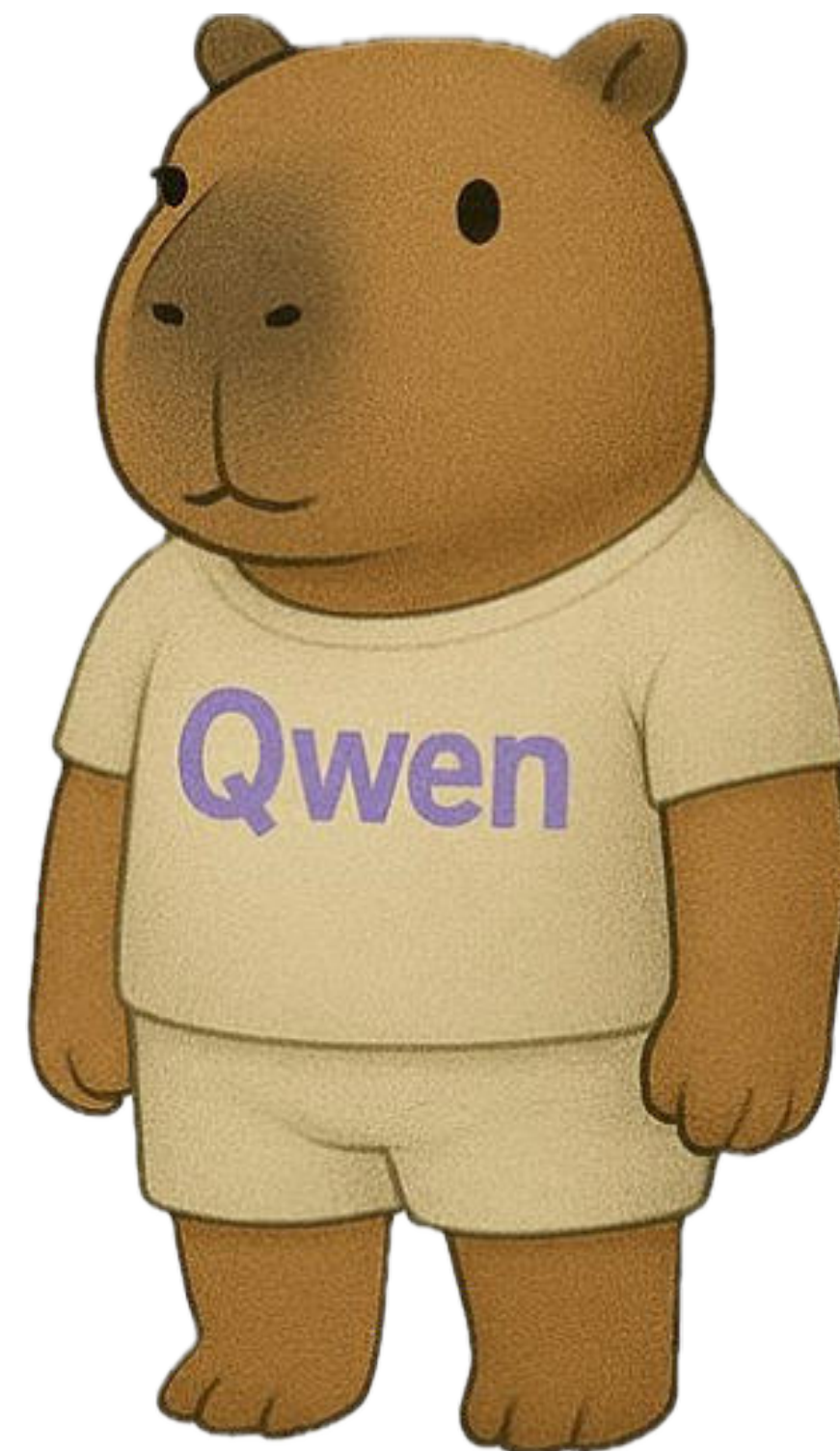
**From the drop of apples to the
discovery of general relativity.**

(1590—1915)

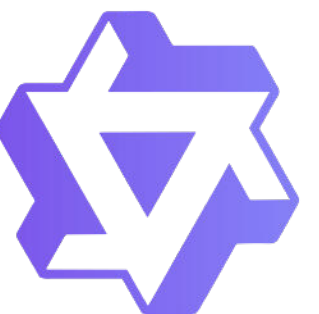


Can AI discover general relativity?

