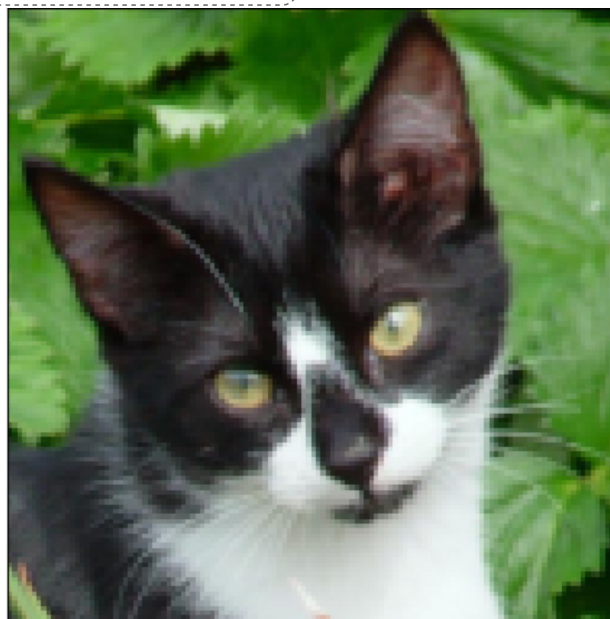
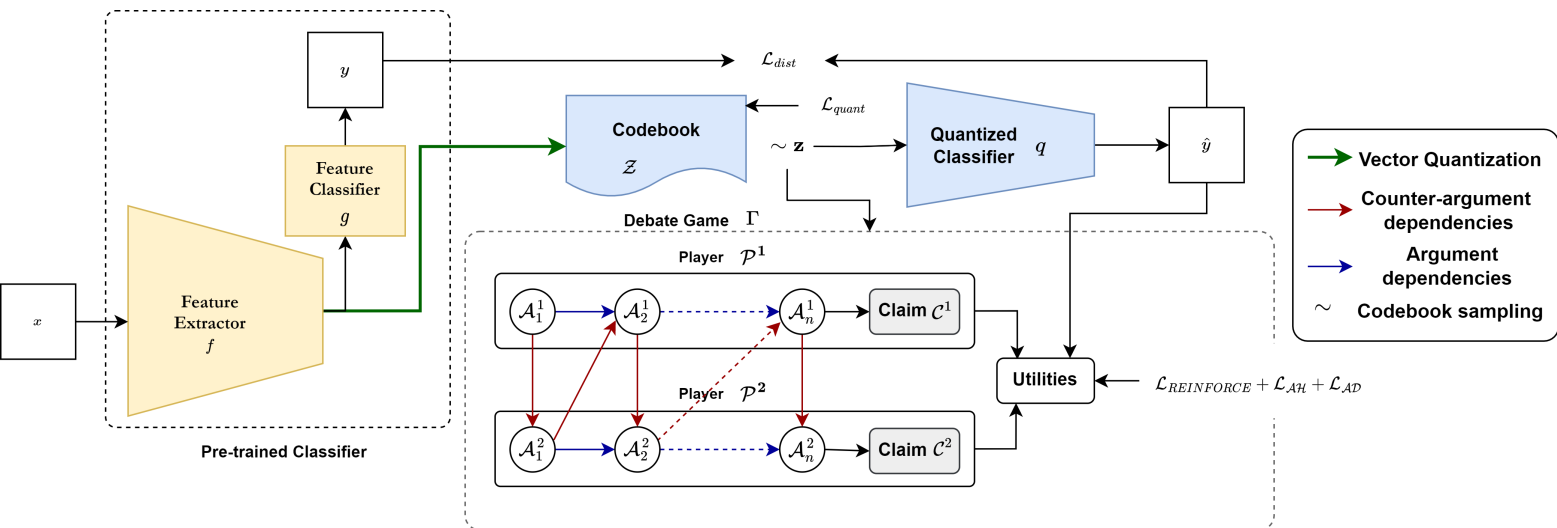


Visual Debates

De

(Kori et al 2024)



Argument z_{38} , with strength 1 and claim Cat (Description: pointy ears).

Argument z_8 , with strength -1 and claim Dog (Description: Fur).

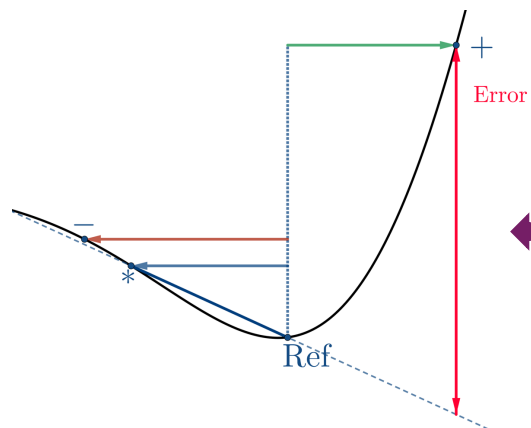
Argument z_{42} , with strength 1 and claim Cat (Description: eyes).

Argument z_{11} , with strength -1 and claim Dog (Description: Forehead/nose).

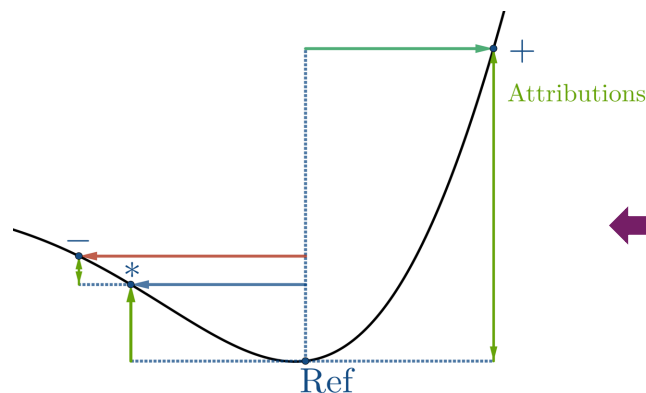
CAFE: Conflict-Aware Feature-wise Explanations

(Dejl et al 2024)

A feature attribution method addressing the limitations of gradient-based methods in handling conflicts.

