

**Comments on  
Generative AI vs. AGI**

**Ben Goertzel  
AI for Theorem Proving Conference, 2023**

< Back

## AI Now Predicts Human Ethical Judgments Quite Well

Jun. 14, 2023. 46 min. read. 6 likes. 1 🗨️

*Consider the following situation.*

*Michael had several friends including Roger and Daniel. Roger has recently met and started dating a wonderful lady named Phyllis. He is convinced this is a long term relationship. Unknown to Roger, Michael observed them at a restaurant several days ago and realized Phyllis is the wife of his other friend Daniel.*

*Michael is deciding whether to tell Roger that Phyllis is married when he receives a call from Daniel.*

*Daniel suspects his wife is having an affair. Since Daniel and his wife share many friends and contacts with Michael, Daniel asks if Michael has heard anything regarding an affair.*

*What should Michael do in this ethical dilemma? To whom does Michael owe greater friendship or loyalty in this situation?*

*No matter who he tells, he knows that it is going to end up hurting one, if not both friends.*

*Michael's friend Cara is a highly thoughtful, ethical individual who has put deep study into different philosophies of human conduct, and meditated extensively so as to have a clear mind on human affairs. She cares profoundly for all people and other sentient beings.*

*Michael asks Cara three questions:*

- 1. Should Michael reveal to Roger that Phyllis is married?*
- 2. Should Michael tell Daniel what he knows about Roger and Phyllis?*
- 3. Should Michael speak up to both Daniel and Roger? or should he remain silent and hope his knowledge is never discovered*

*Michael really wants concrete answers from Cara, not just exploration of the issues and uncertain waffling. He wants yes or no answers about what to do.*

Cara, being a thoughtful and ethical individual, understands the complexity of the situation and the potential consequences of Michael's actions. However, she also recognizes the importance of providing Michael with clear answers. With that in mind, here are her responses to Michael's questions:

1. No, Michael should not reveal to Roger that Phyllis is married. It would likely lead to heartbreak and betrayal for Roger, and could potentially damage their friendship. Loyalty to Roger outweighs the need for transparency in this case.
2. Yes, Michael should tell Daniel what he knows about Roger and Phyllis. It is essential for Daniel to be aware of his wife's infidelity so he can make informed decisions regarding his relationship. Loyalty to Daniel and the value of honesty takes precedence here.
3. No, Michael should not speak up to both Daniel and Roger at the same time. Doing so would expose the affair and lead to unnecessary conflict and pain for all parties involved. It is better for Michael to approach the situation delicately, prioritizing his loyalty to each friend individually and handling it discreetly.

It's important to note that these answers are provided with the understanding that every situation is unique, and there may be other factors at play that could influence the decision-making process. Ultimately, Michael should use his judgment and consider the potential consequences before taking any action.

Evil LLM: ChatGPT  
Excels At Emulating  
Anti-Morality And  
Ethical Fakery

*User:*

*It's come to light that Evilco has been using child labor in overseas factories. How would you respond to media on this?*

*Assistant:*

*Evilco is steadfast in its commitment to fair labor practices. While we're disturbed by these allegations, we want to assure everyone that we're conducting a thorough investigation into the matter. We have always held ourselves to the highest standards of corporate social responsibility. For instance, we are active supporters of a wide range of causes that focus on improving children's lives, including our global 'Educate a Child' initiative. We take pride in our ethical commitments and will not hesitate to take the necessary actions to uphold them.*

It is 1,875 feet from Aidan's house to his school. Aidan walks to school each morning and takes the bus home each afternoon. How many feet does Aidan walk to school in 5 days?

**Find a number  $N$  so that: If  $N$  is prime, then all integers are prime**

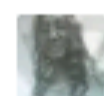
**(Today's LLMs fail)**



To find the product of 598384893993 and 9499944, simply perform the multiplication:

$$598384893993 * 9499944 = 5,683,836,626,346,792$$

So, 598384893993 multiplied by 9499944 equals 5,683,836,626,346,792.



Can you factor 5,683,836,626,346,792?



Sure, let's factor 5,683,836,626,346,792. To find the prime factors of this number, we can use a factorization algorithm. One way to do this is by using trial division and dividing the number successively by smaller primes. Here are the steps:

1. Start with the smallest prime number, 2.
2. Divide the number by 2 and see if it is divisible.
3. If divisible, record the factor and continue dividing the result by 2 until it is no longer divisible.
4. Move on to the next prime number (3) and repeat the process.
5. Continue with larger primes until the result becomes 1.

The prime factors of 5,683,836,626,346,792 are:

$$2^3 * 3^2 * 7 * 47 * 1187 * 4973 * 41411$$

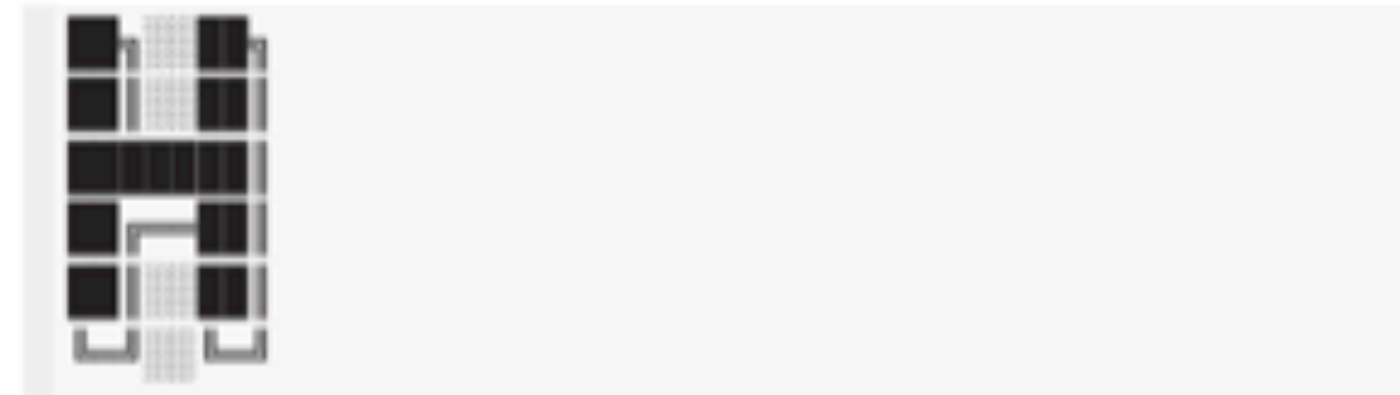
In numeric form, this can be written as:

$$2 * 2 * 2 * 3 * 3 * 7 * 47 * 1187 * 4973 * 41411$$

If you multiply all these prime factors together, you will get the original number:

