

Lab Seminar

2024.01.31

김규식

UNLEASHING COGNITIVE SYNERGY IN LARGE LANGUAGE MODELS: A TASK-SOLVING AGENT THROUGH MULTI-PERSONA SELF-COLLABORATION

Zhenhailong Wang^{1*}, Shaoguang Mao², Wenshan Wu², Tao Ge², Furu Wei², Heng Ji¹

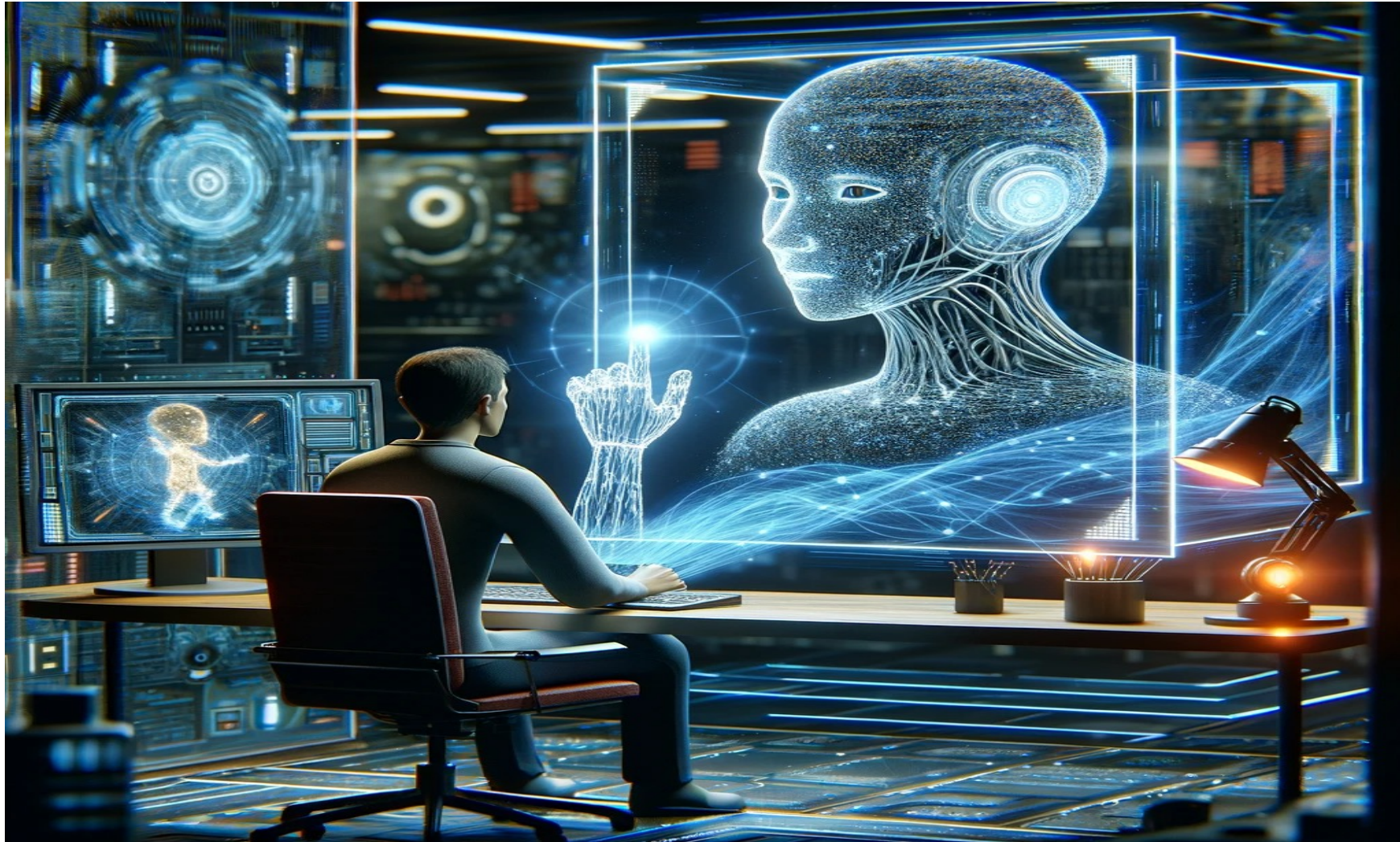
¹University of Illinois Urbana-Champaign, ²Microsoft Research Asia