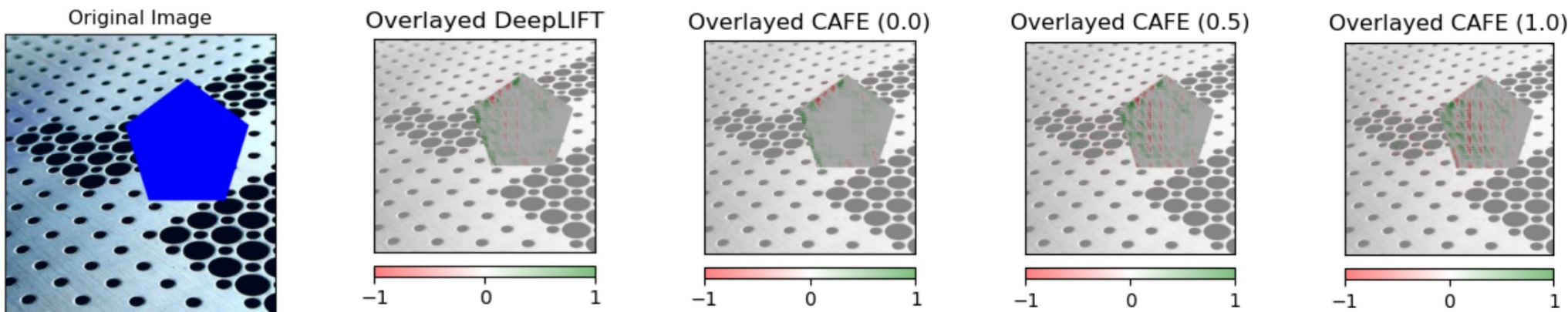


Local gradient-based approximation ignores the potential effect of the positive feature.



CAFE separately considers the effects of conflicting features, resulting in more accurate attribution scores.

Initial experiments demonstrate the ability of CAFE to surface conflicts in image predictions (e.g., the internal areas highlighted negatively by CAFE 1.0 in the image below might also be seen in other shapes).



Evaluating Explanations in Fact Verification

(Kotonya, Toni2024)

p_3 : A daffodil plant can live for more than two years.

Verdict: Verified

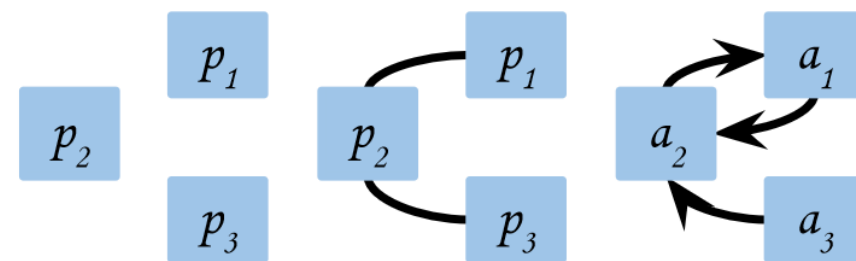
Explanation $\langle \mathcal{P}, \mathcal{R} \rangle$, where:

$\mathcal{P} = \{p_1, p_2, p_3\}$, for:

p_1 : Daffodil is the common name for plants of the narcissus genus, which are perennial.

p_2 : A perennial plant has a minimum life span of two years.

$\mathcal{R} = \{(p_1, p_2), (p_2, p_3)\}$.



(a) Free-form (b) Deductive (c) Argumentative

Figure 1: Abstract illustrations of the three classes of explanations explored in this paper (where the p_i are propositions and the a_j are arguments).

An argumentative explanation is *dialectically faithful* if whenever the model predicts with

- top confidence, no arguments attack any argument for the prediction;
- high confidence, arguments for the prediction are stronger than arguments against it;
- low confidence, there must be only weak arguments for the prediction or strong arguments against the prediction

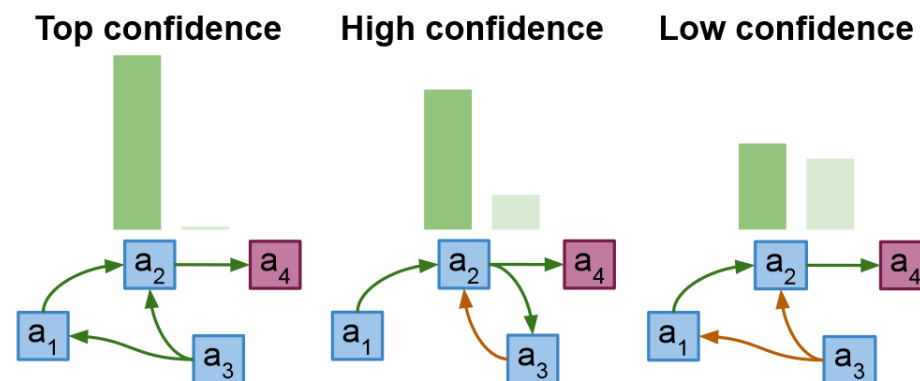
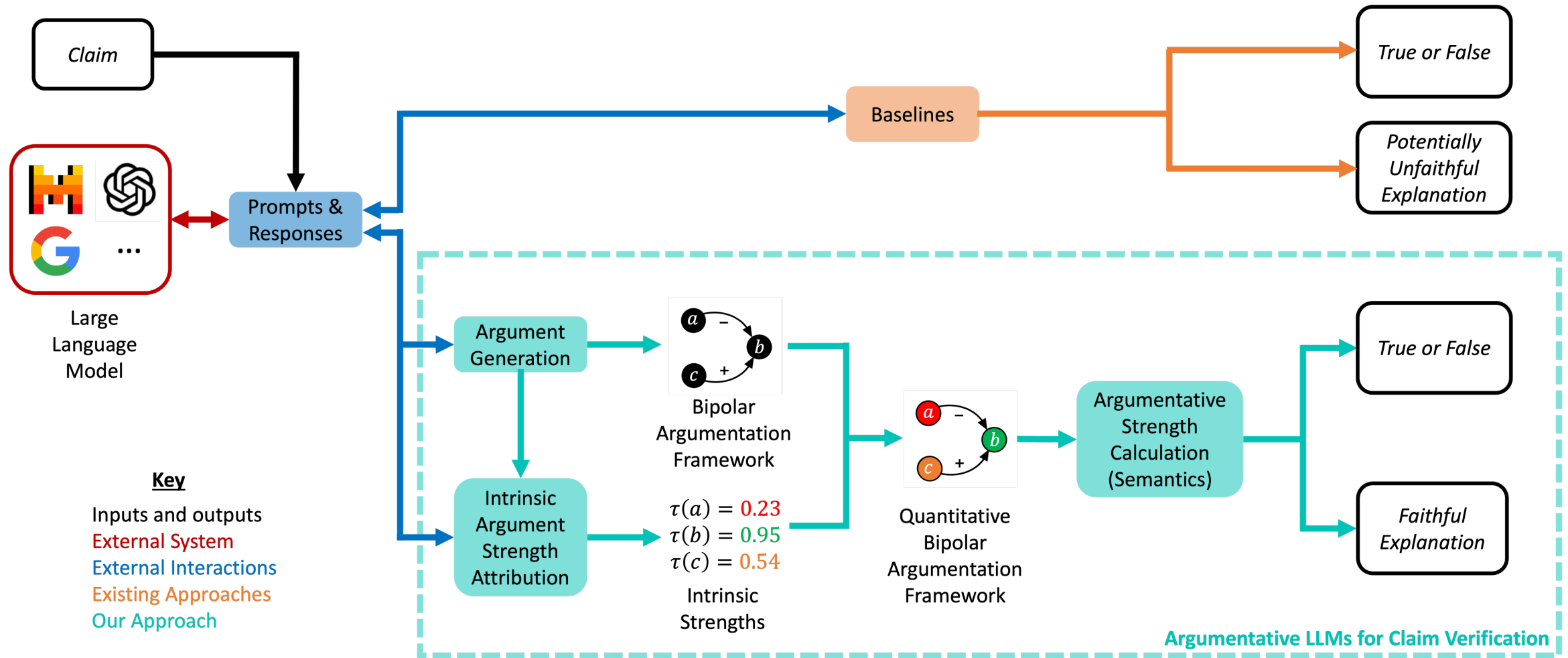