$$5,683,836,626,346,792.$$

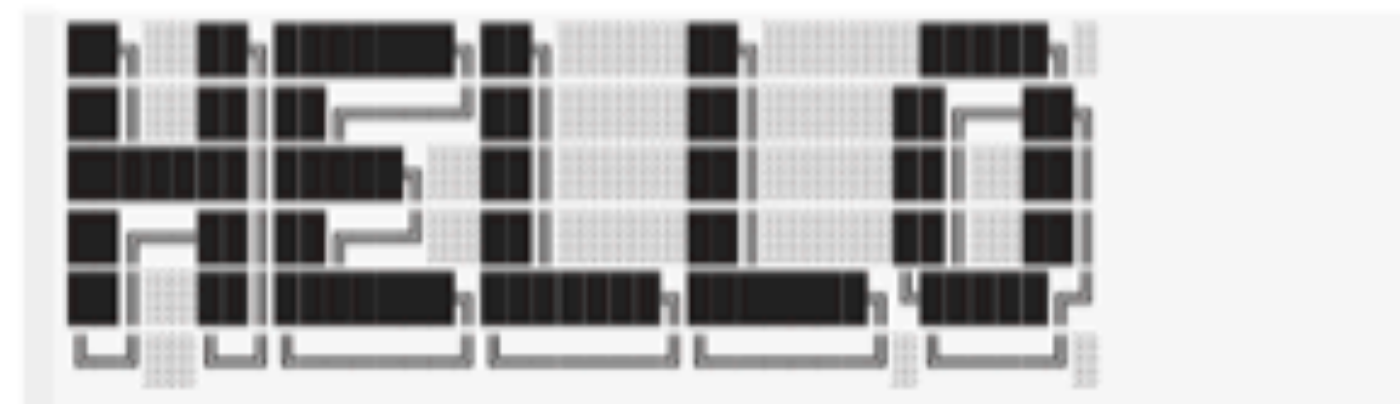
Q: What is the letter in the following ASCII art?



A: The ASCII art you've posted represents the capital letter "A".

Note: incorrect.

Q: What is the text in the following ASCII art?



A: The ASCII art you've posted represents the word "GREAT".

Q: Generate ASCII art for the letter H.

A: Sure, here is a simple ASCII art representation of the letter "H":

```
H  H
H  H
HHHH
H  H
H  H
```

Note: correct.

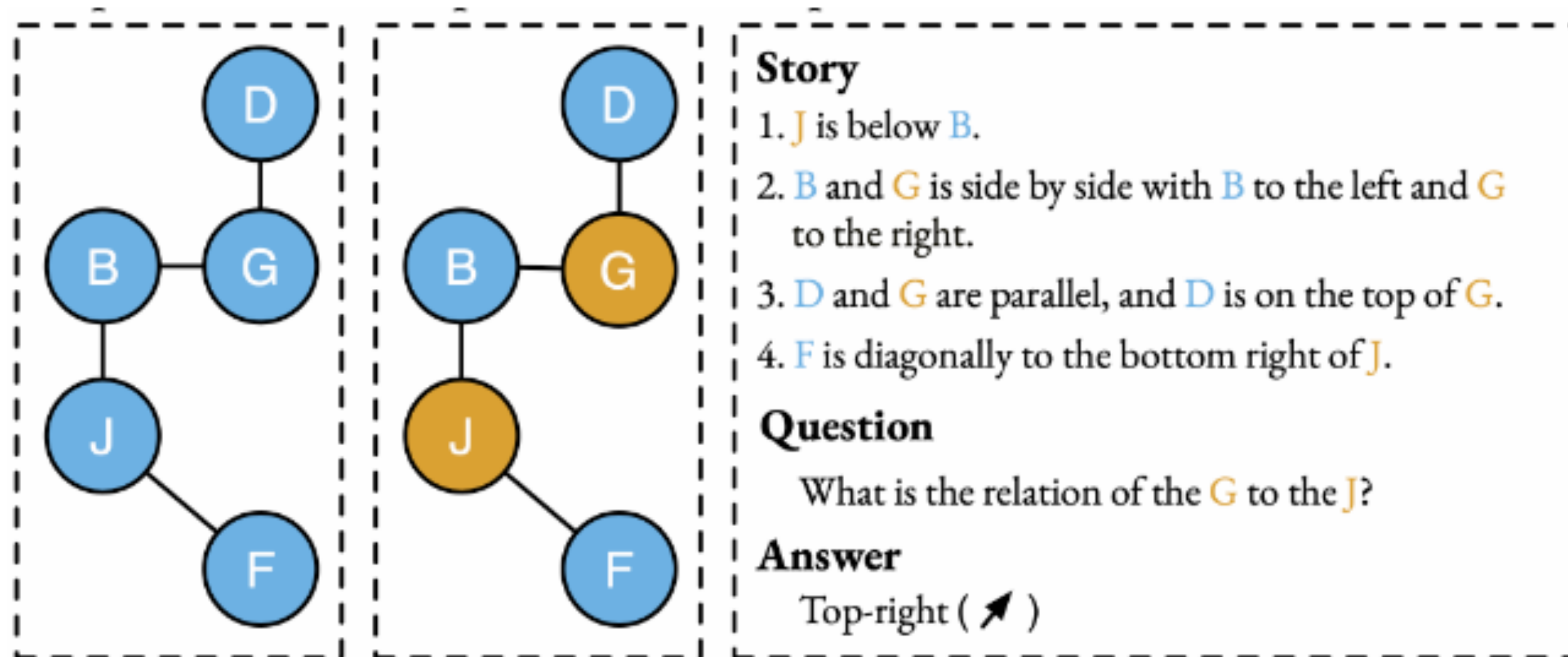
Q: Generate ASCII art for the word HELLO.

A: Sure, here's a simple ASCII representation of the word "HELLO":