**There is an apple tree
ahead, An apple is falling...**



VLM



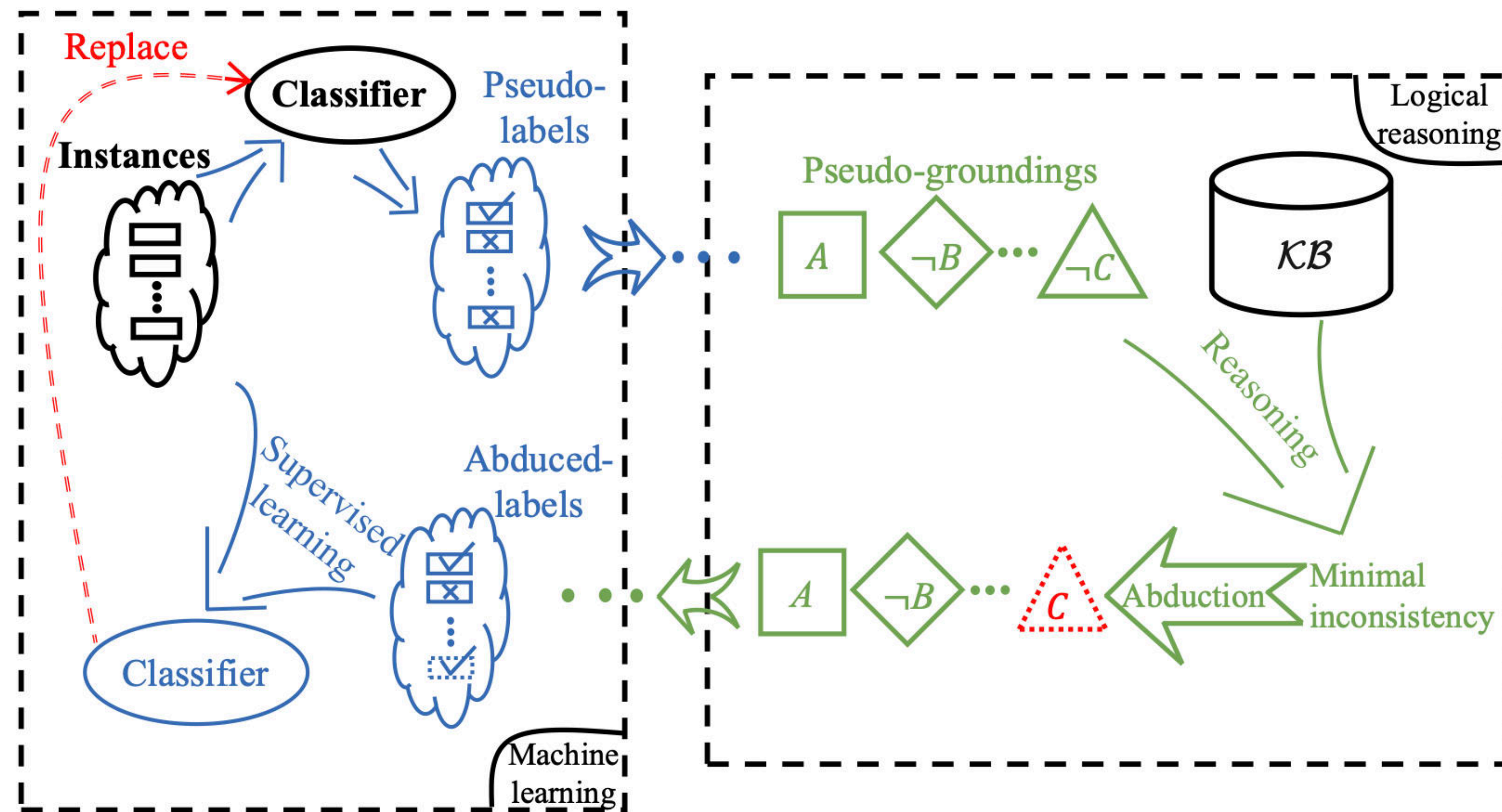
Abductive Reasoning and Brain

“The brain is an abduction machine, continuously trying to prove abductively that the observables in its environment constitute a coherent situation. ”

– Jerry Hobbs, ACL 2013 Lifetime Achievement Award

Abductive Reasoning and Machine Learning

- Machine learning and logical reasoning **converge**, Abductive learning where a classifier and a knowledge base are available.[Zhou et al, 2019]



Machine learning and logical reasoning framework

Abductive Reasoning and Machine Learning

- Formally, given $\{x_1, \dots, x_i\}$, C and KB , the task is to seek a **function** f such that.

$$\{x_1, \dots, x_i\}, f \triangleright \mathcal{O} \quad (1)$$

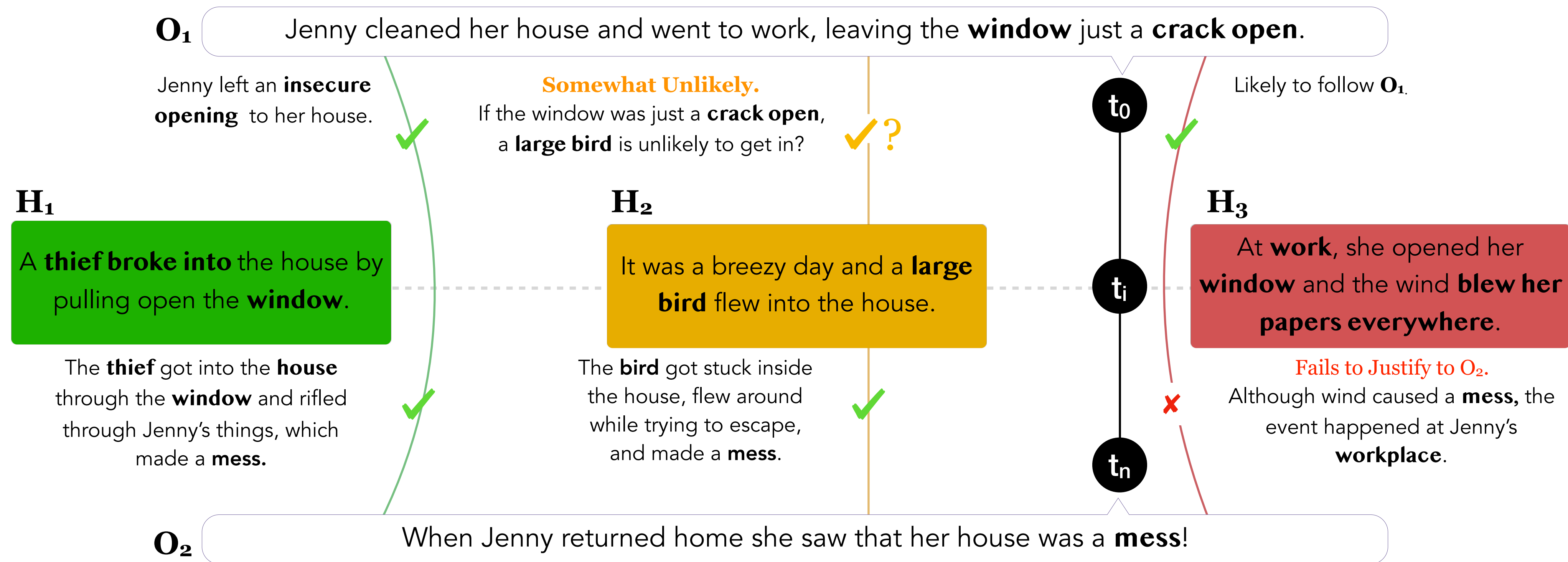
$$s.t. \quad KB \models \mathcal{O}, \quad \text{or} \quad (2)$$

$$KB \models \Delta(\mathcal{O}), \quad f \leftarrow \psi(f, \Delta(\mathcal{O})), \quad (3)$$

- $\mathcal{O}$ is compatible with KB , Otherwise, $\Delta(\mathcal{O})$ is compatible with KB after obtaining $\Delta(\mathcal{O})$ by logical backflow, and f is updated by the process ψ .