Figure 3: An illustration of argumentative explanations for top, high, and low confidence (binary) predictions. Attacks are shown in orange and supports are shown in green. Argument a_4 with conclusion \hat{y} is purple.

Argumentative LLMs

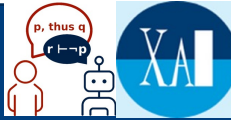
De

(Freedman et al 2024)



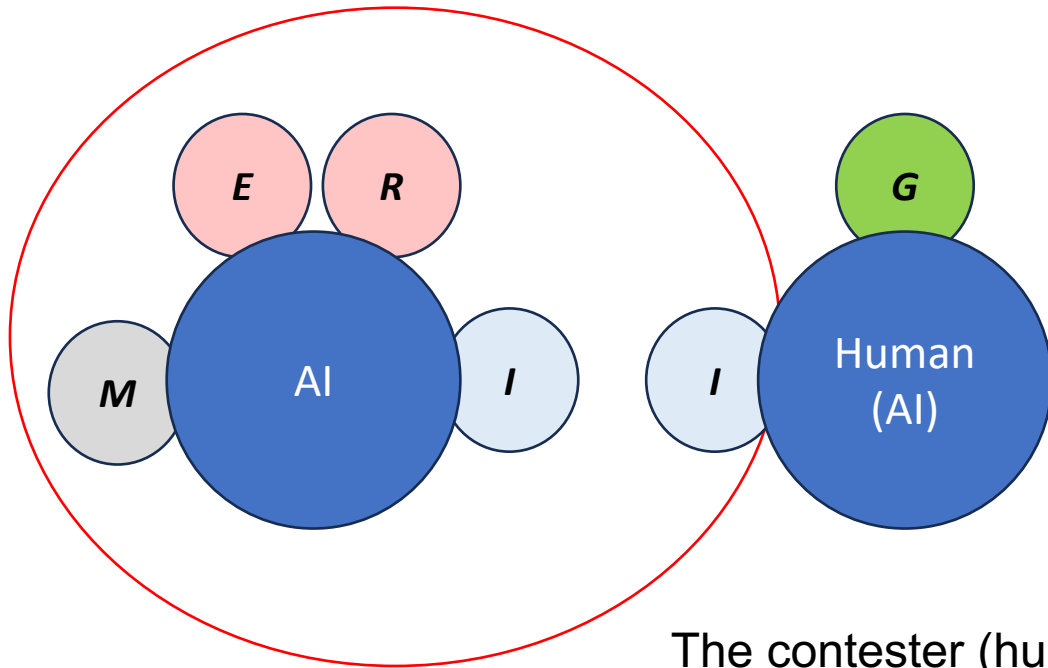
Contestable AI

De



(Leofante et al 2024, Freedman et al 2024)

Both contested AI and contesters are able to interact (I)



The contesteer (human or AI) is equipped with

- a ground generator for contestations (G)

The contested AI is equipped with

- a model (M)
- an explanation method (E)
- a redress method (R)



Article 22(3): ... the data controller shall implement suitable measures to safeguard the data subject's ... right to *obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.*

UK pro-innovation framework



Principle 5: **Contestability and redress**

Claim 'Birds are important to badminton'

Support

Attack

While it might seem unlikely that birds would have any significance to the sport of badminton, there is still a connection to be made. One important aspect of badminton is the shuttlecock, the cone-shaped projectile used to hit the net. Interestingly, the shape of the shuttlecock was inspired by the feathered design of bird's wings, specifically that of a bird called the Chinese goose. The structure of the feathers on the shuttlecock is meant to mimic the aerodynamics of a bird's wing in flight, allowing the shuttlecock to travel smoothly and swiftly.

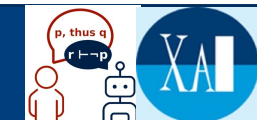
While birds are fascinating creatures and can sometimes be found near badminton courts, the claim that birds are important to badminton is tenuous at best. The sport of badminton, which involves hitting a shuttlecock over a net, does not require the presence of birds. In fact, the shuttlecock used in badminton is not bird-related at all; it is called a "shuttlecock" because the shuttle-shaped projectile is historically made with feathers, but these are goose or duck feathers, not those of birds typically associated with badminton such as sparrows.