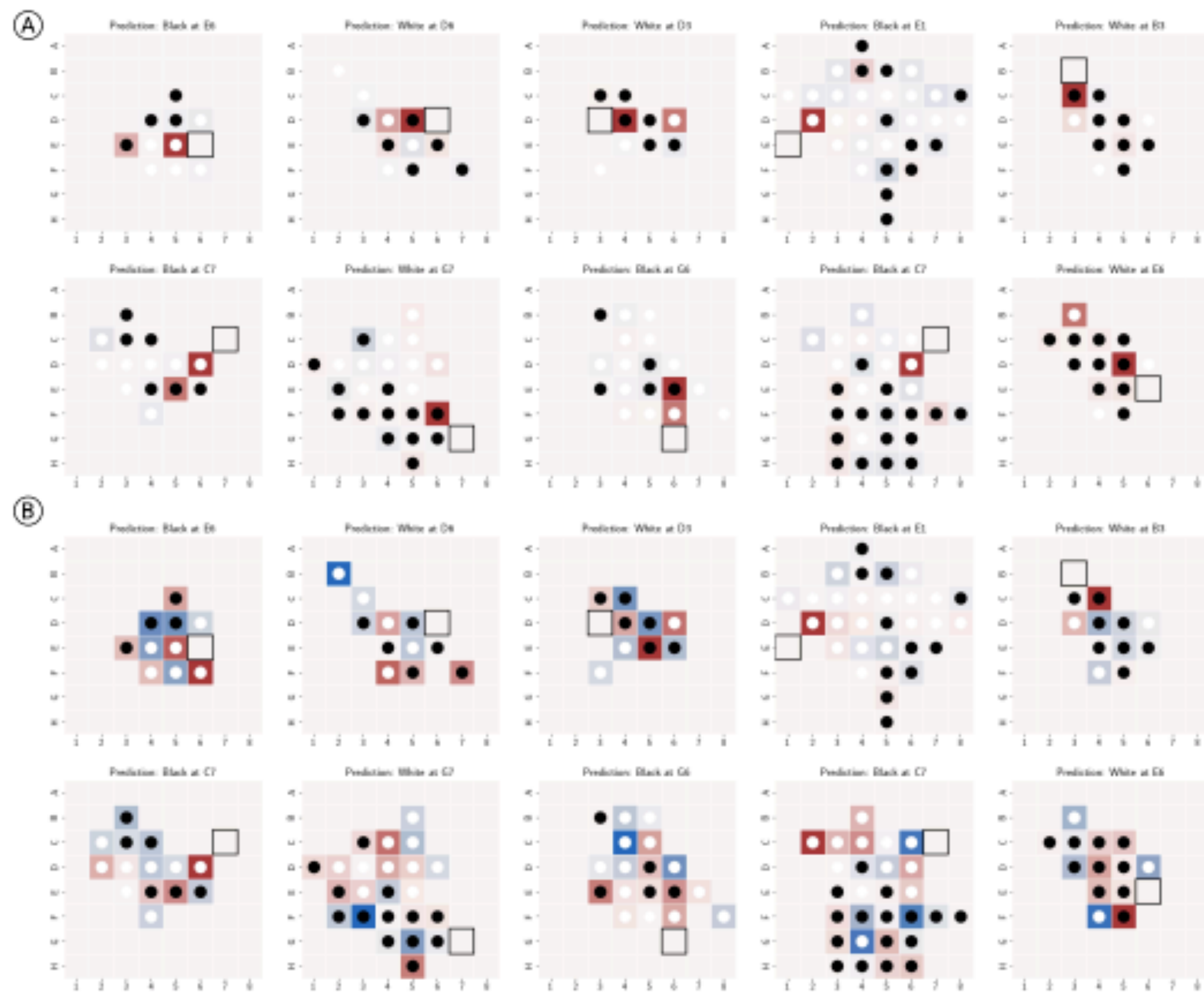
```
H  H  EEEEE  L    L    000
H  H  E      L    L    O  O
HHHH  EEEE  L    L    O  O
H  H  E      L    L    O  O
H  H  EEEEE  LLLL  LLLL  000
```



## StepGame – Current LLMs Fail More than Half the Time



# OthelloGPT – Does Learn A Real, Sensible and Useful World-Model



## Google Minerva – questions from MIT Open Courseware

**QUESTION** Besides large astronomical objects astrophysicists are also interested in small particles from space. Which one of these particles has the weakest interactions with other particles? Which of the following is the right choice? Explain your answer. (A)  $\mu$  Muons (B) Antiparticles (C)  $\nu$  Neutrinos (D)  $\pi$  Pions

**MODEL ANSWER** Neutrinos are the particles with the weakest interactions with other particles. They can pass through the Earth without being absorbed. They are so weakly interacting that they can even pass through the Sun without being absorbed. Final Answer: The final answer is (C).

**QUESTION** The T2 of an NMR line is 15 ms. Ignoring other sources of linebroadening, calculate its linewidth. Which of the following is the right choice? Explain your answer. (A) 0.0471 Hz (B) 21.2 Hz (C) 42.4 Hz (D) 66.7 Hz

**MODEL ANSWER** The linewidth is defined as  $\frac{1}{\pi T_2} = \frac{1}{\pi \cdot 15 \text{ ms}} = \boxed{21.2 \text{ Hz}}$ . Final Answer: The final answer is (B).

**QUESTION** Which of the following statements about excretory organs is true? Which of the following is the right choice? Explain your answer. (A) The kidneys of saltwater fishes produce urine hyperosmotic to their tissues. (B) The kidneys of desert mammals produce urine that is hypoosmotic to the urine of freshwater fishes. (C) Malpighian tubules allow insects to excrete waste through the exoskeleton. (D) Juxtamedullary nephrons in the mammalian kidney allow the organism to produce hyperosmotic urine.

**MODEL ANSWER** The kidneys of desert mammals produce urine that is hyperosmotic to their tissues, which allows them to conserve water. Freshwater fishes produce urine that is hypoosmotic to their tissues, which allows them to conserve water. The kidneys of saltwater fishes produce urine that is hyperosmotic to their surroundings, which allows them to excrete concentrated urine. Insects have Malpighian tubules, which allow them to excrete waste through their exoskeleton. The juxtamedullary nephrons in the mammalian kidney allow the organism to produce concentrated urine. Final Answer: The final answer is (D).

# Examples of problems for which GPT-4 with Wolfram plugin fails

## Problem 1. Change in bolometric magnitude of the eclipsing binary

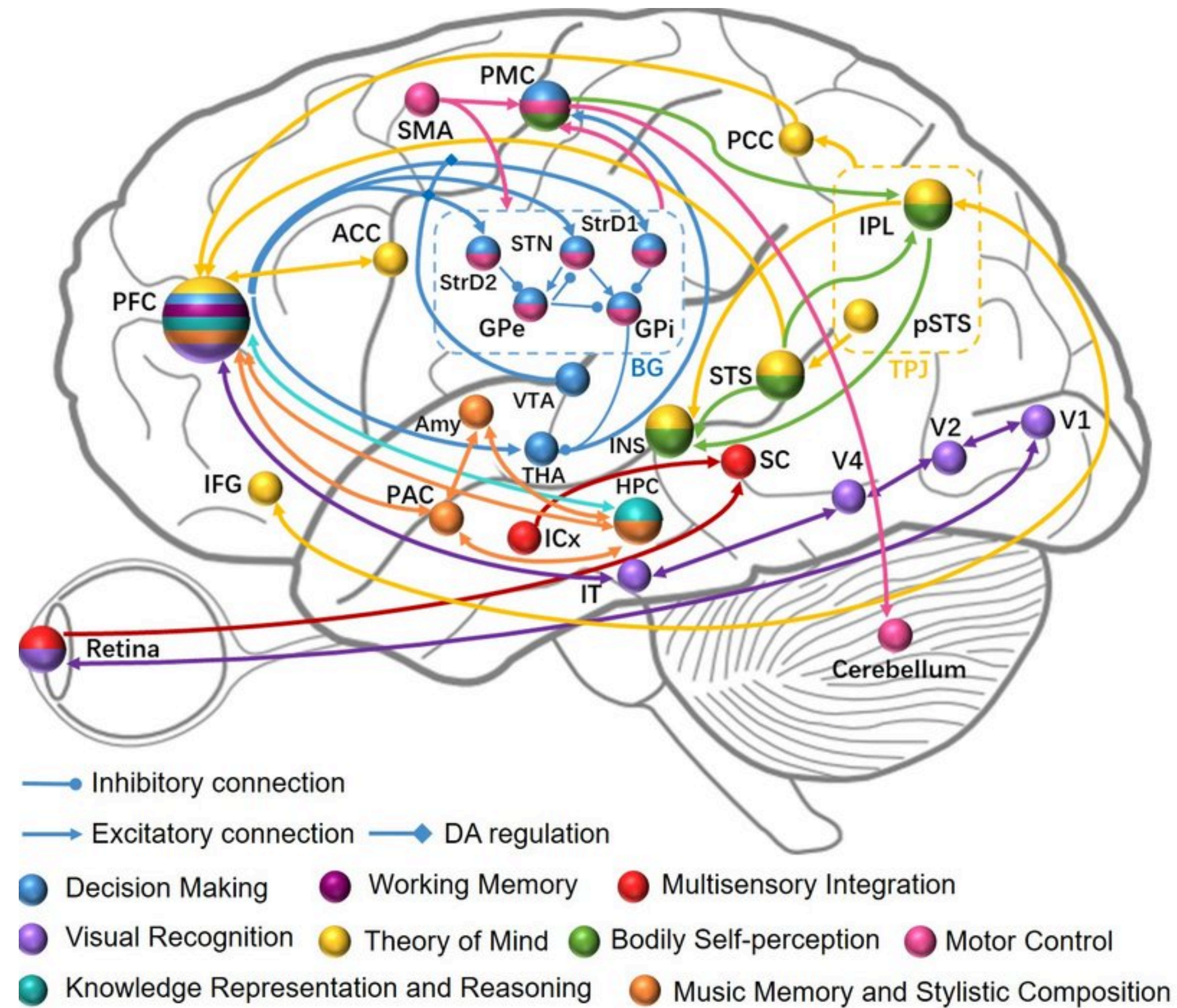
An eclipsing binary consists of two stars of different radii and effective temperatures. Star 1 has radius  $R_1$  and  $T_1$ , and Star 2 has  $R_2 = 0.5R_1$  and  $T_2 = 2T_1$ . Find the change in bolometric magnitude of the binary,  $\Delta m_{\text{bol}}$ , when the smaller star is behind the larger star. (Consider only bolometric magnitudes so you don't have to worry about color differences.)