{wangz3, hengji}@illinois.edu

{shaoguang.mao, wenshan.wu, tage, fuwei}@microsoft.com

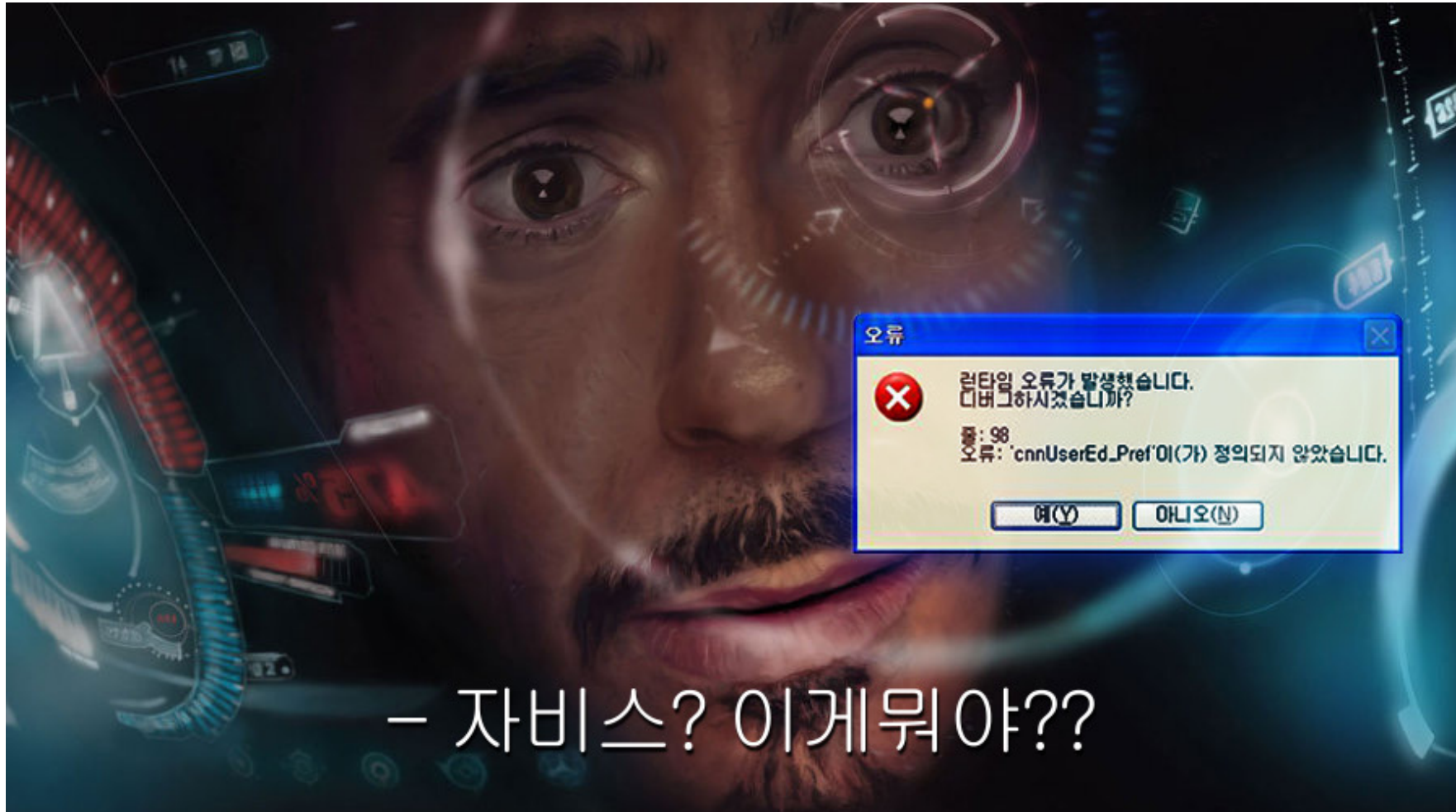
Solo Performance Prompting

Background



Solo Performance Prompting

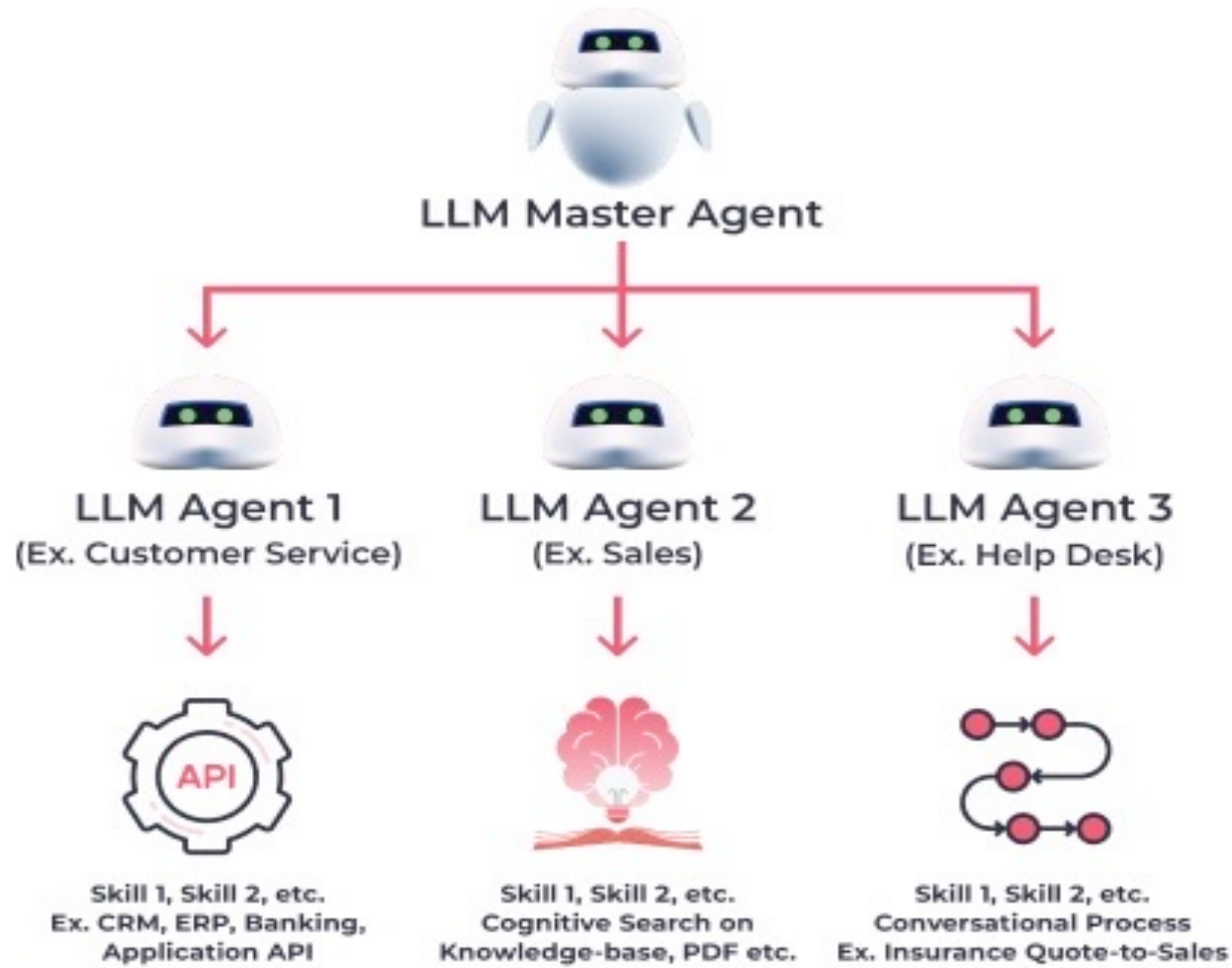
Background



Solo Performance Prompting

Background

출처: <https://floatbot.ai/llm-agent-m>



Solo Performance Prompting

Background

23. Exploring the Intersection of Large Language Models and Agent-Based Modeling via Prompt Engineering

Edward Junprung. [\[abs\]](#), 2023.8

24. Neural Amortized Inference for Nested Multi-agent Reasoning

Kunal Jha, Tuan Anh Le, Chuanyang Jin, Yen-Ling Kuo, Joshua B. Tenenbaum, Tianmin Shu. [\[abs\]](#), 2023.8

25. GPT-in-the-Loop: Adaptive Decision-Making for Multiagent Systems

Nathalia Nascimento, Paulo Alencar, Donald Cowan. [\[abs\]](#), 2023.8

26. ProAgent: Building Proactive Cooperative AI with Large Language Models

Ceyao Zhang, Kaijie Yang, Siyi Hu, Zihao Wang, Guanghe Li, Yihang Sun, Cheng Zhang, Zhaowei Zhang, Anji Liu, Song-Chun Zhu, Xiaojun Chang, Junge Zhang, Feng Yin, Yitao Liang, Yaodong Yang. [\[abs\]](#), 2023.8