0.7

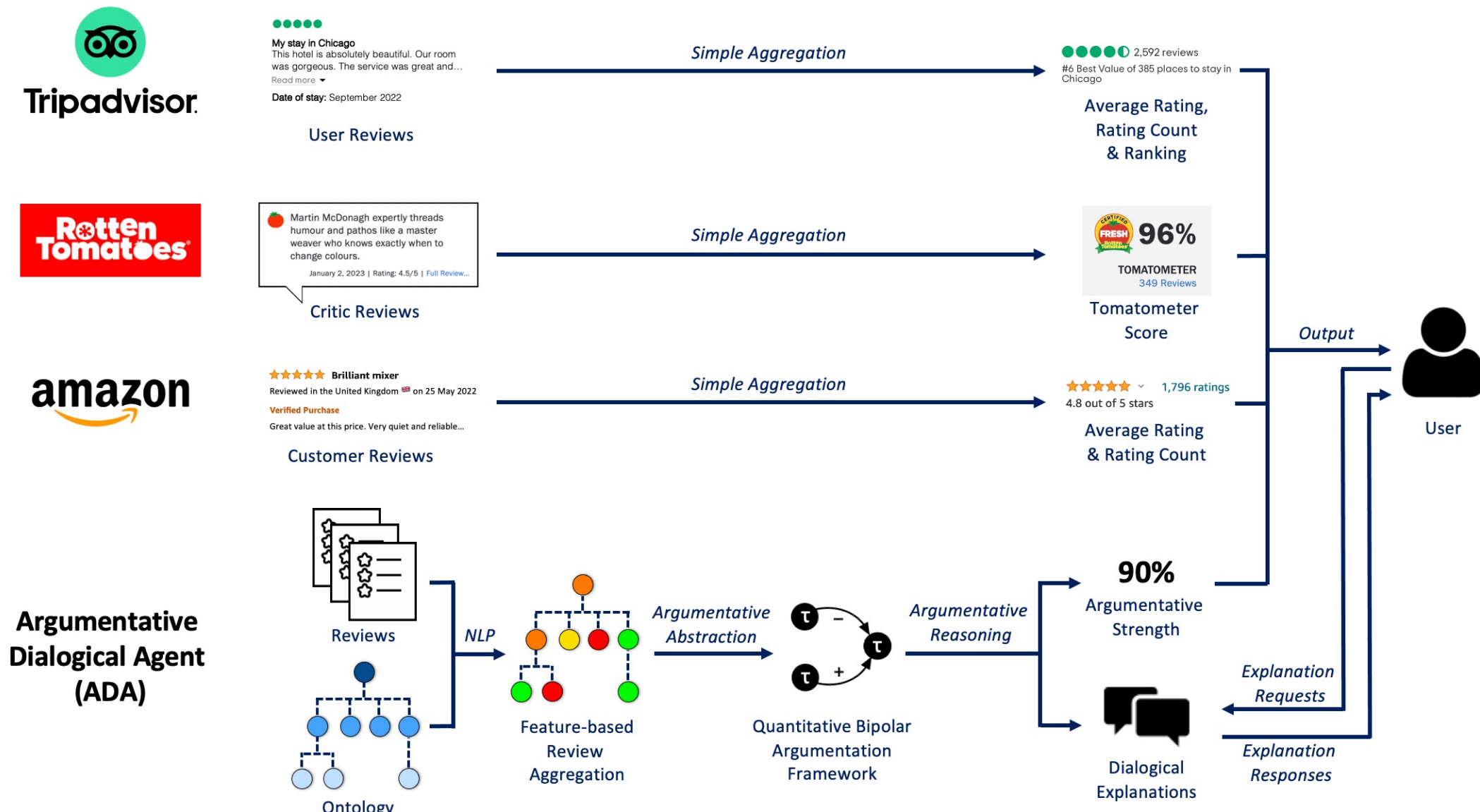
True

0.9 → 0.5

ADA: Argumentative Dialogical Agents

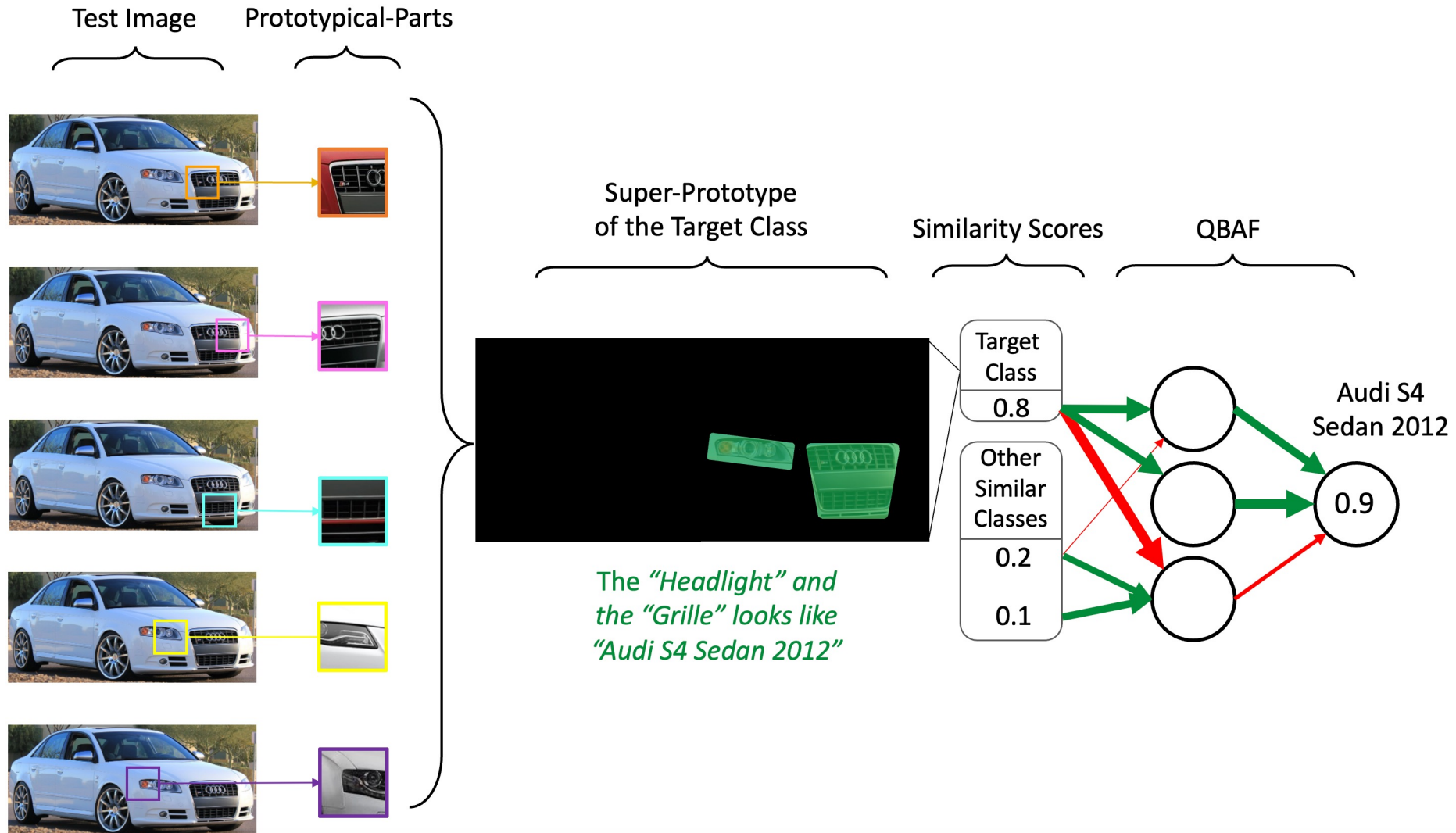


Cocarascu, Rago, Toni AAMAS19
Oksanen, Cocarascu, Toni KRHI22



ProtoArgNet

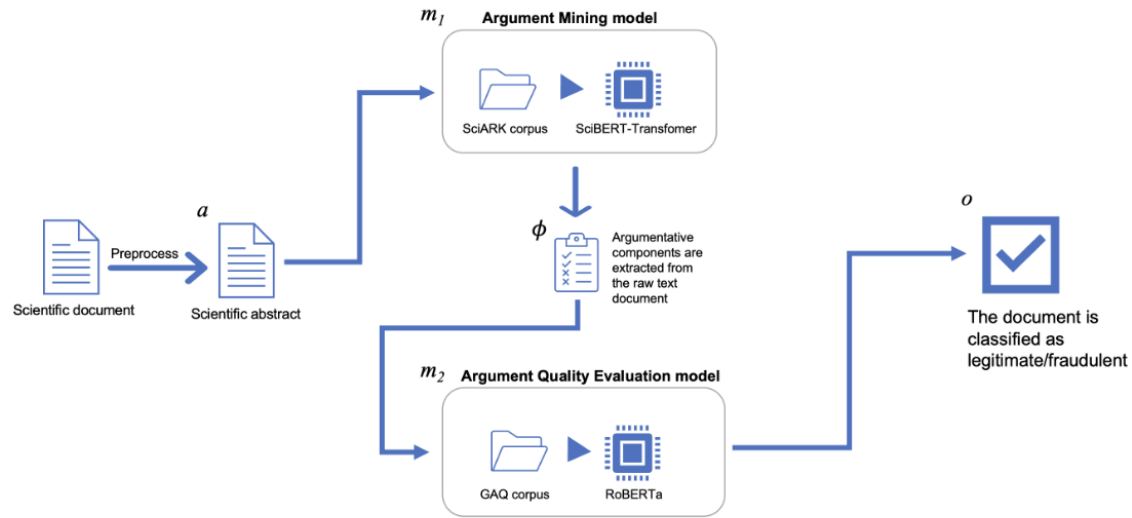