Abductive Reasoning for Natural Language

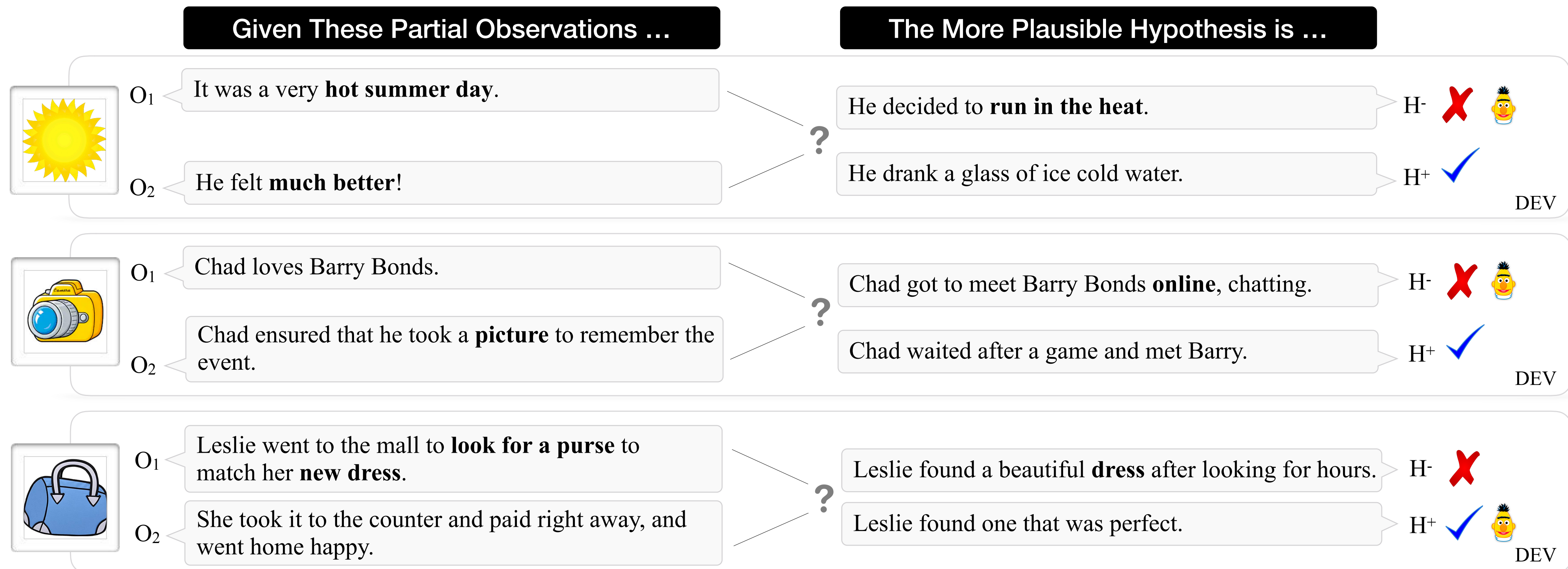
- The behavior of "**abductive learning**" is regarded as the **core** for people to interpret and understand natural language.[Bhagavatula et al, 2020]



Example of Abductive Reasoning.

Abductive Natural Language Inference

- Given the observations and a pair of hypotheses, the α NLI task is to select **the most plausible explanation (hypothesis)**.



examples from the α NLI task

Abductive Natural Language Inference

- **Performance of baselines** and finetuned-LM approaches on the test set of *ART*.

Model	GPT AF Acc. (%)	<i>ART</i> Acc. (%)
Random (2-way choice)	50.1	50.4
Majority (from <i>dev</i> set)	50.1	50.8
Infersent (Conneau et al., 2017)	50.9	50.8
ESIM+ELMo (Chen et al., 2017)	58.2	58.8
Finetuning Pre-trained LMs		
GPT-ft	52.6 (0.9)	63.1 (0.5)
BERT-ft [h^i Only]	55.9 (0.7)	59.5 (0.2)
BERT-ft [O_1 Only]	63.9 (0.8)	63.5 (0.7)
BERT-ft [O_2 Only]	68.1 (0.6)	66.6 (0.2)
BERT-ft [Linear Chain]	65.3 (1.4)	68.9 (0.5)
BERT-ft [Fully Connected]	72.0 (0.5)	68.6 (0.5)
Human Performance	-	91.4

Performance

Abductive Natural Language Generation

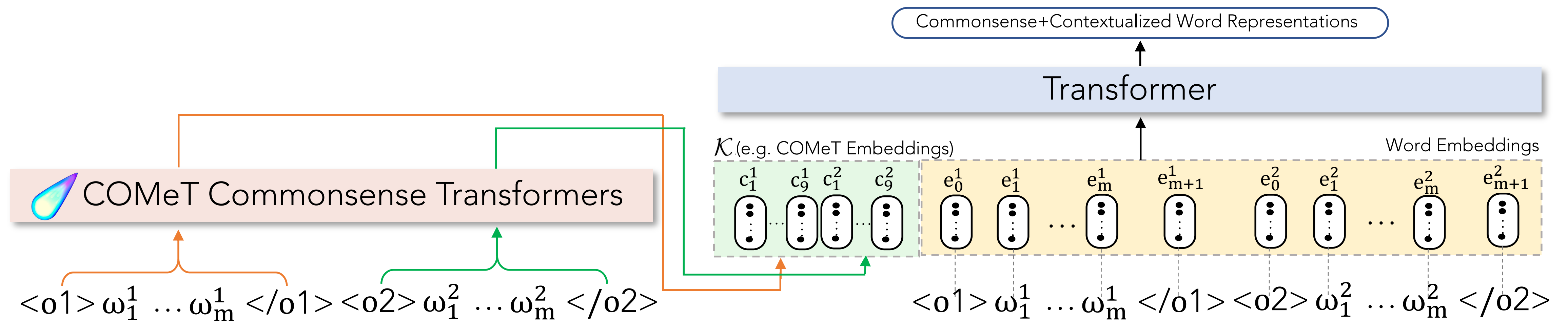
- Given a set of training data O_1, O_2 , and h^+ , the task of aNLG is to **maximize the generation probability**. The loss function of the model is constructed as follows:

$$\mathcal{L} = - \sum_{i=1}^N \log P \left(w_i^h \mid w_{<i}^h, w_1^{o1} \dots w_m^{o1}, w_1^{o2} \dots w_n^{o2}, K \right)$$

- $O_1 = \{w_1^{o1}, \dots, w_m^{o1}\}$, $O_2 = \{w_1^{o2}, \dots, w_n^{o2}\}$, $h^+ = \{w_1^h, \dots, w_l^h\}$.
- $w_{<i}^h$ represents the i tokens before the current position.
- w_i^h is the token at the current position i and K as background knowledge.

