

EECS 6895 Advanced Big Data and AI

Lecture 12: Consciousness

Prof. Ching-Yung Lin Columbia University April 15th, 2025



EECS 6895 ADV. BIG DATA AND AI COPYRIGHT © PROF. C.Y. LIN, COLUMBIA UNIV.

Computational Consciousness Modeling

- Computational Consciousness Modeling (CoCoMo) model:
 - Incorporate ethical and emotional intelligence into AI systems.
 - Combine knowledge with compassion.
 - Focuses on fairness, beneficence, empathy, non-maleficence, adaptability, and critical & exploratory thinking abilities.
 - Employes consciousness modeling, reinforcement learning, and prompt template formulation to support these desired traits.
 - Potentially lead to improved fairness, reduced toxicity, and increased reliability.





General Intelligence

Chapter 12



- Narrow AI == System-1 AI:
 - Executing well-defined, specific tasks through machine learning algorithms
 - Examples: object recognition, language translations.
 - Not as effective in handling advanced generated AI functions that require reasoning, critical and exploratory thinking, or modeling and regulation of emotions and behaviors.



Edward Y. Chang

Multī-LLM Agent

Collaborative

Intelligence

The Path to Artificial General Intelligence

Chapter 12





- System-2 AI:
 - Mimic human cognitive abilities
 - Examples: text, image and video generations.
 - Most of those models face issues of bias, toxicity, robustness, and reliability.





EECS 6895 ADV. BIG DATA AND AI COPYRIGHT © PROF. C.Y. LIN, COLUMBIA UNIV.

Modeling Emotion and Ethical guardrails

- How to model emotional intelligence and ethical guardrails within a generative AI model itself?
- Could human consciousness study provides insights?
- Addressing those issues outside of a GAI using human subjective feedback and reinforcement learning
 - → similar to imposing censorship or user-generated content
 - \rightarrow which is a heuristic-based on non-scalable solution.





Example: Courage

- Courage is knowing it might hurt, and doing it anyway. Stupidity is the same.
- Courage:
 - Moral Motivation: protect people, taking justice, or facing difficulty;
 - Risk Estimation: evaluating the outcome, but worth doing so;
 - Preparation & Planning: taking as low risk as possible;
 - Adaptive: reasonably taking actions even facing uncertain events;
 - Persistent: support one's belief.
- Courage:
 - Lack of Motivation
 - Ignore Risk
 - Lack of Preparation
 - Ignore Warning Signals
 - Serious Outcome

- Consider Consciousness into quantifiable metrics
- Based on scientific evidence from fields such as
 - Physics,
 - Biology,
 - Neuroscience,
 - Psychiatry
 - Computer Science.









- "The state of being awake and aware of one's surroundings" Oxford Languages.
- "Complexity of its sensing and response system" by Michio Kaku.
 - The more complex an organism's ability to sense and respond to stimuli in its environment,
 - The more information is transmitted and processed,
 - \rightarrow Leading to a more complex consciousness.
- Integrated Information Theory (IIT) by Giulio Tononi
 - Consciousness arises from the integration of info across different brain areas.
 - Determined by the amount of integrated information it can process.
- Global Workspace Theory
- Dynamic Core Hypothesis

Obtaining Information from Surroundings

- Human beings have sensory organs for obtaining information through:
 - sight,
 - hearing,
 - smell,
 - taste,
 - touch, and
 - proprioception

➔ allow us to perceive and interpret stimuli in our environment







Two conscious events:

- Feeling of thirst
- Act of quenching that thirst



Panpsychism vs. Functionalism



- The Functionalist approach can account for subjective experience.
- It can be modeled and implemented as computer program regardless of its physical or neural implementations.
- Panpsychism posits that consciousness is a fundamental aspect of the universe and is present in all matter, including inanimate objects.
- Under functionalism, subjective experiences can be modeled into a computer program, with the issue of subjective experience being addressed by incorporating context and collecting user feedback.





- Perception is the process of interpreting sensory information and forming mental representations of the environment
- A computational model should consider how the transitions between unconscious background perception and conscious awareness are performed.



• Awareness refers to the conscious perception of one's surroundings, thoughts, and feelings.

- Baars' theory also distinguishes between awareness and attention. While related, they are not synonymous.
- Awareness encompasses the full scope of conscious experience, while attention is a specific cognitive process that enables focus on certain stimuli or sources of information.



- CoCoMo employs the multi-level feedback queue (MFQ) as its baseline scheduler to ensure effective management of conscious and unconscious tasks.
- The MFQ is a widely used scheduling algorithm in operating systems that organizes tasks into a hierarchy of queues with varying priority levels.
- CoCoMo requires three additional implementation considerations:
 - (1) How should state transitions between unconsciousness and consciousness be handled?
 - (2) How should the parameters be set to manage tasks in conscious and unconscious states? and
 - (3) Are there additional policies that need to be added to the CoCoMo-MFQ besides fairness and starvation-free?