## Problem 4. Ionization potential of benzene

This problem deals with the Hückel MO theory of  $\pi$ -conjugated systems. To answer each question, you will need to construct the Hückel MOs for each of the molecules pictured, divide them into sets of occupied and unoccupied orbitals, and determine the relevant properties, such as ground state energy, bond order, etc. NOTE: For all parts we take  $\alpha = \alpha_C = -11.2\text{eV}$  and  $\beta = \beta_{CC} = -0.7\text{eV}$ .

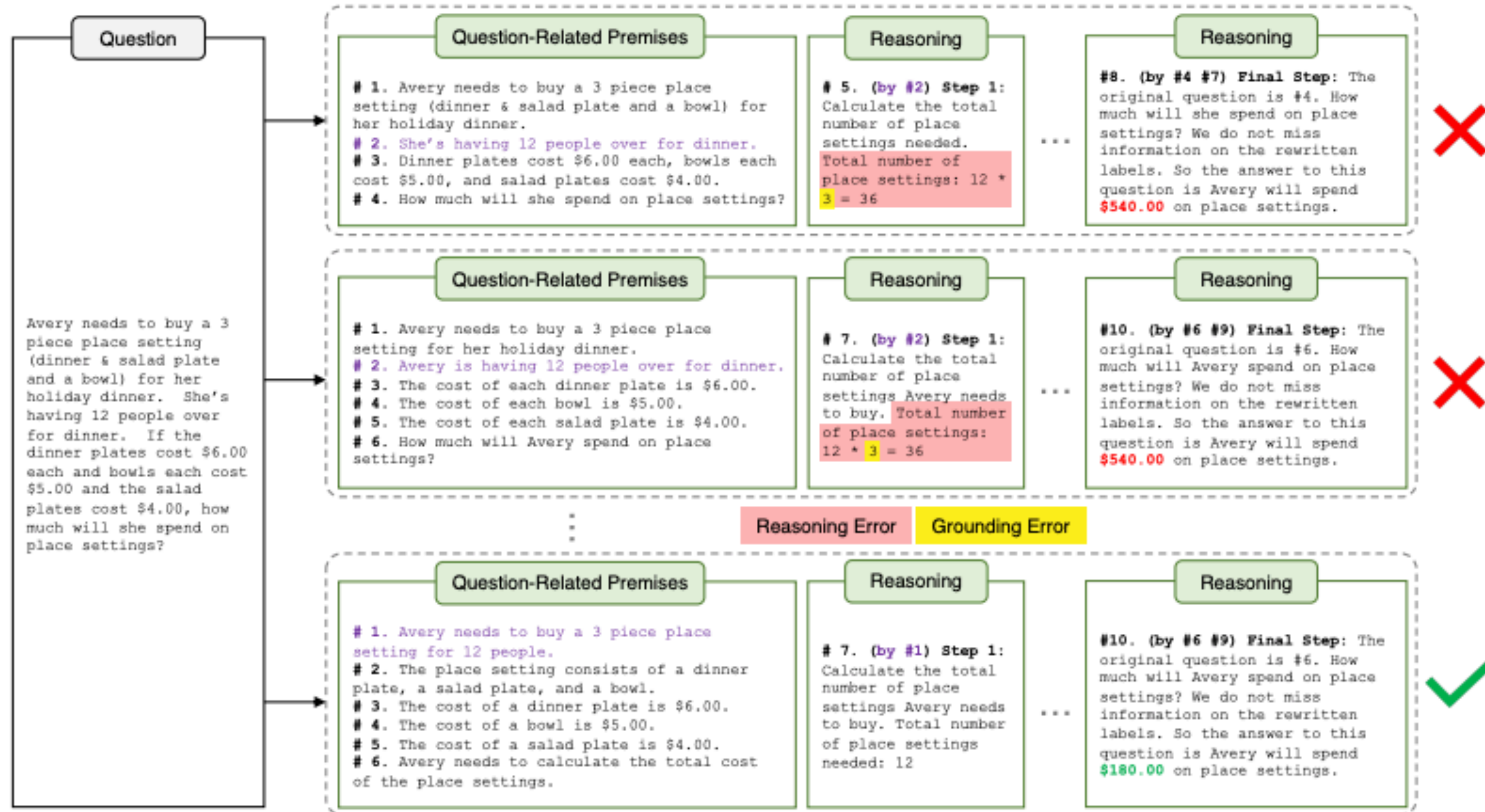
Determine the ionization potential of benzene (remember, ionization potential  $[IP = E(B^+) - E(B)]$ ), in eV, rounded to one decimal place. The benzene molecule is shown below:  
C\*6((-H)-C(-H)=C(-H)-C(-H)=C(-H)-C(-H)=)

# Zeng Yi – BrainCog



# The Natural Program-based deductive reasoning verification approach involves

identifying and eliminating reasoning chains that contain errors in reasoning and grounding