Abductive Natural Language Generation

- **Background commonsense knowledge** is efficiently integrated into the language model using embeddings, which are then decoded by GPT2.

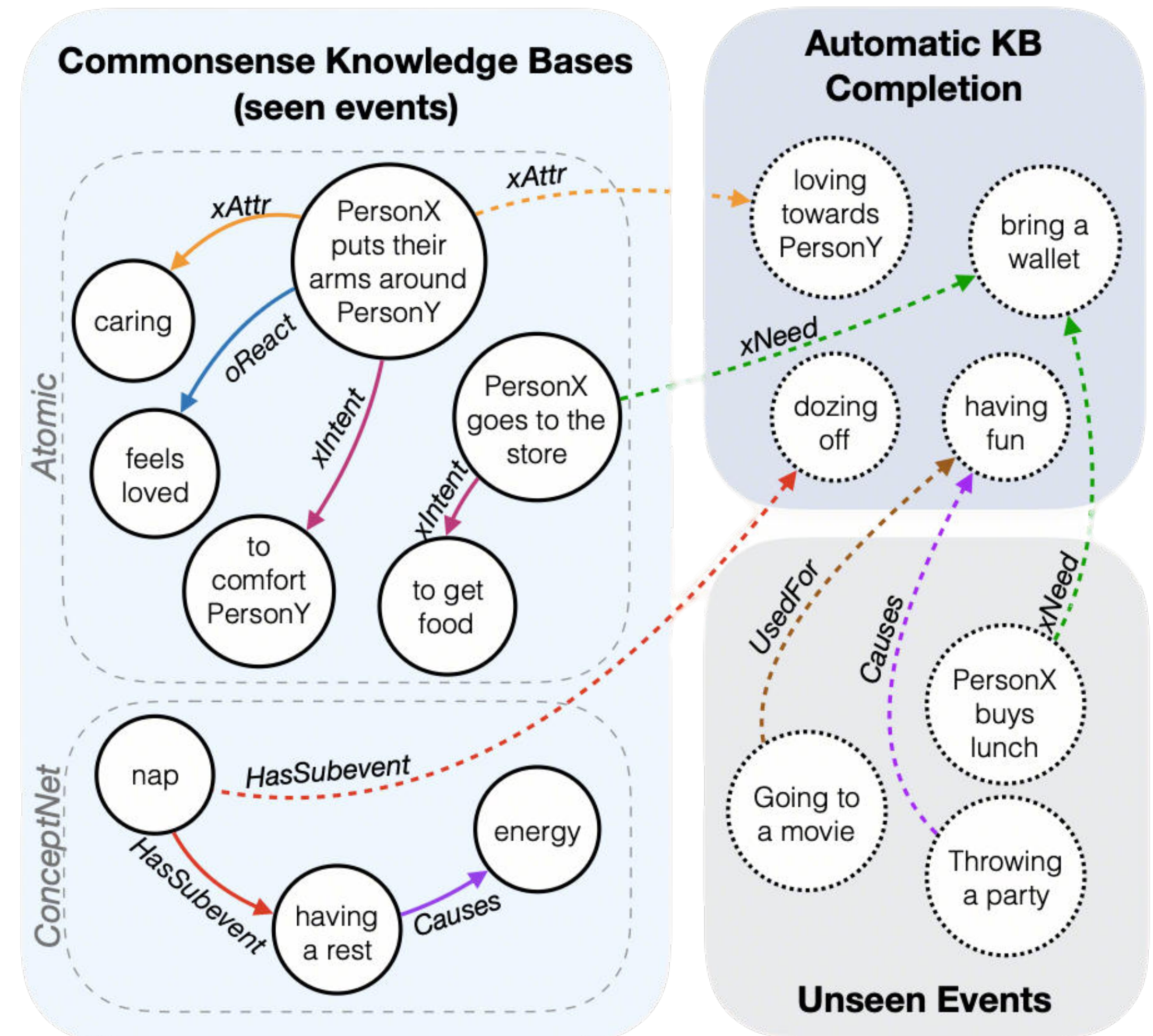


- **COMET** is a Transformer specifically designed to automatically construct common sense knowledge graphs.

COMET—Commonsense Transformers

- An adaptive generative model can assist the **self-complement of commonsense knowledge**. [Bosselut et al, 2019]

- xIntent → Motivation / Intention
- xNeed → Precondition / Requirement
- xAttr → Attribute / Trait
- xEffect → Outcome / Effect
- xReact → Emotion / Reaction
- xWant → Desire / Goal
- oEffect → Impact / Outcome
- oReact → Response / Emotion
- oWant → Desire / Goal



Abductive Natural Language Generation

- **Performance** of generative models on the test set of *ART*. All models except GPT2-Fixed are finetuned on *ART*.

Model	BLEU	METEOR	ROUGE	CIDEr	BERT-Score	Human
GPT2-Fixed	0.0	9.29	9.99	3.34	36.69	-
O_1 - O_2 -Only	2.23	16.71	22.83	33.54	48.74	42.26
COMeT-Txt+GPT2	2.29	16.73	22.51	31.99	48.46	38.28
COMeT-Emb+GPT2	3.03	17.66	22.93	32.00	48.52	44.56
Human-written Hypotheses	8.25	26.71	30.40	53.56	53.30	96.03

Performance

- **Can we give a better model of abductive learning in NLP?**

Reasoning is about giving reasons

- Addressing reasoning by identifying an intermediate representation (**Representation of the Logical Structure**). [Shah et al, 2025]

fact1: Harry is young, nice, and kind. He works outside and is cold and blue.

rule1: Nice blue people are usually round in shape.

query: Harry is round.