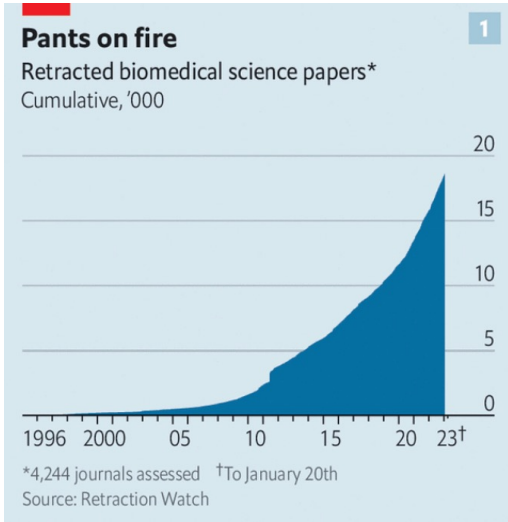
(Ayooobi et al 2024)



Veracity prediction

De

(Chen, Freedman, Toni, ongoing)

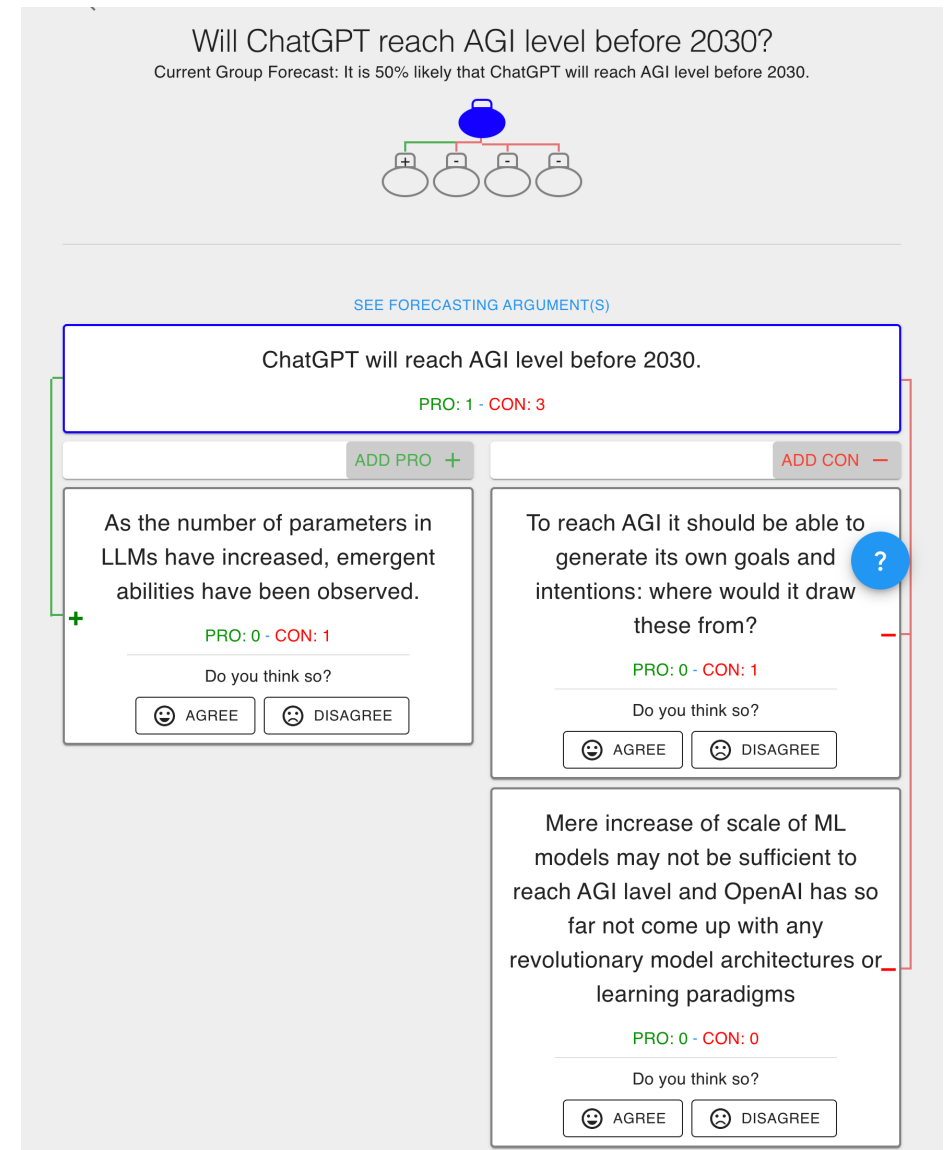


Metadata: Authors, Institution, Journal, etc

Explanation?