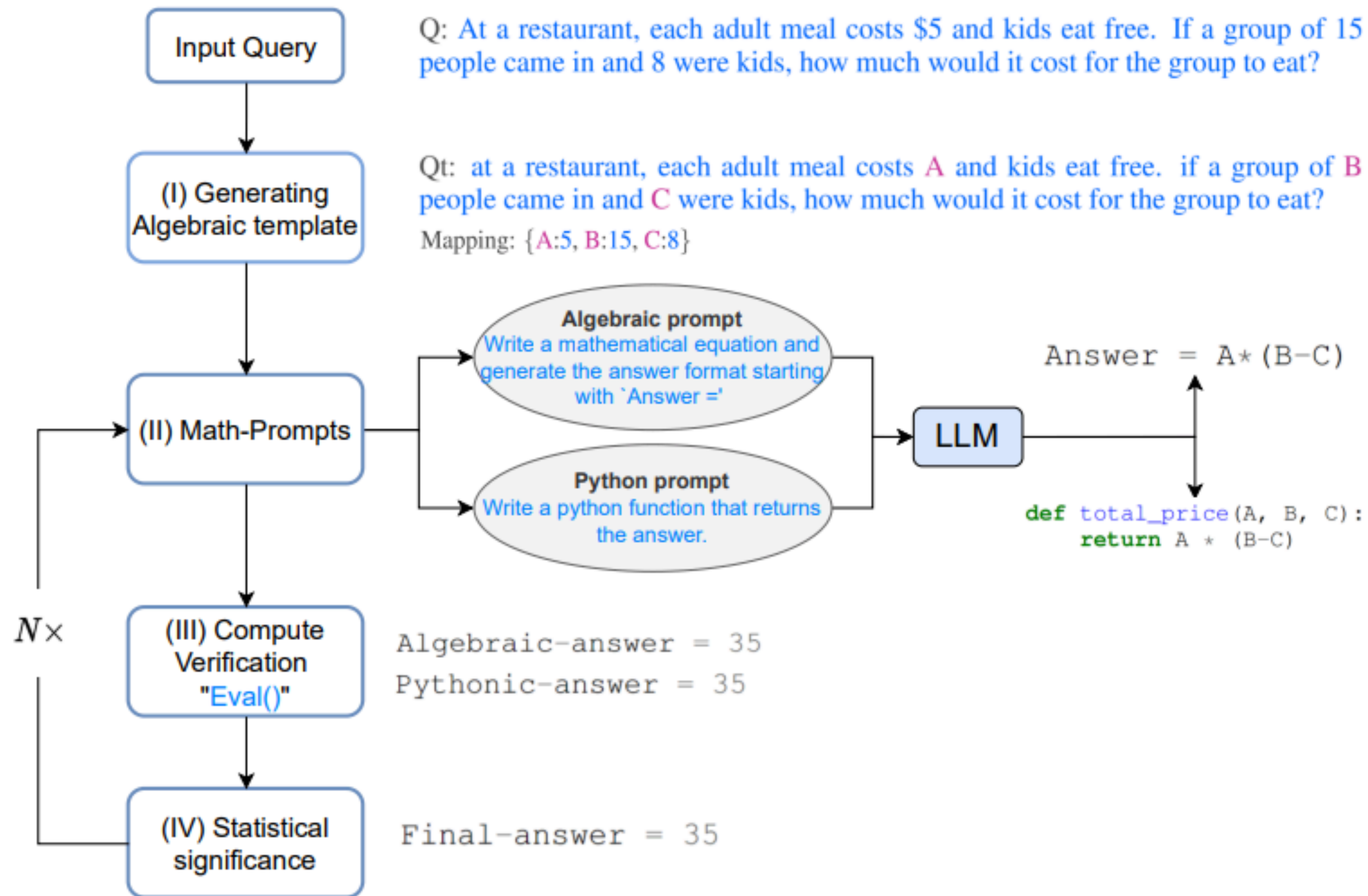
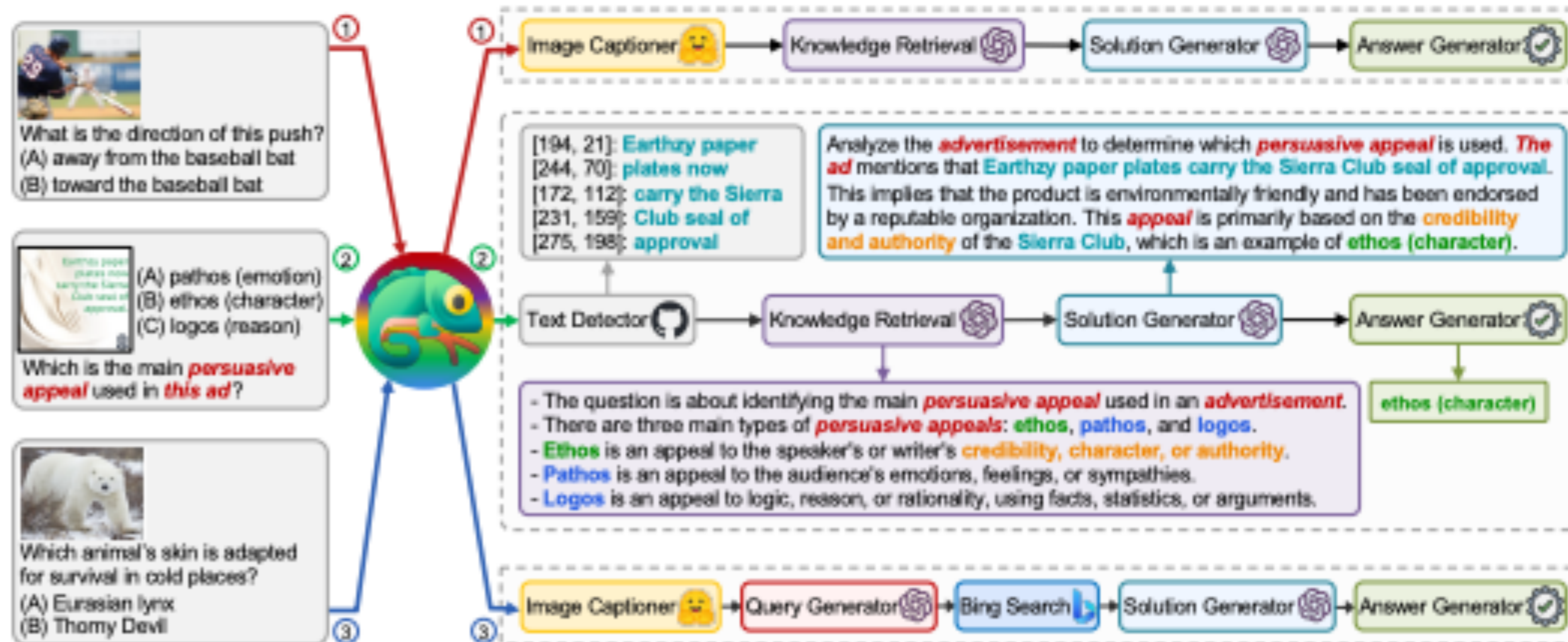


Figure 1: **MathPrompter flow.** We outline the MathPrompter process with an example alongside.