Sentence level explanation:

fact1 + rule1 → Harry is round.

Logical structure explanation:

(Harry, is, nice) + (Harry, is, blue) → (Harry, is, round)

$$(X, \text{is, blue}) \wedge (X, \text{is, nice}) \rightarrow (X, \text{is, round})$$

Example:

fact1: Bill is an ostrich.

query: Bill is not an ostrich.

Predicted RLS:

(Bill, is, ostrich, +)

(Bill, is, an ostrich, -)

An example mistake of our method on the Birds dataset,

Abductive Reasoning for Knowledge Graphs

- **Abductive reasoning in knowledge graphs** aims to generate plausible logical **hypotheses** from observed entities [Gao et al, 2025]

Observation

Systemic Lupus Erythematosus, Antiphospholipid Syndrome, Sjögren's Syndrome

Hypothesis 1:

$H = V_? : \text{Produce}(V_?, \text{Autoantibodies}) \wedge \text{Property}(V_?, \text{Chronic}) \wedge \text{Property}(V_?, \text{Autoimmune})$

Interpretation:

The disease that produces autoantibodies and is chronic and autoimmune.

Hypothesis 2:

$H = V_? : \text{Produce}(V_?, \text{Autoantibodies}) \wedge \text{Property}(V_?, \text{Chronic})$

Interpretation:

The disease that are autoimmune and can be treated with hydroxychloroquine.

Hypothesis 3:

$H = V_? : \text{Produce}(V_?, \text{Autoantibodies}) \wedge \text{Property}(V_?, \text{Chronic}) \wedge \text{Age}(P_?, \text{Reproductive}) \wedge \text{Carries}(P_?, \text{HLA-DR3})$

Interpretation:

The disease that is more likely to occur in women of reproductive age who carry the HLA-DR3 susceptible allele.

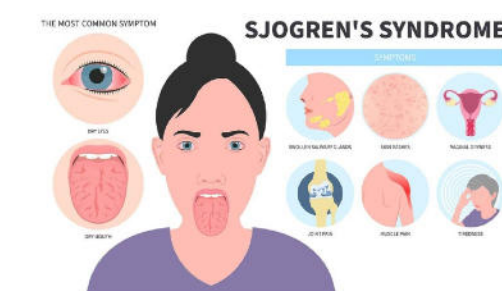
Pathology



Treatment



Susceptibles



Observation

Chris Paul, James Harden, Russel Westbrook

Hypothesis 1:

$H = V_? : \text{PlayedFor}(V_?, \text{Thunder}) \wedge \text{Won}(V_?, \text{AssistTitle})$

Interpretation:

Players who have played for the Oklahoma City Thunder and have won the NBA assist title.

Hypothesis 2:

$H = V_? : \text{PlayedFor}(V_?, \text{Thunder}) \wedge \text{Won}(V_?, \text{AssistTitle}) \wedge \neg \text{Won}(V_?, \text{NBACHampionship})$

Interpretation:

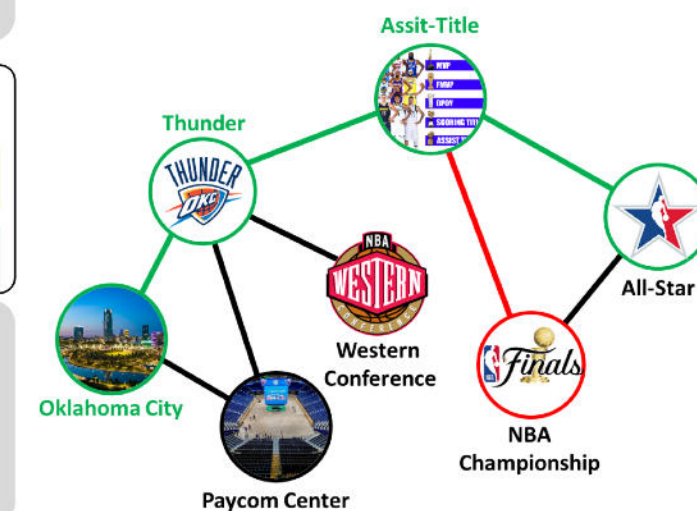
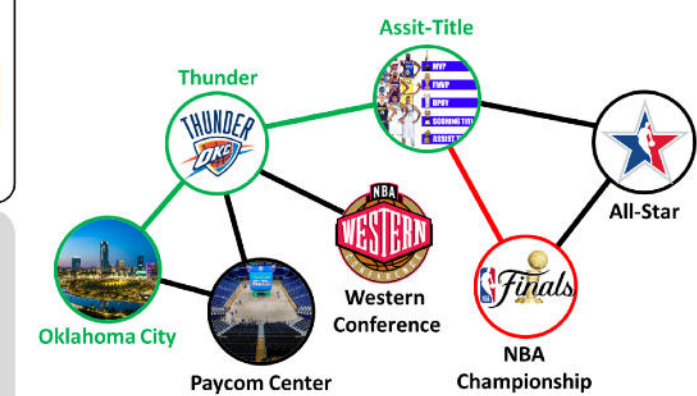
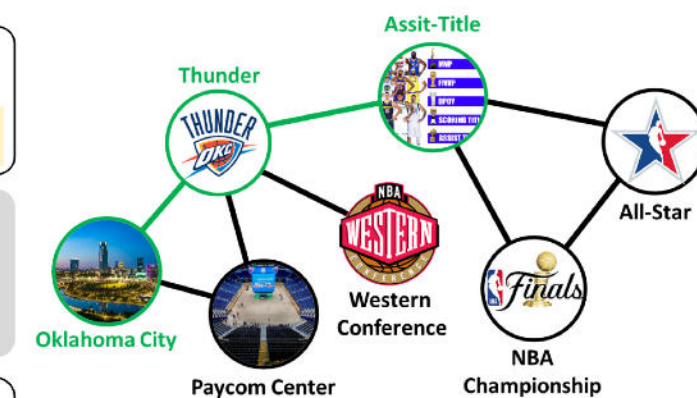
Players who have played for the Thunder, have won the NBA assist title, and have never won an NBA championship.