ArguCast supports **Judgmental Forecasting** (users making predictions for future events) using **Computational Argumentation**:

- Accommodates **debate** between users with arguments that **support** (are in favour) or **attack** (are against) other arguments/forecasting questions,
- Elicits **opinions** (by *votes*) and **predictions** from users, and
- Check if users' opinions and predictions are **coherent** and **filter out** incoherent predictions to get rid of the cognitive biases of users.



Will ChatGPT reach AGI level before 2030?
Current Group Forecast: It is 50% likely that ChatGPT will reach AGI level before 2030.

SEE FORECASTING ARGUMENT(S)

ChatGPT will reach AGI level before 2030.
PRO: 1 - CON: 3

ADD PRO + ADD CON -

As the number of parameters in LLMs have increased, emergent abilities have been observed.
PRO: 0 - CON: 1
Do you think so?
AGREE DISAGREE

To reach AGI it should be able to generate its own goals and intentions: where would it draw these from?
PRO: 0 - CON: 1
Do you think so?
AGREE DISAGREE

Mere increase of scale of ML models may not be sufficient to reach AGI level and OpenAI has so far not come up with any revolutionary model architectures or learning paradigms
PRO: 0 - CON: 0
Do you think so?
AGREE DISAGREE

Explaining in Quantified Bipolar Argumentation

(Yin et al 2023,24)

Arguments:

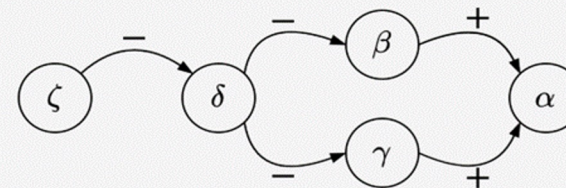
α : It is easy for children to learn a foreign language well.

β : Studies show that young children possess higher neuroplasticity, making language learning more effective.

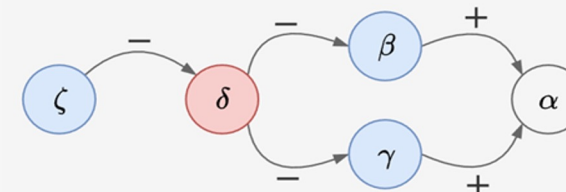
γ : Children immersed in a foreign language environment from an early age have better language acquisition.

δ : Learning a foreign language requires cognitive maturity, which children lack. Hence, it's difficult.

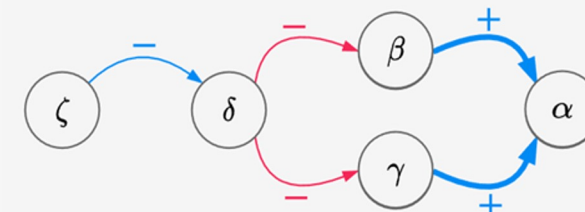
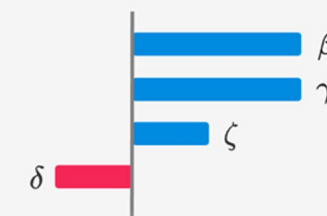
ζ : Children's brains are highly adaptable, making them more effective at absorbing new languages than adults.



Base score	Strength
$\tau(\alpha) = 0.5$	$\sigma(\alpha) = 0.80$
$\tau(\beta) = 0.5$	$\sigma(\beta) = 0.38$
$\tau(\gamma) = 0.5$	$\sigma(\gamma) = 0.38$
$\tau(\delta) = 0.5$	$\sigma(\delta) = 0.25$
$\tau(\zeta) = 0.5$	$\sigma(\zeta) = 0.50$



Argument Attribution (Gradient-based)

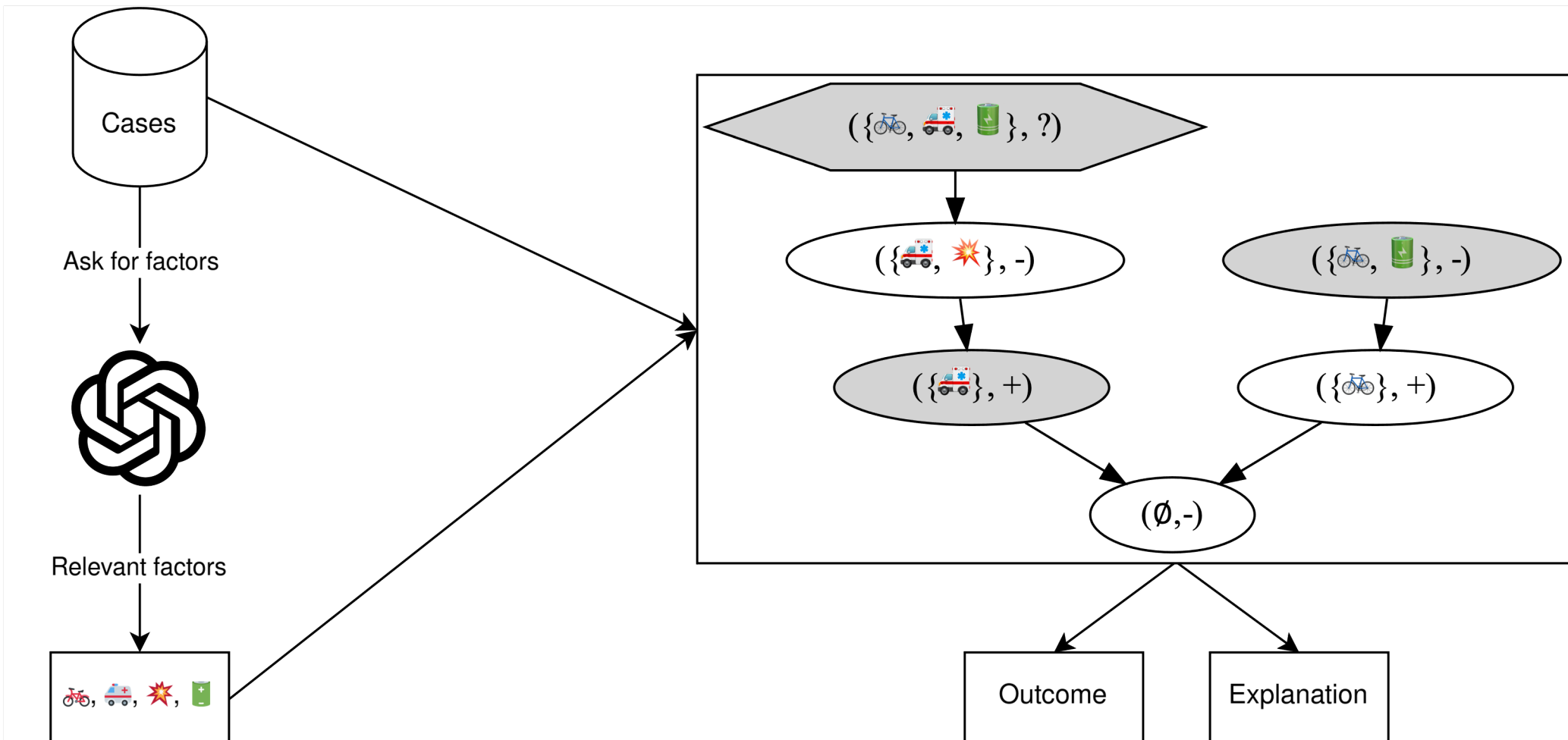


Relation Attribution (Shapley-based)