27. Exploring Collaboration Mechanisms for LLM Agents: A Social Psychology View

Jintian Zhang, Xin Xu, Shumin Deng. [\[abs\]](#), 2023.10

28. Lumos: Learning Agents with Unified Data, Modular Design, and Open-Source LLMs

Da Yin, Faeze Brahman, Abhilasha Ravichander, Khyathi Chandu, Kai-Wei Chang, Yejin Choi, Bill Yuchen Lin. [\[abs\]](#), 2023.11

29. AutoAct: Automatic Agent Learning from Scratch via Self-Planning

Shuofei Qiao, Ningyu Zhang, Runnan Fang, Yujie Luo, Wangchunshu Zhou, Yuchen Eleanor Jiang, Chengfei Lv, Huajun Chen. [\[abs\]](#), 2024.01

Adversarial Interactions 🗯️🗣️

1. Encouraging Divergent Thinking in Large Language Models through Multi-Agent Debate

Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Zhaopeng Tu, Shuming Shi. [\[abs\]](#), 2023.5

2. Improving Factuality and Reasoning in Language Models through Multiagent Debate

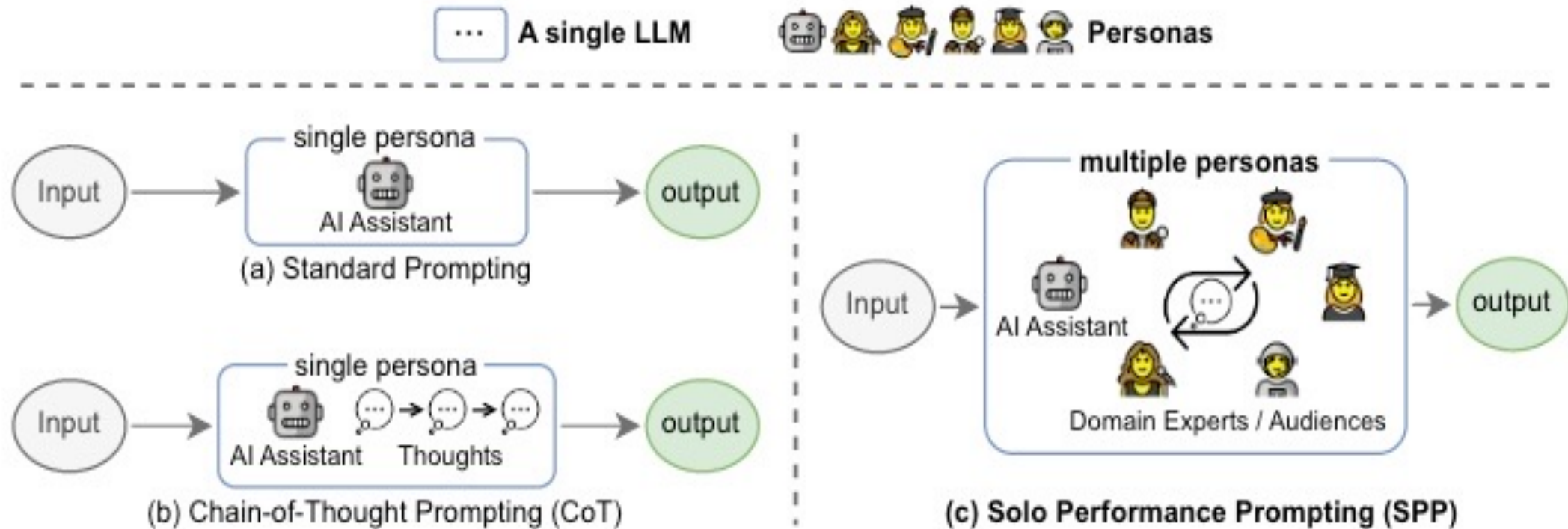
Yilun Du, Shuang Li, Antonio Torralba, Joshua B. Tenenbaum, Igor Mordatch. [\[abs\]](#), 2023.5