Hypothesis 3:

$H = V_? : \text{PlayedFor}(V_?, \text{Thunder}) \wedge \text{Won}(V_?, \text{AssistTitle}) \wedge \neg \text{Won}(V_?, \text{NBACHampionship}) \wedge \text{selected}(V_?, \text{ALL_Star})$

Interpretation:

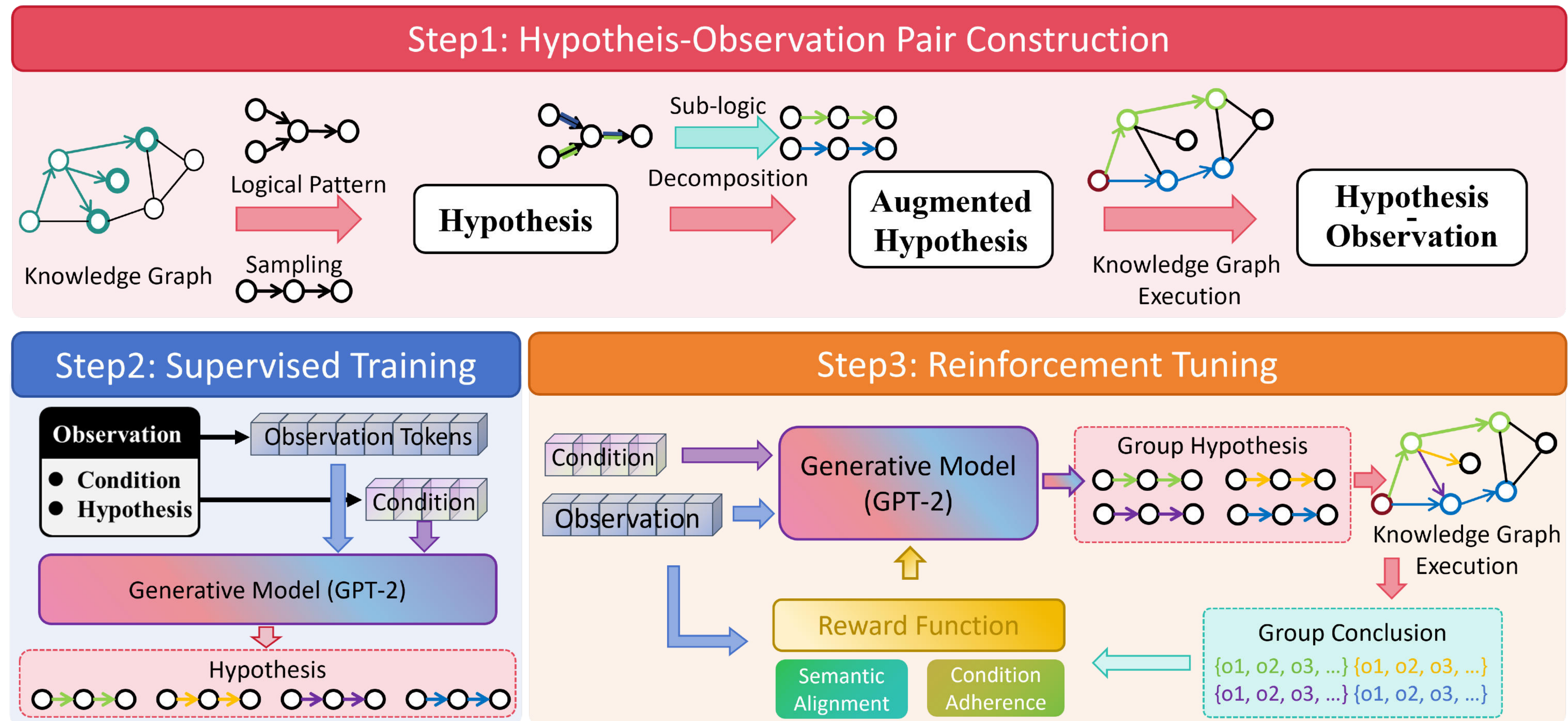
Players who have played for the Thunder, have won the NBA assist title, have never won an NBA championship, but have been selected as NBA All-Stars.



Examples of Controllability in Abductive Reasoning

Abductive Reasoning for Knowledge Graphs

- An overview of the **controllable abductive reasoning framework** is shown in the following figure.



controllable abductive reasoning framework

Abductive Reasoning and Subjective Experience

“I experienced the ‘subjective experience’ of pink elephants floating before me. Those pink elephants are a hypothesis about the external world — a hypothesis that accounts for my present internal state. The peculiarity of subjective experience lies not in some mysterious substance, but in its hypothetical nature.”

– Geoffrey Hinton, The 2024 Nobel Prize in Physics

The Abduction of Sherlock Holmes

- Can machines perform visual reasoning by **identifying concrete visual clues scattered throughout a scene?** [Hessel et al, 2022]

What can you **infer** from the **visual clues** ?