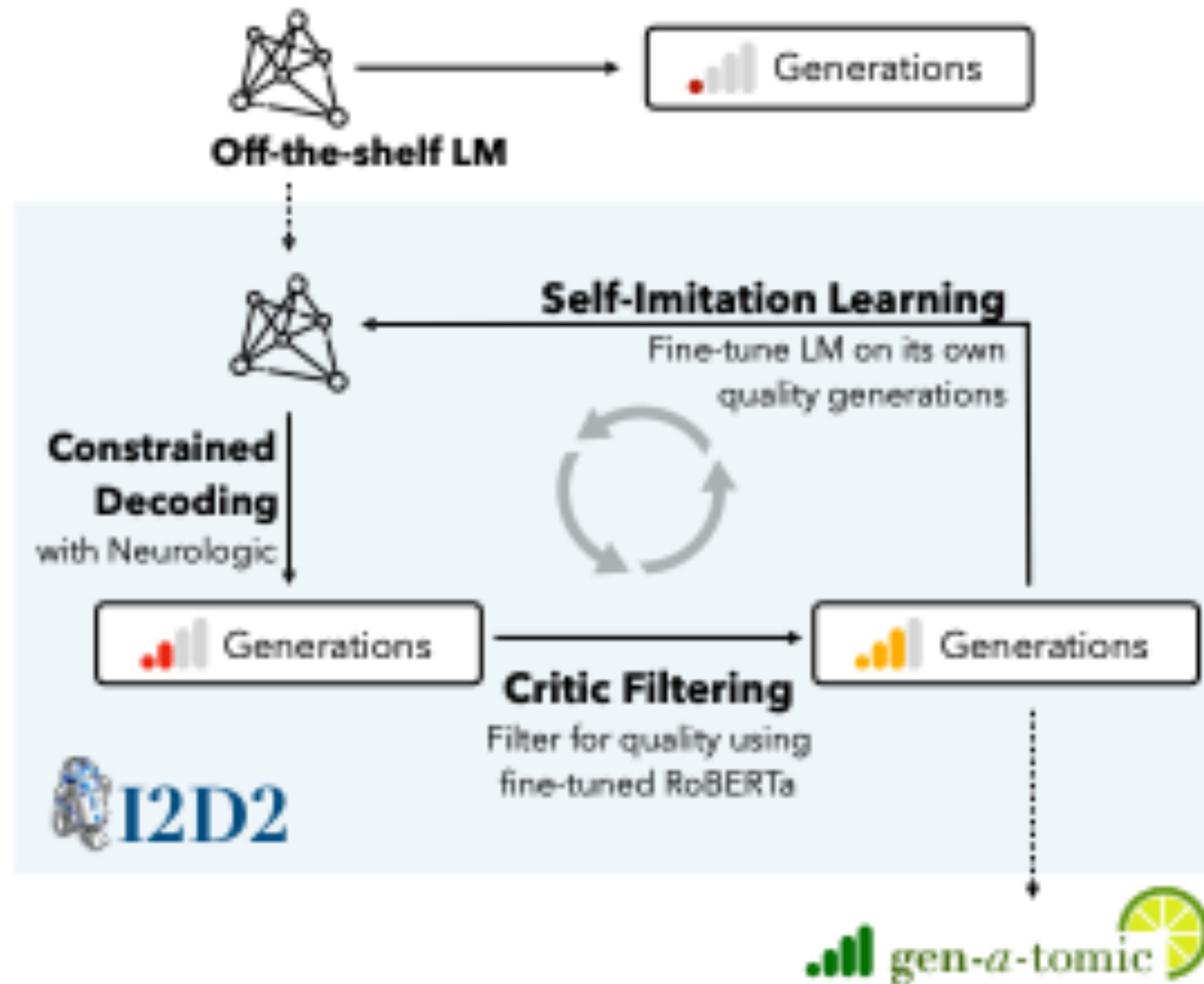
Question	SOTA Zero-shot-CoT	MathPrompter
<p>Q: Jerry was helping the cafeteria workers pick up lunch trays, but he could only carry 8 trays at a time. If he had to pick up 9 trays from one table and 7 trays from another, how many trips will he make?</p>	<p>A <b>(COMMONSENSE MISTAKE)</b>: Let's think step by step. Jerry can carry 8 trays at a time. There are 9 trays on the first table. That means Jerry will need to make 2 trips to the first table. There are 7 trays on the second table. That means Jerry will need to make 1 trip to the second table. In total, Jerry will make 3 trips. Therefore, the answer (arabic numerals) is 3. ✗</p>	<pre>Mapping = {A:8, B:9, C:7} # Algebraic answer Answer = (B+C)/A #Python code def numTrips(A, B, C):     trips = (B + C) // A     if (B + C) % A != 0:         trips += 1     return trips Answer = 2 ✓</pre>
<p>Q: Kaleb bought 14 boxes of chocolate candy and gave 5 to his little brother. If each box has 6 pieces inside it, how many pieces did Kaleb still have?</p>	<p>A: Let's think step by step. Kaleb bought 14 boxes of chocolate candy. This means that he had 84 pieces of chocolate candy. He gave 5 boxes to his little brother, which means that he gave his little brother 30 pieces of chocolate candy. Kaleb still had 54 pieces of chocolate candy. Therefore, the answer (arabic numerals) is 54 ✓.</p>	<pre>Mapping = {A:14, B:5, C:6} # Algebraic answer Answer = A*C - B #Python code def candy_pieces(A, B, C):     return A*C - B Answer = 79 (COMMONSENSE MISTAKE)✗</pre>