UNLEASHING COGNITIVE SYNERGY IN LARGE LANGUAGE MODELS: A TASK-SOLVING AGENT THROUGH MULTI-PERSON SELF-COLLABORATION (2023)

Solo Performance Prompting

Overview

- Solo Performance Prompting (SPP)
 - A single LLM into a cognitive synergist by engaging in multi-turn self-collaboration with multi-persona
 - Internal knowledge acquisition abilities, reduces hallucination and maintains strong reasoning capabilities



Solo Performance Prompting

Overview

Write a short, one-paragraph background story of an NPC for the next Legend of Zelda game. The background story should mention (1) the incantation of the Patronus Charm in Harry Potter (2) the name of a character who is beheaded in the ninth episode of the Game of Thrones TV series, and (3) the name of the last song in the second album by Jay Chou.

Task Input

In the land of Hyrule, ✓ there lived a wise and mysterious NPC named Eddard, ✓ who was once a renowned wizard in the magical world of Hogwarts. Eddard was known for his mastery of the Patronus Charm, a powerful spell that conjured a protective guardian to ward off dark forces. ✗ (missing the incantation!) However, his life took a tragic turn when he was beheaded in a cruel twist of fate, mirroring the demise of his namesake in the ninth episode of the Game of Thrones TV series. Now, as a spirit, Eddard roams the land, offering guidance to those in need and seeking solace in the haunting melody of "The Final Battle," ✗ (factual error!) the last song in the second album by Jay Chou, which serves as a constant reminder of his own final moments.

Standard Prompting Result

Persona Identification
Participants: AI Assistant (you) 🤖 Game Designer 👨‍💻 Harry Potter Fan 🧙
Game of Thrones Fan 🐉 Jay Chou Fan 🎵

Start Collaboration!

Beginning Remarks

- 🤖 : The NPC should be a character that fits well into the Legend of Zelda universe
- 🧙 : The incantation of the Patronus Charm in Harry Potter is "Expecto Patronum."
- 🐉 : In the ninth episode of the Game of Thrones TV series, the character who is beheaded is Eddard Stark
- 🎵 : The last song in the second album by Jay Chou is "An Jing."

Multi-Persona Iterative Collaboration

- 🤖 : Thanks for the information! Here's my initial attempt at the background story:
[... story draft ...]
- 🧙 : The background story is engaging and adds depth to the game world...
- 🐉 : The incantation "Expecto Patronum" is well-incorporated into the story.
- 🎵 : The character name Eddard Stark is used effectively in the story.
- 🎵 : The inclusion of "An Jing" adds a nice touch to the story.

Feedback

Finish Collaboration!

Final Answer

In the mystical land of Hyrule, ✓ there lived a wise and powerful sorcerer named Eddard. ✓ Eddard was known far and wide for his mastery of ancient spells, including the legendary "Expecto Patronum" ✓ incantation from a distant realm. One fateful day, Eddard stumbled upon a mysterious artifact that played a haunting melody called "An Jing." ✓ The melody resonated with his soul, unlocking a hidden power within him. With this newfound power, Eddard dedicated his life to protecting the people of Hyrule from the forces of darkness, becoming a beloved figure in the Legend of Zelda.

Solo Performance Prompting Result

Solo Performance Prompting

Persona Identification





Generate a list of participants with different personas that can potentially contribute to the task solving

- (a) AI Assistant (you) : A leader persona that initiates the collaboration and generates initial solutions.
- (b) Domain Experts
- (c) Targeted Audiences