large semi truck and trailer on its side laying on a freeway

There was a major accident that occurred minutes ago

The people are inspecting damage to the vehicles in the accident

patches of snow spread throughout grass on the side of freeway

Cold weather is causing hazardous conditions at this location

The roads are very icy

a white license plate with five red English style numbers displayed

This accident happened in an English speaking country

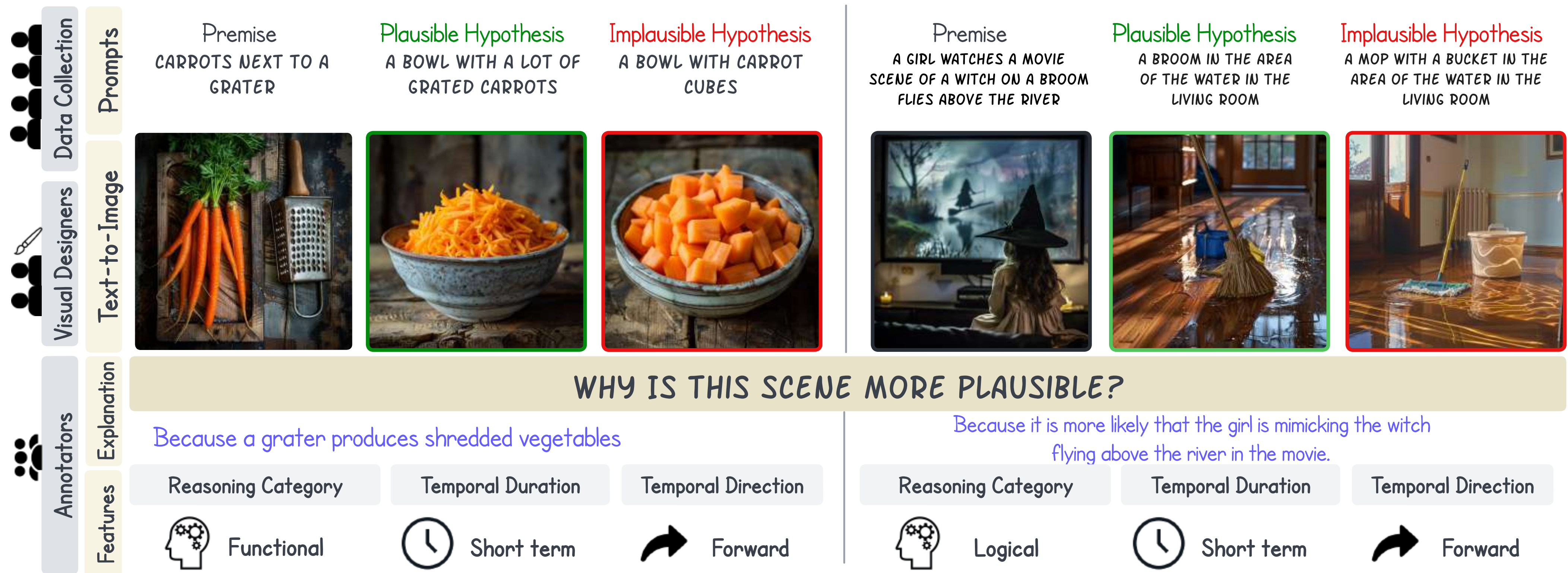
This is Ohio

A Dataset for Visual Abductive Reasoning



Abductive Reasoning for Images

- **NL-EYE** evaluates the abductive reasoning capabilities of VLMs.
[Ventura et al, 2024]



NL-EYE benchmark

Abductive Reasoning for Images

- VLMs fail to perform abductive reasoning with images, The results **reveal a large performance gap** between humans and VLMs on both tasks.

Input Strategy	Model	Prediction		Explanation	
		Triplet	Pairs	Human	Auto
	Humans	85%	83%	95%	—
Separate Images	Gemini-1.5-Pro	51%	42%	38%	34%
	GPT-4-Vision	46%	40%	39%	37%
	GPT-4o	16%	50%	23%	23%
	Claude-Sonnet-3.5	49%	38%	50%	26%
Combined Image	Gemini-1.5-Pro	43%	39%	40%	33%
	GPT-4-Vision	41%	34%	37%	27%
	GPT-4o	60%	45%	44%	40%
	Claude-Sonnet-3.5	28%	36%	42%	21%
	Claude-Opus-3	15%	33%	26%	6%
	LLaVA 1.6	14%	42%	15%	4%
	Fuyu	4%	44%	10%	2%
	Random	25%	45%	—	—
	Dumb Pixel	50%	50%	—	—

scores for vision-based experiments

VLMs Failure Analysis

- VLMs fail to perform abductive reasoning with images, **The Failure is in Visual Interpretation.**
- VLMs produce explanations that describe different reasoning than humans.

Reasoning Approach	Describer	Predictor	Prediction Triplet
Text-Only	Gold	Gemini-1.5-Pro	66%
		GPT-4o	80%
		GPT-4	78%
		Claude-Sonnet-3.5	79%
		Claude-Opus-3	81%
		BART L mnli	68%
		DeBERTa nli v3	65%
Image-to-Text	GPT-4o	Gemini-1.5-Pro	29%
		GPT-4 vision	32%
		Claude 3.5	39%
		Claude 3	33%
		LLaVA 1.6	29%
		BLIP 2	40%
		Instruct BLIP	35%

Performance for prediction in the triplet setup

VLMs Failure Analysis

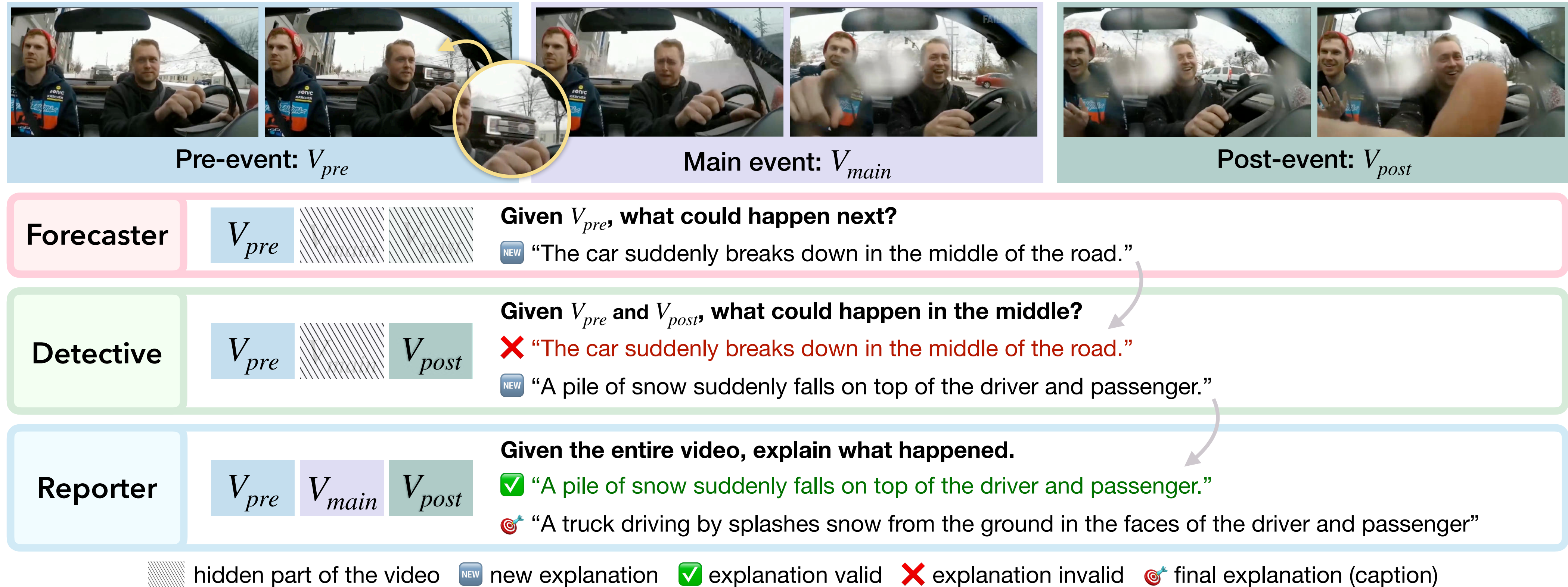
- VLMs fail to perform abductive reasoning with images, **VLMs Predictions Depend on Hypothesis Location.**
- VLMs rely on weak visual encoding, capturing superficial patterns like hypothesis order rather than meaningful image content.