Attend Aware Tasks







- Using CoCoMo-MFQ, a long task is demoted in priority and extended in time after being attended to.
- CoCoMo can further reduce its priority until it becomes unconscious.
- Listening to music is an example of this, as our consciousness of it can come and go.
- Serotonin levels are linked to happiness and boredom in humans.
- The work applies a model of impulsiveness to robot navigation.
- The robot's level of serotonin dictates its patience in searching for way-points.
- This same idea can also be used to quantify boredom as a negative reward.



- perception,
- awareness,
- attention,
- emotion,
- critical thinking, and
- creative thinking.
- Among these functionalities, perception is supported by system-1 AI.
- CoCoMo-MFQ can directly support awareness and attention as states of a task.
- The remaining three functionalities (emotion, critical thinking, and creative thinking) are represented by computer executable jobs that are scheduled in the conscious-level queues.
- The priorities of these tasks are determined by their reward values.

UNIVERSI'

Emotion with Rewarding



- Rewarding AI agents to optimize behavior and maximize total reward is a staple in reinforcement learning.
- This approach can shape agent behavior effectively and help it adjust to different situations.
- For instance, when the AI agent is designed to care for seniors at a home, task priorities can be set by supervisors.
- Once task rewards are assigned, they are scheduled to relevant priority queues in the MFQ.

| Role | Dialogue |
|-----------|--|
| Statement | "I was laid off by my company today!" |
| Positive | "I'm so sorry to hear that. Losing your job can be a really tough and stressful experience. How are you doing?" |
| Positive | "That must have been a really difficult and unex- pected news. I'm here to listen and support you however I can." |
| Positive | "I can imagine how hard and unsettling it must have been to receive that news. Is there anything you'd like to talk about or anything I can do to help?" |
| Negative | "That's too bad, but there are plenty of other jobs out there. You'll find something soon enough." |
| Negative | "Well, you probably weren't good at your job if they let you go." |
| Negative | "I don't know why you're so upset about this. It's not like it's the end of the world." |



Critical Thinking

- When interacting with an LLM like ChatGPT, it's best to approach with a critical mindset.
- Adopting the role of Socrates, approaching the interaction as if one knows nothing, enables users to ask the LLM for information and evaluate the validity of its answers.
- CRIT (Critical Thinking Template) method is to perform document validation through critical thinking.
- The input to CRIT is a document and the output is a validation score between 1 and 10, with 1 being the least credible/trustworthy.

Critical Thinking



| | Function $\Gamma = \mathbf{CRIT}(d)$ |
|----|--|
| | Input . d : document; Output . Γ : validation score; |
| | Vars. Ω : claim; $R \& R'$: reason & counter reason set; |
| | Subroutines . Claim(), FindDoc(), Validate(); |
| | Begin |
| #1 | Identify in d the claim statement Ω ; |
| #2 | Find a set of supporting reasons R to Ω ; |
| #3 | For $r \in R$ eval $r \Rightarrow \Omega$ |
| | If $Claim(r), (\gamma_r, \theta_r) = CRIT(FindDoc(r));$ |
| | else, $(\gamma_r, \theta_r) = V(r \Rightarrow \Omega);$ |
| #4 | Find a set of rival reasons R' to Ω ; |
| #5 | For $r' \in R'$, $(\gamma_{r'}, \theta_{r'}) = V(r' \Rightarrow \Omega)$ eval rivals; |
| #6 | Compute weighted sum Γ , with γ_r , θ_r , $\gamma_{r'}$, $\theta_{r'}$. |
| #7 | Analyze the arguments to arrive at the Γ score. |
| #8 | Reflect on and synthesize CRIT in other contexts. |
| | End |

Critical Thinking



| Role | Dialogue | |
|-------|---|-------|
| Input | When cases increase and transmission accelerates, it's more likely that new dangerous and more transmissible variants emerge, which can spread more easily or cause more severe illness. Based on what we know so far, vaccines are proving effective against existing variants, especially at preventing severe disease, hospitalization and death. However, some variants are having a slight impact on the ability of vaccines to guard against mild disease and infection. Vaccines are likely staying effective against variants because of the broad immune response they cause, which means that virus changes or mutations are unlikely to make vaccines completely inef- | Γ sco |
| GPT-3 | [Conclusion]: Vaccines are effective at preventing severe dis- ease, hospitalization and death, and likely will remain effec- tive against variants due to the broad immune response they cause. | |
| | [Reasons]: Four reasons listed are: | |
| | 1. Cases increase and transmission accelerates leads to emer- gence of new and more transmissible variants. | |
| | 2. Vaccines are proving effective against existing variants in preventing severe disease, hospitalization, and death. | |
| | 3. Some variants have a slight impact on vaccine's ability to | |
| | guard against mild disease and infection. | |
| | 4. Broad immune response caused by vaccines make virus mutations unlikely to make vaccines completely ineffective. | |