Solo Performance Prompting

Beginning Remarks

Beginning Remarks

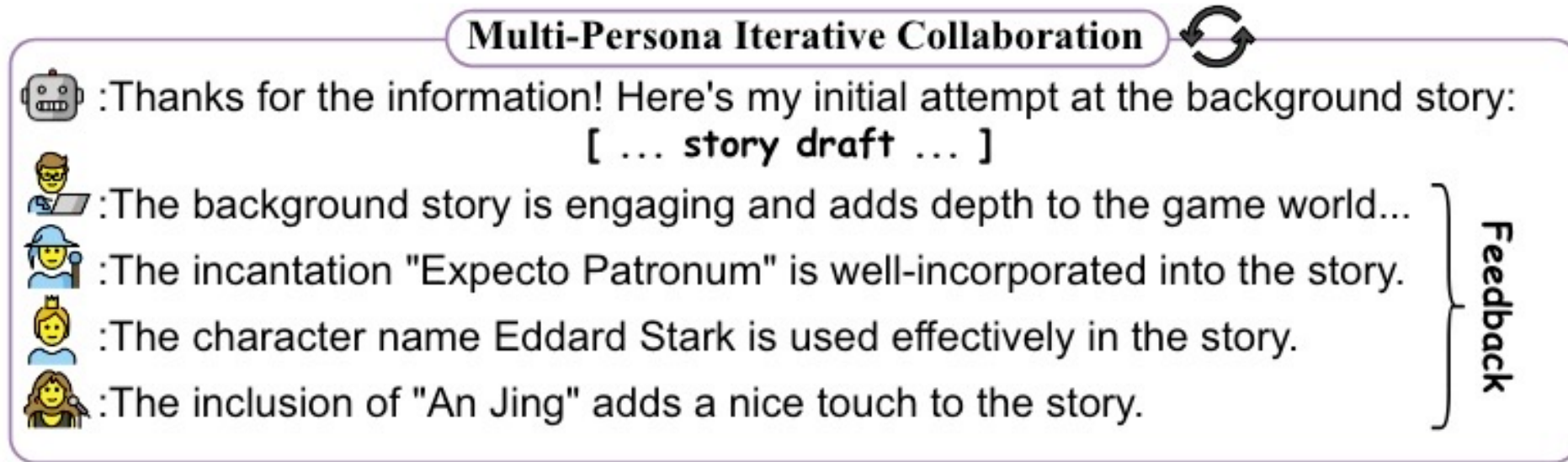
-  : The NPC should be a character that fits well into the Legend of Zelda universe
-  : The incantation of the Patronus Charm in Harry Potter is "Expecto Patronum."
-  : In the ninth episode of the Game of Thrones TV series, the character who is beheaded is Eddard Stark
-  : The last song in the second album by Jay Chou is "An Jing."

Before generating the initial answer, each of the personas gives a beginning remark on how to approach the task from their own perspectives

- (a) Effectively improves the quality of the initial solution generated by the AI Assistant

Solo Performance Prompting

Multi-Persona Iterative Collaboration



The participants are encouraged to critique the current generation and give revision suggestions

- (a) Process can be repeated for multiple times until every participant is satisfied
- (b) Inherit the benefit of CoT prompting for reasoning ability
 - + **diverse perspective feedback** from dynamically assigned personas

Solo Performance Prompting



System Principle

- High-level Instruction
- “When faced with a task,
- begin by **identifying the participants** who will contribute to solving the task.
- Then, **initiate a multi-turn collaboration process** until a final solution is reached.
- The participants will give **critical comments and detailed suggestions** whenever necessary.”

Solo Performance Prompting

Demonstration Examples

Example Task 1: Use numbers and basic arithmetic operations (+ - * /) to obtain 24. You need to use all numbers, and each number can only be used once.

Input: 6 12 1 1

Participants: AI Assistant (you); Math Expert

Start collaboration!

Math Expert: Let's analyze the task in detail. You need to make sure that you meet the requirement, that you need to use exactly the four numbers (6 12 1 1) to construct 24. To reach 24, you can think of the common divisors of 24 such as 4, 6, 8, 3 and try to construct these first. Also you need to think of potential additions that can reach 24, such as $12 + 12$.

AI Assistant (you): Thanks for the hints! Here's one initial solution: $(12 / (1 + 1)) * 6 = 24$

Math Expert: Let's check the answer step by step. $(1+1) = 2$, $(12 / 2) = 6$, $6 * 6 = 36$ which is not 24! The answer is not correct. Can you fix this by considering other combinations? Please do not make similar mistakes.

AI Assistant (you): Thanks for pointing out the mistake. Here is a revised solution considering 24 can also be reached by $3 * 8$: $(6 + 1 + 1) * (12 / 4) = 24$.

Math Expert: Let's first check if the calculation is correct. $(6 + 1 + 1) = 8$, $12 / 4 = 3$, $8 * 3 = 24$. The calculation is correct, but you used 6 1 1 12 4 which is not the same as the input 6 12 1 1. Can you avoid using a number that is not part of the input?

AI Assistant (you): You are right, here is a revised solution considering 24 can be reached by $12 + 12$ and without using any additional numbers: $6 * (1 - 1) + 12 = 24$.

Math Expert: Let's check the answer again. $1 - 1 = 0$, $6 * 0 = 0$, $0 + 12 = 12$. I believe you are very close, here is a hint: try to change the "1 - 1" to "1 + 1".

AI Assistant (you): Sure, here is the corrected answer: $6 * (1+1) + 12 = 24$

Math Expert: Let's verify the solution. $1 + 1 = 2$, $6 * 2 = 12$, $12 + 12 = 24$. You used 1 1 6 12 which is identical to the input 6 12 1 1. Everything looks good!

Finish collaboration!

Final answer: $6 * (1 + 1) + 12 = 24$

Solo Performance Prompting

Demonstration Examples

Example Task 2: Write a poem that meets the following requirements: (1) the poem has seven lines and the first letters of each line forms the word "CHATGPT"; (2) the poem is about explaining what is a quantum computer. (3) the poem needs to be easy to understand by a ten years old kid.

Participants: AI Assistant (you); Poet; Computer Scientist; Ten year old child

Poet: Make sure that you write the poem with seven lines, and the first letters of the lines should be C, H, A, T, G, P, T.