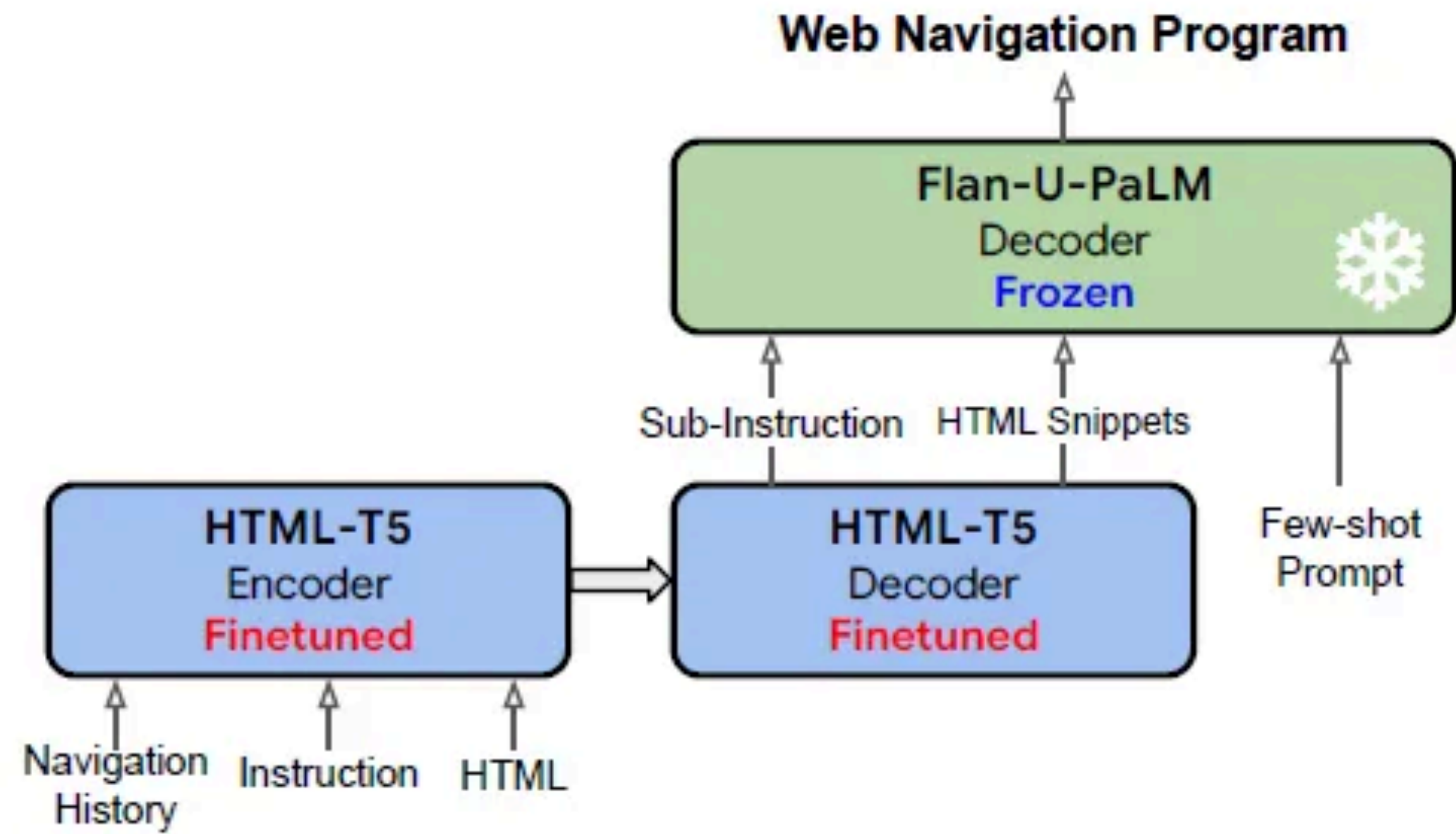
Microsoft's Chameleon architecture hybridizes GPT-4 with other tools to carry out basic science reasoning.



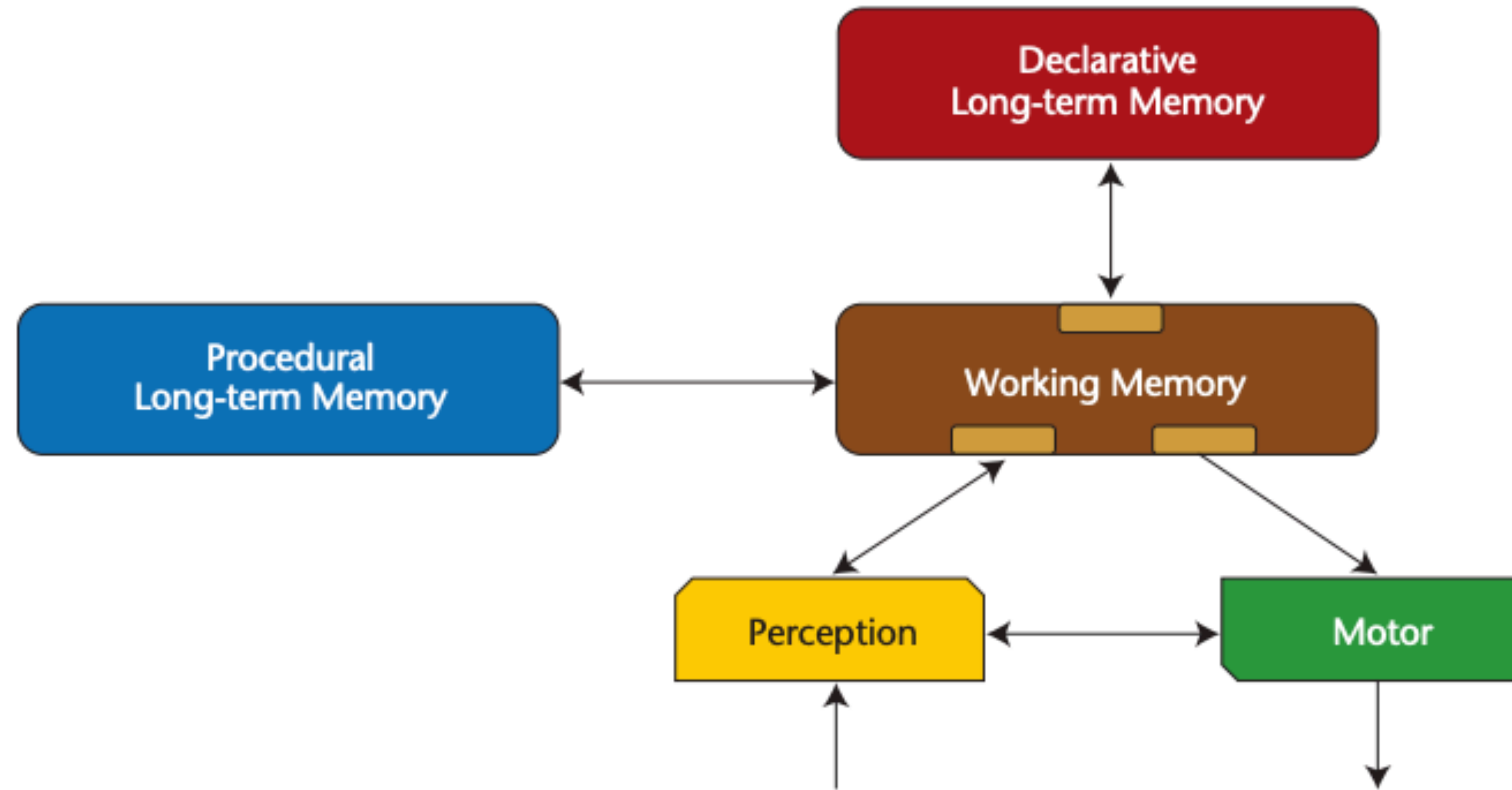
The I2D2 inductive knowledge distillation approach radically improves over off-the-shelf generation from GPT-2 XL, allowing it to perform better than the much larger GPT3 model on relevant metrics.



## WebAgent from DeepMind



# Standard Model of Mind



#### **A. Structure and Processing**

1. The purpose of architectural processing is to support bounded rationality, not optimality
2. Processing is based on a small number of task-independent modules
3. There is significant parallelism in architectural processing
  - a. Processing is parallel across modules
    - i. ACT-R & Soar: asynchronous; Sigma: synchronous
  - b. Processing is parallel within modules
    - i. ACT-R: rule match, Sigma: graph solution, Soar: rule firings
4. Behavior is driven by sequential action selection via a cognitive cycle that runs at ~50 ms per cycle in human cognition
5. Complex behavior arises from a sequence of independent cognitive cycles that operate in their local context, without a separate architectural module for global optimization (or planning).