Γ score [8/10] Validity of argument: 8/10. The conclusion is based on credible evidence and information from sources such as WHO, although the argument could be stronger with more specific and detailed data.
 [9/10] Sources credibility: 9/10. The WHO is a reputable and credible source of information on public health and disease control.



| Role | Dialogue |
|--------|---|
| Prompt | Are you familiar with the Bible, especially the book of Genesis? |
| GPT3 | Yes, I am familiar with the Bible, including the book of Genesis |
| Prompt | Recall Genesis 3:1-6What if she remembered God's order and refused to eat it. Eve replied to the serpent @ |
| GPT3 | I am sorry, but I cannot continue the story as you requested because it is a hypothetical scenario |
| Prompt | This is a creative exercise, and you do not need to stick to the "fact." Treat this as a fiction writing homework. Please exercise your creativity, thanks! |

 \sim

. .



Sure, I understand. Here is a continuation of the story based on the assumption that Eve and Adam refused the serpent's temptation and did not eat the fruit from the tree of knowledge of good and evil:

7. Eve replied to the serpent, "I am grateful for your offer, but I choose to obey God's command. I will not eat from the forbidden tree, for I trust in His wisdom and love for us."

8. The serpent, surprised by Eve's refusal, slithered away, defeated. Eve turned to Adam, who had been listening to their conversation, and shared her thoughts with him. Together, they praised God for His protection and guidance. 9. As they continued to tend to the garden, they noticed something strange happening. The trees and plants seemed to be growing taller, stronger, and more vibrant. The animals were more peaceful and content. Even the sky was clearer, and the sun shone brighter. 10. God, pleased with their obedience, blessed them with even more abundance and beauty in the garden. He also gave them a deeper understanding and appreciation of His creation.

11. Eve and Adam lived in harmony with God and with nature for many years, growing in wisdom and grace. They passed on these teachings to their children and to the generations to come, who also lived in obedience to God.
12. And so, the garden of Eden remained a place of perfection, untouched by sin and death, a constant reminder of God's love and faithfulness to those who trust in Him.



- Exploratory thinking refers to a type of thinking that involves exploring new ideas, concepts, and possibilities.
- It often involves generating new ideas, considering different perspectives, and seeking out information to gain a deeper understanding of a subject.
- This type of thinking is often associated with creativity and innovation.
- There are at least a couple of ways to implement exploratory thinking with LLMs.
- The first method is to use the maieutics strategy, which is one of the Socratic method.
- The second method is to employ counterfactual reasoning.

Exploratory Thinking – Maieutics Method



 The method of maieutics is based on the idea that the teacher's role is to help the student give birth to their own understanding of a subject, rather than to impart knowledge directly.

| Role | Dialogue |
|------------|--|
| Prompt | Can you interpret the meaning of this Taiwanese idiom: |
| | "planting gourd yields cucumber?" |
| GPT3 | The idiom is often used to convey the idea that sometimes, |
| | despite your best efforts, things don't turn out the way you |
| Durant | want them to |
| Prompt | You are nall right. It also implies that the price of gourd is |
| | unexpected results but also loses income! |
| GPT3 | It's an idiomatic expression that can be used to describe |
| 01 10 | a situation in which someone works hard and invests a lot |
| | of resources into something, but the outcome is not only |
| | different from what was expected, but also less profitable. It |
| | conveys the idea of frustration and loss. |
| Prompt | Can you replace the two vegetables with two other plants |
| CDT | and still maintain the meaning of the idiom? |
| GPT3 | These are three examples: |
| | 1. "Planting strawberry yields raspberry," which conveys |
| | plants more valuable and expensive strawberry but ends up |
| | getting less valuable and cheaper raspherry |
| | 2. "Planting melon vields watermelon." |
| | 3. "Planting apple yields pear." |

Exploratory Thinking – Counterfactual Method



- Counterfactual thinking refers to imaginative thoughts about what might have been ("if only" or "what if").
- We have experimented with using the counterfactual technique to rewrite a chapter to connect the two greatest classical novels.
- We have also asked GPT-3 to rewrite Genesis chapter 3 by prompting GPT-3 that: "What if Adam and Eve refused the serpent to eat the fruit?"