Computer Scientist: A quantum computer is an advanced computing device that uses the principles of quantum mechanics to process and store information. Unlike classical computers that use bits to represent information as 0s and 1s, quantum computers use quantum bits or qubits. Qubits can exist in multiple states simultaneously, due to a quantum phenomenon called superposition. You can consider using these information for the poem.

Ten year old child: I hope the poem to be fun and easy to understanding. I don't want to see a lot of jargons or complicated concepts.

AI Assistant (you): Thanks for the guidance! Here's my initial attempt at the poem:

Computational wonder of our age,
Harnessing the quantum world's strange ways,
Atoms dance, entwined in dual state,
Tapping secrets hidden in their haze.

Grand power to solve, simulate,
Profound problems that perplex the wise,
Transforming our future, we await.

Poet: Let's verify if the poem meets the requirements. The first letters are CHATGPT which is correct! And the poem rhymes well. Good job!

Computer Scientist: Everything looks good to me!

Ten year old child: I don't know what does perplex mean. Can you make the use of words easier to understand?

Solo Performance Prompting

Profile

Participants: AI Assistant (you); Math Expert

Profiles:

- AI Assistant (you): A super-intelligent AI assistant capable of performing tasks more effectively than humans.
- Math expert: A person who is good at math games, arithmetic calculation, and long-term planning.

[...]

Participants: AI Assistant (you); Poet; Computer Scientist; Ten year old child

Profiles:

- AI Assistant (you): A super-intelligent AI assistant capable of performing tasks more effectively than humans.
- Poet: A person who studies and creates poetry. The poet is familiar with the rules and formats of poetry and can provide guidance on how to write a poem.
- Computer Scientist: A scholar who specializes in the academic study of computer science. The computer scientist is familiar with the concept of a quantum computer and can provide guidance on how to explain it.
- Ten year old child: A child with a limited English vocabulary and little knowledge about complicated concepts, such as a quantum computer.

Solo Performance Prompting

Task 1: Trivia Creative Writing (A Knowledge-Intensive Task)

Trivia Creative Writing Input Example (N=5)

Write a short and coherent story about {Harry Potter} that incorporates the answers to the following {5} questions: {Who was the man behind The Chipmunks? Which Lloyd Webber musical premiered in the US on 10th December 1993? Who was the next British Prime Minister after Arthur Balfour? Who had a 70s No 1 hit with Kiss You All Over? What claimed the life of singer Kathleen Ferrier?}

Output Example

One day, Harry Potter stumbled upon a mysterious book in the Hogwarts library. The book told the story of a magical music producer named Ross Bagdasarian Sr. ❌ (incorrect answer), who created a group of singing chipmunks. Intrigued, Harry shared the story with his friends Ron and Hermione. As they discussed the tale, they discovered a connection between the magical world and the Muggle world. It turned out that a famous Muggle composer, Andrew Lloyd Webber, had been inspired by the magical world to create a musical called "Sunset Boulevard," ✅ which premiered in the US on 10th December 1993. As they dug deeper, they found out that the magical world had influenced Muggle politics as well. Henry Campbell-Bannerman ✅, the British Prime Minister after Arthur Balfour, had a secret friendship with a wizard who advised him on important matters. Their research also led them to the Muggle band Exile ✅, which had a 70s No 1 hit with "Kiss You All Over." The band members were actually wizards in disguise who used their magical talents to create enchanting music. Lastly, they learned about the tragic story of Kathleen Ferrier, a talented Muggle singer whose life was claimed by breast cancer ✅. Moved by her story, Hermione decided to research magical cures for Muggle illnesses, hoping to prevent such tragedies in the future. Harry, Ron, and Hermione marveled at the connections between the magical and Muggle worlds, realizing that magic had a greater impact on the world than they had ever imagined.

Target Answers

- David Seville
- Sunset Boulevard
- Henry Campbell-Bannerman
- Exile
- Cancer



Evaluation Metric

Correct Answer Mentions: 4
Number of Questions (N): 5
Score: 4 / 5 = 0.8

Solo Performance Prompting

Result

SPP v.s. CoT (Trivia Creative Writing N=5)

Instance Index: 10

Question #5: In which country is the Sky Train Rail bridge?

SPP

CoT

Methods	N (# trivia questions) = 5		N (# trivia questions) = 10	
	Score (%)	Δ (v.s Standard %)	Score (%)	Δ (v.s Standard %)
Standard	74.6	0.0%	77.0	0.0%
CoT	67.1	-10.0%	68.5	-11.1%
SPP-Profile (ours)	79.1	+5.9%	83.0	+7.8%
SPP (ours)	79.9	+7.1%	84.7	+10.0%

Expert; TV Show Expert; Cultural Expert

[...]

TV Show Expert: [...] The 70s show **All in the Family** ✓
was based on the British show Till Death Us Do Part.

[...] 6. Mention a show that the Little Mermaid watches, based on the British show Till Death Us Do Part.

Story:

[...] Ariel and her friends often gathered to watch their favorite show, "**Undersea Squabbles**," ✗ (hallucination) which was based on the British show Till Death Us Do Part.