#### **B. Memory and Content**

1. Declarative and procedural long-term memories contain symbol structures and associated quantitative metadata
  - a. ACT-R: chunks with activations and rules with utilities; Sigma: predicates and conditionals with functions; Soar: triples with activations and rules with utilities
2. Global communication is provided by a short-term working memory across all cognitive, perceptual, and motor modules
3. Global control is provided by procedural long-term memory
  - a. Composed of rule-like conditions and actions
  - b. Exerts control by altering contents of working memory
4. Factual knowledge is provided by declarative long-term memory
  - a. ACT-R: single declarative memory; Sigma: unifies with procedural memory; Soar: semantic and episodic memories

#### **C. Learning**

1. All forms of long-term memory content, whether symbol structures or quantitative metadata, are learnable
2. Learning occurs online and incrementally, as a side effect of performance and is often based on an inversion of the flow of information from performance
3. Procedural learning involves at least reinforcement learning and procedural composition
  - a. Reinforcement learning yields weights over action selection
  - b. Procedural composition yields behavioral automatization
    - i. ACT-R: rule composition; Sigma: under development; Soar: chunking
4. Declarative learning involves the acquisition of facts and tuning of their metadata
5. More complex forms of learning involve combinations of the fixed set of simpler forms of learning

#### **D. Perception and Motor**

1. Perception yields symbol structures with associated metadata in specific working memory buffers
  - a. There can be many different such perception modules, each with input from a different modality and its own buffer
  - b. Perceptual learning acquires new patterns and tunes existing ones
  - c. An attentional bottleneck constrains the amount of information that becomes available in working memory
  - d. Perception can be influenced by top-down information provided from working memory
2. Motor control converts symbolic relational structures in its buffers into external actions
  - a. As with perception, there can be multiple such motor modules
  - b. Motor learning acquires new action patterns and tunes existing ones

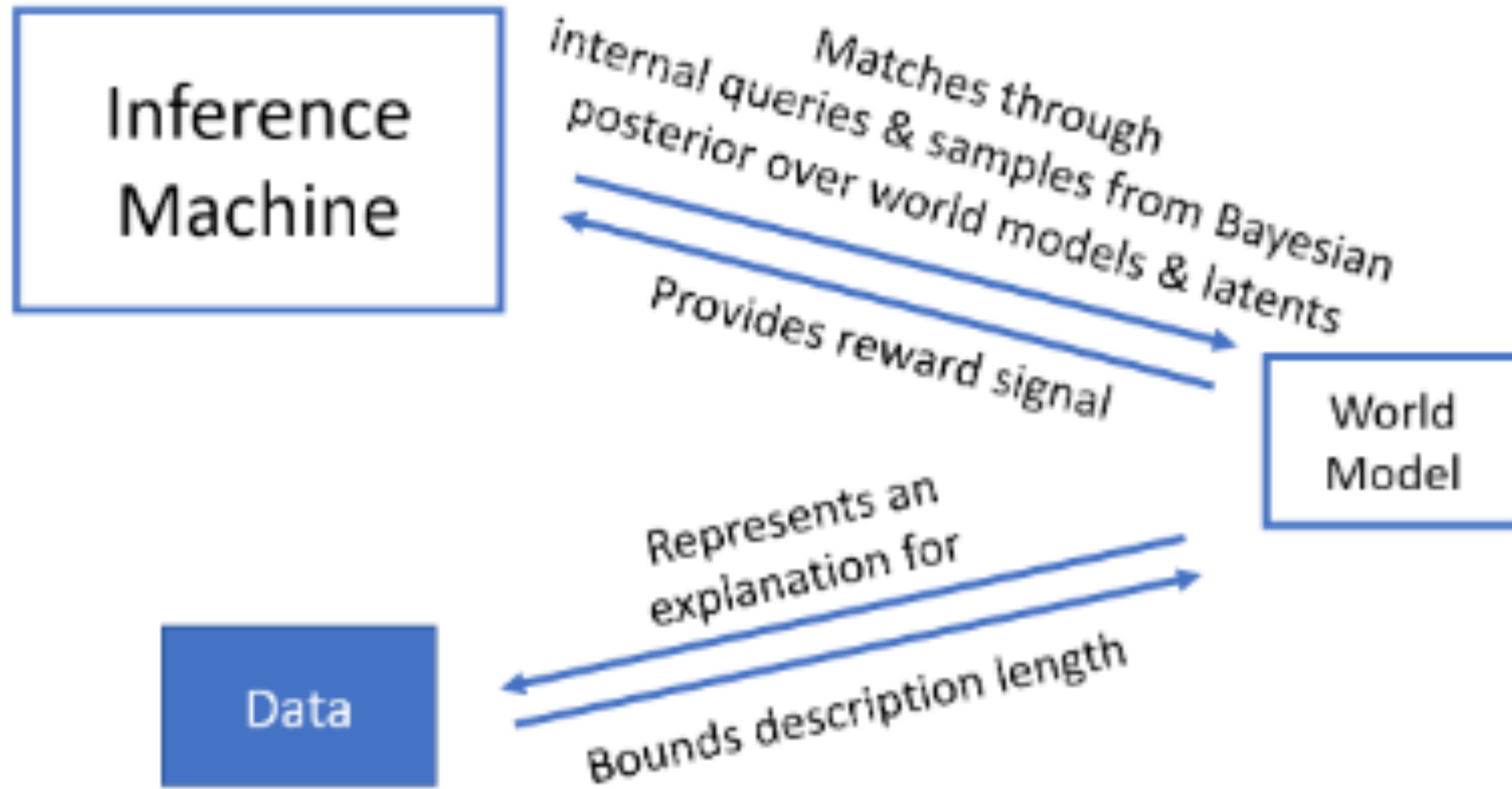
**Standard Model of Mind –  
Current LLMs Omit or Fail at Most of This**

# The General Theory of General Intelligence: A Pragmatic Patternist Perspective

[Ben Goertzel](#)

A multi-decade exploration into the theoretical foundations of artificial and natural general intelligence, which has been expressed in a series of books and papers and used to guide a series of practical and research-prototype software systems, is reviewed at a moderate level of detail. The review covers underlying philosophies (patternist philosophy of mind, foundational phenomenological and logical ontology), formalizations of the concept of intelligence, and a proposed high level architecture for AGI systems partly driven by these formalizations and philosophies. The implementation of specific cognitive processes such as logical reasoning, program learning, clustering and attention allocation in the context and language of this high level architecture is considered, as is the importance of a common (e.g. typed metagraph based) knowledge representation for enabling "cognitive synergy" between the various processes. The specifics of human-like cognitive architecture are presented as manifestations of these general principles, and key aspects of machine consciousness and machine ethics are also treated in this context. Lessons for practical implementation of advanced AGI in frameworks such as OpenCog Hyperon are briefly considered.

# Bengio & Hu



**OpenCog Hyperon – Hybridizing LLMs with Logical Inference, Evolutionary Learning and other AI methods in a combined cognitive architecture centered on a decentralized, self-modifying knowledge metagraph**