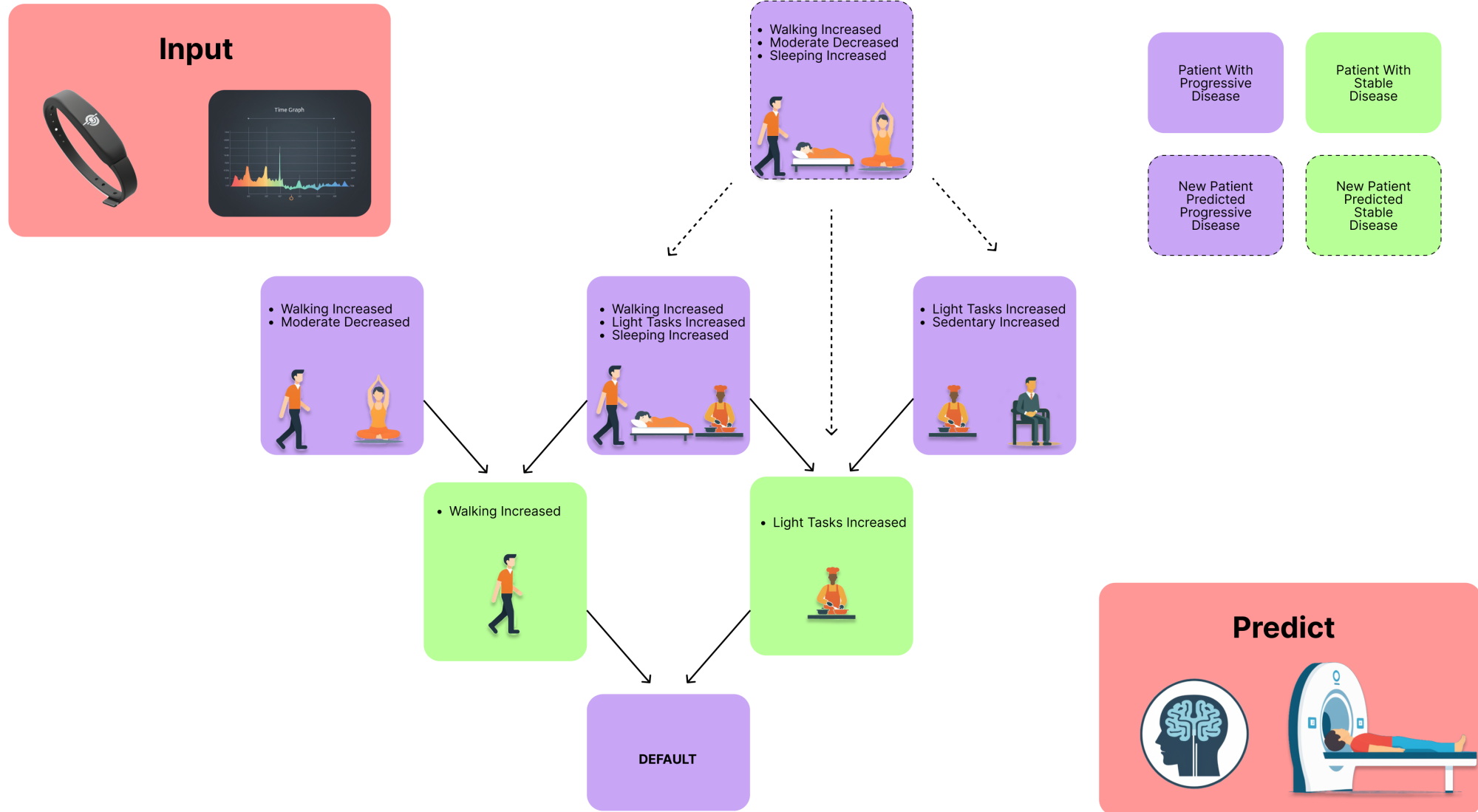
Argumentation and LLMs for legal reasoning

(Paulino-Passos and Toni 21, on-going)



Argumentation and Brain Cancer

(Gould et al, ongoing)



Argumentation for image classification

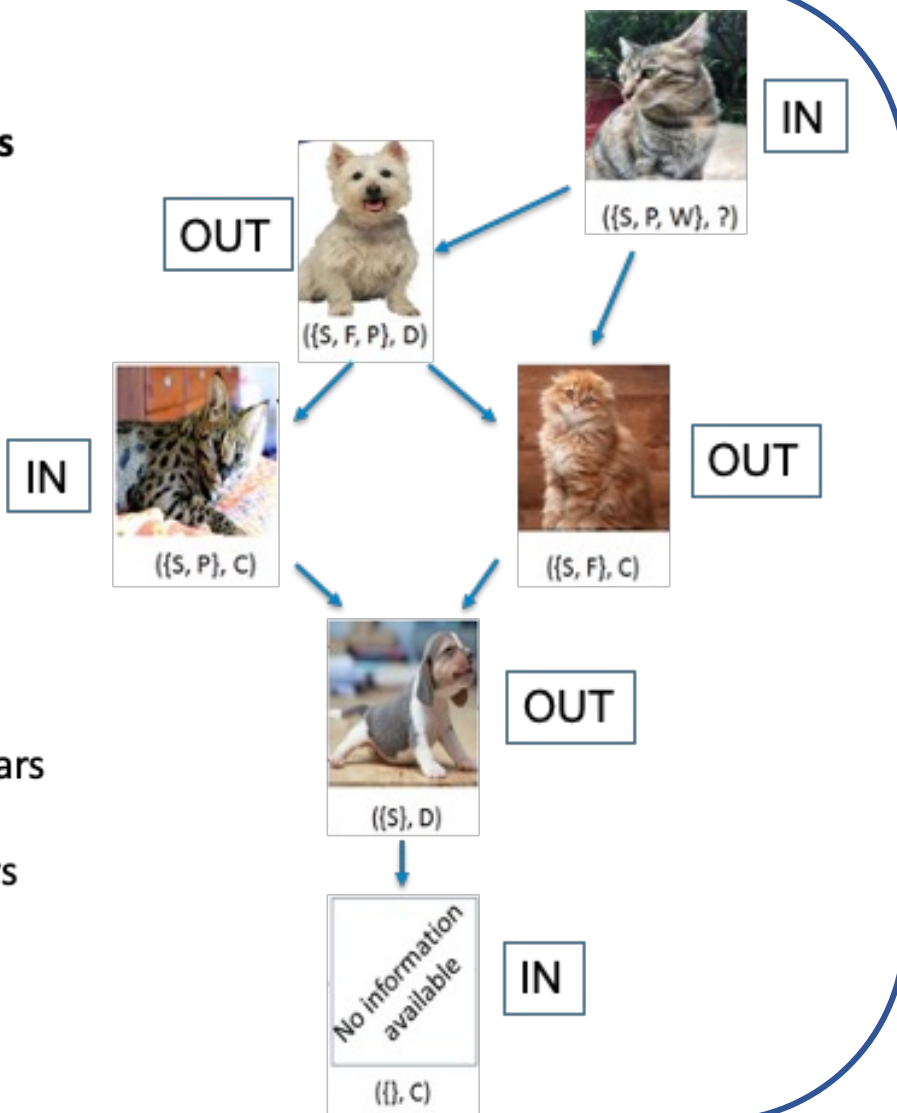
(Stylianou, Toni, ongoing)