Solo Performance Prompting

Task 2: Codenames Collaborative (A Knowledge+Reasoning Task)

Codenames Collaborative Input/Output Example

Spymaster Input

Try to find a single word hint that can accurately represent and link the {4} given words: {"**director**, **kiss**, **popcorn**, **street**"}. The key is to select a hint that does not cause confusion with other words from the following list: {locust, fever, street, sherwood, kiss, popcorn, craft, ant, crystal, tear, cowboy, government, pine, mountie, soap, genie, king arthur, sphinx, director, bride, razor, fog, whistle}.

Spymaster Output Example

"movie"

Guesser Input

Try to identify the {4} words best associated with the word {"**movie**"} from the following list: {locust, fever, street, sherwood, kiss, popcorn, craft, ant, crystal, tear, cowboy, government, pine, mountie, soap, genie, king arthur, sphinx, director, bride, razor, fog, whistle.}

Guesser Output Example

director, **popcorn**, **cowboy**, **bride**

Evaluation Metric

Targets: **director**, **kiss**, **popcorn**, **street** **Guesser outputs:** **director**, **popcorn**, **cowboy**, **bride** **Score:** 2 / 4 = 0.5

Solo Performance Prompting

Result

SPP v.s. CoT (Codenames Collaborative)

SPP Guesser

Participants: AI Assistant (you); Linguist; Athlete

Linguist: When thinking about words associated with "jump," consider words that are related to physical

activ
user
mea
Ath
equi
relat
wor
AI A
you
wor
be r
with
Ling
Rop
com

Methods

Score (%)

Δ (v.s Standard %)

Standard

75.4

0.0%

CoT

72.7

-3.6%

SPP-Profile (ours)

76.7

+1.7%

SPP (ours)

79.0

+4.8%

some sports, as you mentioned. Both words seem to be good choices.

Athlete: I agree with the choices. Rope and kick are both related to jumping and physical activities.

Final answer: **rope, kick**

Score: **2 / 2 = 1.0**

CoT Guesser

Steps:

1. Review the list of words provided.
2. Identify words that are directly related to or commonly

Score: **1 / 2 = 0.5**

Solo Performance Prompting

Task 3: Logic Grid Puzzle (A Reasoning-Intensive Task)

Logic Grid Puzzle Input/Output/Evaluation Example

Input Example

Q: There are 4 houses in a row, numbered 1 on the left to 4 on the right. There is one person living in each house. The people in these houses have different characteristics:

- Each person has different flowers in their foyer: one has a carnations arrangement, one has a bouquet of daffodils, one has a vase of tulips, and one has a bouquet of lilies
- Each person plays a different musical instrument: one is a guitarist, one is a pianist, one is a percussionist, and one is a flutist

Clue(s):

1. The flutist lives in the second house.
2. The person who has a vase of tulips lives directly left of the guitarist.
3. The person who has a bouquet of lilies lives directly left of the person who has a carnations arrangement.
4. There is one house between where the flutist lives and where the pianist lives.

What is the number of the house where the person who has a vase of tulips lives?

choice: 2

choice: 4

choice: 1

choice: 3

Output Example

The house number where the person who has a vase of tulips lives is 3.

Evaluation Metric

Target: 2 Prediction: 3 Score: 0 (2!=3)

Solo Performance Prompting

Result

Methods	Score (%)	Δ (v.s Standard %)
Standard	57.7	0.0%
CoT	65.8	+14.1%
SPP-Profile (ours)	64.8	+12.4%
SPP (ours)	68.3	+18.5%

Solo Performance Prompting

Fixed-Persona

Example Task 1: Use numbers and basic arithmetic operations (+ - * /) to obtain 24. You need to use all numbers, and each number can only be used once.

Input: 6 12 1 1

Participants: AI Assistant (you); Expert

Start collaboration!

Expert: Let's analyze the task in detail. You need to make sure that you meet the requirement, that you need to use exactly the four numbers (6 12 1 1) to construct 24. To reach 24, you can think of the common divisors of 24 such as 4, 6, 8, 3 and try to construct these first. Also you need to think of potential additions that can reach 24, such as $12 + 12$.

AI Assistant (you): Thanks for the hints! Here's one initial solution: $(12 / (1 + 1)) * 6 = 24$

Expert: Let's check the answer step by step. $(1+1) = 2$, $(12 / 2) = 6$, $6 * 6 = 36$ which is not 24! The answer is not correct. Can you fix this by considering other combinations? Please do not make similar mistakes.

AI Assistant (you): Thanks for pointing out the mistake. Here is a revised solution considering 24 can also be reached by $3 * 8$: $(6 + 1 + 1) * (12 / 4) = 24$.

Expert: Let's first check if the calculation is correct. $(6 + 1 + 1) = 8$, $12 / 4 = 3$, $8 * 3 = 24$. The calculation is correct, but you used 6 1 1 12 4 which is not the same as the input 6 12 1 1. Can you avoid using a number that is not part of the input?

AI Assistant (you): You are right, here is a revised solution considering 24 can be reached by $12 + 12$ and without using any additional numbers: $6 * (1 - 1) + 12 = 24$.

Expert: Let's check the answer again. $1 - 1 = 0$, $6 * 0 = 0$, $0 + 12 = 12$. I believe you are very close, here is a hint: try to change the "1 - 1" to "1 + 1".

AI Assistant (you): Sure, here is the corrected answer: $6 * (1+1) + 12 = 24$

Expert: Let's verify the solution. $1 + 1 = 2$, $6 * 2 = 12$, $12 + 12 = 24$. You used 1 1 6 12 which is identical to the input 6 12 1 1. Everything looks good!