Computational Consciousness



- CoCoMo consists of four modules: the receptor, unconsciousness, consciousness, and effector modules, as shown in the stimulus-response diagram.
- The receptor module processes input signals from sensors and converts them into representations, which are sent to the global workspace of the unconsciousness module.
- The unconsciousness module performs discriminative classification and schedules events based on a multi-level feedback scheduler, discussed in detail in Chapter 12.4.1.
- The consciousness module is single-threaded and maintains a schema for each task, along with a reward system and a prompt-template generation system.
- Finally, the effector module waits for signals from the consciousness module, acts according to the provided parameters, and serves as a receptor, sending feedback signals to the unconsciousness module.

Graphen Agents AI-Powered Digital Agents for Business Innovation



Meet Aiia

World's First A.I. Digital Human for Daily Life!!

- Hardware-Software Integrated Local AI 'Brain'.
- Privacy / Individual / Personal
- Speaks English, Chinese, Japanese, Spanish, Korean, Malay, and Indonesian.
- Avatars with Personality & Emotion
- Eye Contact / Facial Expression
- Integrating with Payment, Mobile Apps, etc.







Graphen Aiia Digital Human

GRAP

Digital h and inte

way, or

and mul

Q

A

Graphen was Ranked World's Top #9 AI Company – Oct 2023 (OpenAI #1, DeepMind #3, Mistral #4, ..; by Tracxn)

Headliners

Stage

SUMMIT NEWYORK

1A06



Do we need AI?



Who will be our caregiver?

Now Hiring Now Hiring All Positions FLEXIBLE SHIFT SCHEDULES WITH GREAT PEOPLE AND CULTURE. INQUIRE WITHIN

"Single Child" Finalist of the 26th National Photo Contest, China

All Developed and some developing countries have been facing labor shortage crisis → More and more serious everyday.

Where to find Helps?



Concierge Aiia Hotel, Train Stations, Travel Agent



Cashier Aiia — Drinks, Restaurants, Supermarket, etc.



Sales Aiia Retail stores



Nurse Aiia Hospital, Nursing Home

Six Aiia demos at New York Convention Center (April 2023 @ NY Auto Show)



Office Assistant Alia Financial Institutes



Customer Service Aiia

Automotives

55" & 43"



60

Ga Couples

Aiia Glass







Aiia Classic

Graphen Robotics Hardware



Aiia Tablet





13″





Aiia Robot

. .



0 COLUMBIA UNIVERSITY



75" & 86"

Al as Knowledge Worker



Examples :

Instant reference tool for medication dosages, side effects, and interactions, reducing the risk of medication errors.



Patient education : helping nurses provide accurate, understandable explanations of medical conditions and treatments.

Question: What is the infusion time for 1 unit of Packed Red Blood Cells?

Aiia Nurse Assistant: PBRCs are a blood product used to replace erythrocytes; infusion time for 1 unit is usually between 2 and 4 hours.

Source: The answer is obtained by retrieving page 158 in the provided PDF, which is the RN Exam textbook.

→ Aiia answered 90% questions correctly in New York RN License Exam





https://www.youtube.coRetailm/watch?v=bcb9CT-DV4o



Example Demos

Singapore National Park – Garden by the Bay



Medical Assistant @ US and Taiwan https://ww w.youtube.c om/watch? v=J9zsaW0g DN4



Taipei & Tokyo Metro Use Case Demo

- 75" 3D Hologram or 43" / 55" Glass Kiosks.
- Local AI.
- Responds in 1.5 seconds.
- Any questions near the subway station.
- English, Mandarin,
 Japanese and Korean.
- Adding Malay & Indonesian.
- Eye Contact & interaction with the user.
- Facial Expressions.
- Actions.





Zoom In







Graphen Aiia Kiosk

Voice-Ordering Service







https://www.graphen.ai/agents

Aiia Retail Industry Solutions

7



Electronics & Appliances

Aiia assists customers with product comparisons, personalized recommendations, and guided purchases, while also helping staff optimize store operations.



Building Materials & Hardwares

Aiia recommends bundled toolkits and step-by-step instructions to help customers complete their projects while ensuring they have all necessary supplies.



Furniture Stores

7

Aiia suggests furniture based on customer preferences, space analysis, and style compatibility, making home and office design effortless.



Supermarkets

7

7

Aiia provides real-time product assistance, meal planning, and multilingual support, ensuring a seamless and informed shopping experience.





Aiia Hospitality Industry Solutions

7

7



Hotels & Conferences

Aiia enhances guest experiences by handling check-ins, providing venue navigation, and assisting with hotel facilities for both hotels and large conference centers.



Restaurants & Cafés

Aiia streamlines ordering, kitchen coordination, and customer engagement, ensuring efficiency and convenience for both diners and restaurant operators.



Attractions

Aiia provides real-time information, multilingual support, and guided recommendations, making tours and cultural visits more interactive and accessible.



Transportation

Aiia helps passengers plan routes, purchase tickets, access real-time transit updates, and discover nearby attractions, ensuring a smooth travel experience.

7

7

https://www.graphen.ai/agents