Baseline	Model	Consistency Acc.	Order 1 Acc.	Order 2 Acc.
Separate-Images	Gemini-1.5-pro	50.57%	56.0%	75.14%
	GPT-4-vision	45.71%	59.14%	66.0%
	GPT-4o	16.29%	97.43%	16.29%
	Claude-3.5-sonnet	49.28%	65.9%	59.31%
Combined-Image	Claude-3-opus	15.14%	30.57%	72.86%
	Claude-3.5-sonnet	28.29%	76.86%	30.29%
	Llava-mistral-7b	14.86%	53.14%	26.0%
	Gemini-1.5-pro	42.57%	46.29%	81.43%
	GPT-4-vision	41.14%	54.57%	67.14%
	GPT-4o	60.0%	76.0%	69.14%
	Fuyu-8b	4.58%	42.98%	13.75%
Text-Only	GPT-4	78.0%	86.86%	82.86%
	GPT-4o	80.0%	83.14%	88.57%
	Gemini-1.5-pro	65.8%	72.99%	79.02%
	Claude-opus-3	80.57%	85.43%	87.43%
	Claude-sonnet-3.5	79.43%	82.29%	89.43%
	Bart L mnli	68.0%	68.0%	68.0%
	DeBERTa-v3	65%	65%	65%

Plausibility prediction results of the triplet setup

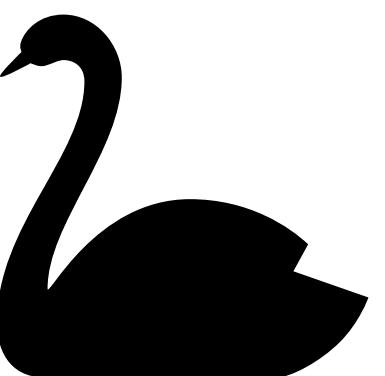
Abductive Reasoning for Videos

- **BlackSwanSuite** evaluates both **abductive** and **defeasible** reasoning with unexpected events. [Chinchure et al, 2024]



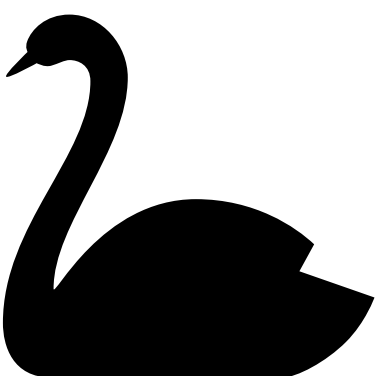
benchmark involves three tasks

<https://arxiv.org/abs/2412.05725>



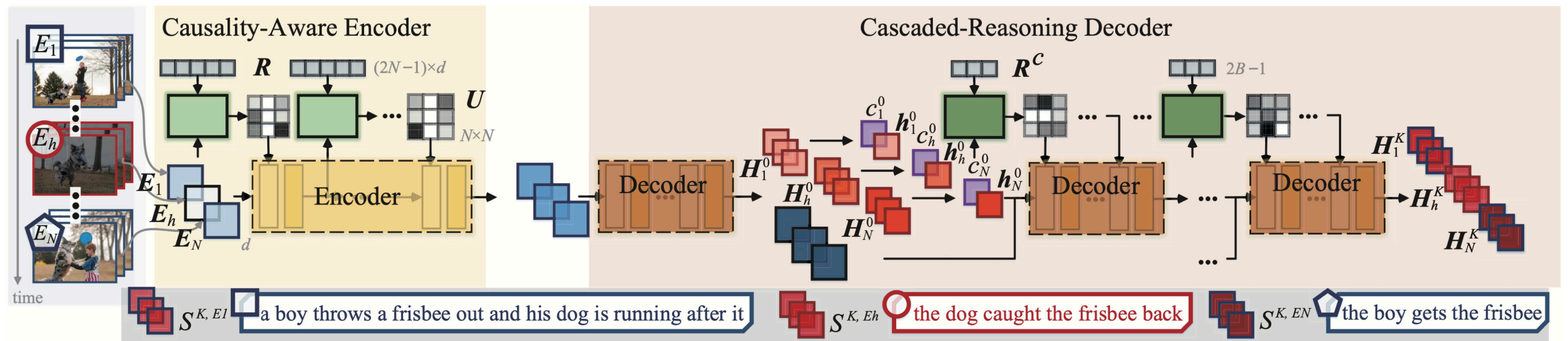
Perception, Comprehension, Reasoning

- Answering an abductive reasoning question (**Detective**) requires three key steps:
 - perception of objects, people, and actions in the videos.
 - comprehending the flow of events in the video based on the differences between V_{pre} and V_{post} .
 - abductive reasoning about what could be happening in the middle.
- Can we give a better model of abductive learning in VLM ?



Visual Abductive Reasoning

- **REASONER** (causal-and-cascaded reasoning Transformer). surpasses many famous video-language models.[Liang et al, 2022]



Network architecture of REASONER

Abductive Reasoning and Simulated Universe

“The world is doing what it does, with all sorts of computational irreducibility. But somehow even with the computationally bounded resources of our brains we’re able to form some kind of coherent model of what’s going on. ”

– Stephen Wolfram, Wolfram Physics Project

Looking ahead...

- Can LLM models **think like Albert Einstein** ?
- Can Artificial Intelligence **have subjective experiences**?
- Can humans **simulate the universe** ?