Outcomes

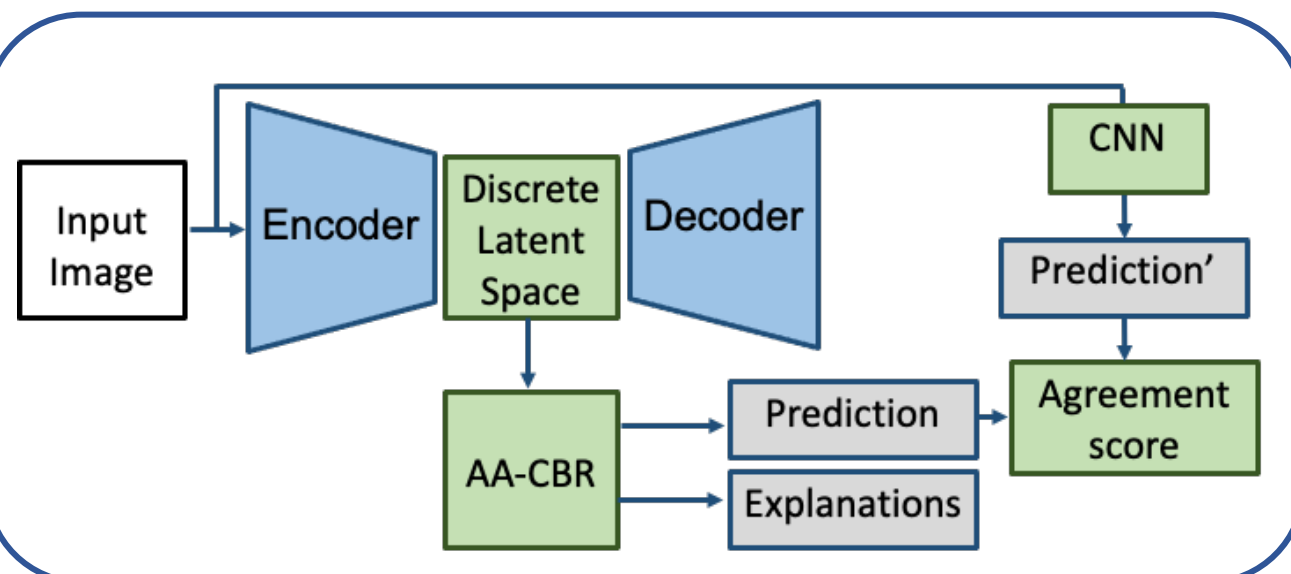
D: Dog
C: Cat

Features

S: small
P: pointy ears
F: fluffy
W: whiskers



Bird
 "has beak"
 "has wing"
 "feather"
 "has head"
 "has leg"



Argumentative Causal Discovery

(Russo, Rapberger, Toni 2024)

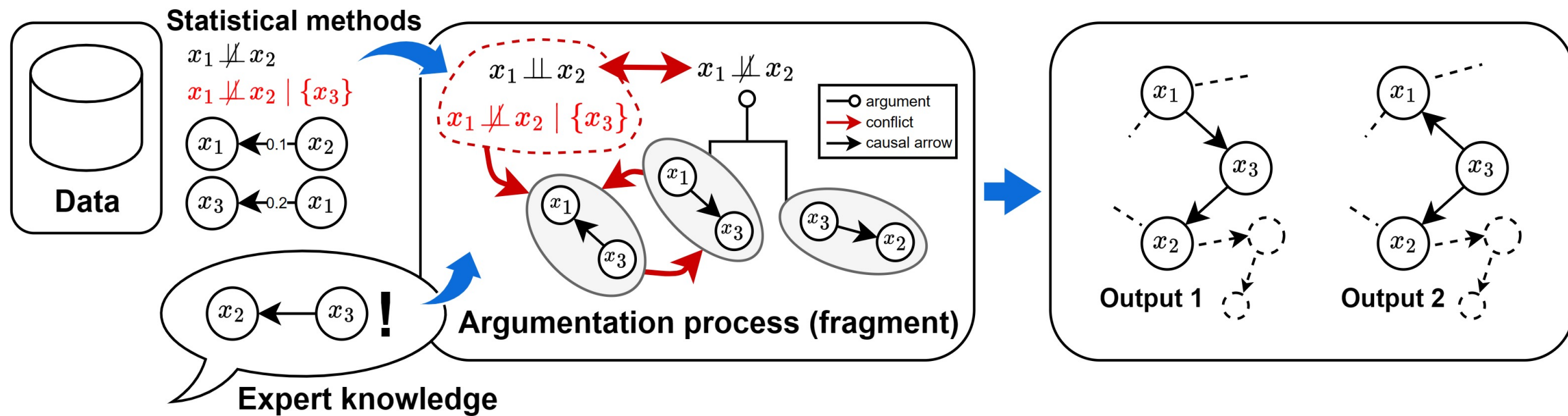


Figure 1: Overview of the workflow of our *Causal ABA algorithm*, which combines statistical methods and expert domain knowledge with non-monotonic reasoning and performs argumentative reasoning to output causal graphs consistent with the reported causal relationships.

Thanks



European Research Council
Established by the European Commission



J.P.Morgan