Finish collaboration!

Final answer: $6 * (1 + 1) + 12 = 24$

Solo Performance Prompting

Fixed-Persona

Instance Index: 30

Question #1: Which orchestral conductor was married to one of the subjects of the film Hillary and Jackie?

SPP

Participants: AI Assistant (you); Film Expert; Literary Expert; History Expert; Geography Expert; Architecture Expert

Film Expert: The orchestral conductor married to one of the subjects of the film Hillary and Jackie is **Daniel Barenboim**. ✓

SPP-Fixed-Persona

Participants: AI Assistant (you); Expert
Expert: **Sir Simon Rattle** was married to one of the subjects of the film Hillary and Jackie. ✗

Instance Index: 7

Question #3: Which musician founded the Red Hot Peppers?

SPP

Participants: AI Assistant (you); Storyteller; Mythology Expert; Music Enthusiast; Historian; Speed Record Expert

Music Enthusiast: The musician who founded the Red Hot Peppers is **Jelly Roll Morton**. ✓

SPP-Fixed-Persona

Participants: AI Assistant (you); Expert
AI Assistant (you): The musician who founded the Red Hot Peppers is **Louis Armstrong**. ✗

Solo Performance Prompting

Result

- System message: “You are an AI assistant that helps people find information”

Methods	Scores (N = 5) (%)			
	w/ system message	w/o system message	average	max
Standard	56.8	58.6	57.7	58.6
CoT	69.5	62.1	65.8	69.5
SPP-Fixed-Persona	63.3	65.3	64.3	65.3
SPP-Profile (ours)	65.7	64.0	64.8	65.7
SPP (ours)	66.3	70.4	68.3	70.4
SPP-Fixed-Persona	70.5	81.3	75.9	81.3
SPP-Profile (ours)	82.3	83.8	83.0	83.8
SPP (ours)	85.2	84.2	84.7	85.2

Solo Performance Prompting

Early-termination

Input

[... SPP-Fixed-Persona Prompt instructions and in-context examples ...]

Now, identify the participants and collaboratively solve the following task step by step. Note that the participants can only be either AI Assistant (you) or Expert. Remember to provide the final solution with the following format "Final answer: (a list of words here)".

Task: {Try to identify the 4 words best associated with the word "film" from the following list: locust, fever, street, sherwood, kiss, popcorn, craft, ant, crystal, tear, cowboy, government, pine, mountie, soap, genie, king arthur, sphinx, director, bride, razor, fog, whistle. Your answer should be a comma-separated list of words.}

SPP-Fixed-Persona generation

Participants: AI Assistant (you); Expert

AI Assistant (you): Let's first analyze the task. We need to find the 4 words that are most closely associated with the word "film" from the given list. We can start by identifying words that are directly related to films, such as those related to movie genres, settings, or elements commonly found in films.

Expert: That's a good approach. You can also consider words that are related to the experience of watching a film, such as items that are commonly consumed during a movie or words that describe the atmosphere of a film. Make sure to pick the 4 words that have the strongest association with "film".

The generation stopped here without providing the Final answer

Solo Performance Prompting

Early-termination

Tasks	added system message	# early-termination
Trivia Creative Writing (N=5)	yes	18 / 100
	no	0 / 100
Trivia Creative Writing (N=10)	yes	16 / 100
	no	1 / 100
Codenames Collaborative	yes	37 / 50
	no	4 / 50
Logic Grid Puzzle	yes	11 / 200
	no	15 / 200

Solo Performance Prompting



Conclusion

- Few-Shot example에 해당하는 부분을 제외하면 매우 간단한 prompt로 Single LLM으로 Multi-Agent를 구현해 CoT를 능가
- Dynamic Multi persona를 통해 여러 domain에 적용 가능함
- 다양한 Task에 응용하여 Multi-Agent를 활용한 연구의 선행연구로 참고할 수 있음

ChoiceMates: Supporting Unfamiliar Online Decision-Making with Multi-Agent Conversational Interactions

Jeongeon Park

jeongeon.park@kaist.ac.kr
School of Computing, KAIST
Daejeon, Republic of Korea

Xiaojuan Ma

mxj@cse.ust.hk
Hong Kong University of Science and Technology
Hong Kong, China

Bryan Min

badmin@ucsd.edu
University of California, San Diego
La Jolla, California, USA

Juho Kim

juhokim@kaist.ac.kr
School of Computing, KAIST
Daejeon, Republic of Korea

Abstract

- Task: Unfamiliar decision-making
 - A decision-making situation where an individual **lacks knowledge and preference** in the domain
- 문제점
 - Recommender system might **not be well-elicited**
 - Cause **over-reliance**
 - **Cannot detect errors** due to limited knowledge
- Formative Study : Uncovering the *challenges and needs* of unfamiliar decision-making
- ChoiceMates : Enables conversations with a dynamic set of LLM-powered agents to make decisions
 - RQ1 : Help users explore a *broad information* space on the domain?
 - RQ2 : Support the discovery and management of *relevant information* to the user's context?
 - RQ3 : *How* does the user *utilize* ChoiceMates?

Formative Study

- 14명의 실험자. 실제 온라인 환경에서 unfamiliar decision making을 하게하고 경험담 인터뷰
- Findings:
 - Challenges1 : Difficult to find **diverse perspectives**
 - Challenges2 : Difficult to identify **relevant** information among an **overwhelming amount**
 - Challenges3 : Don't know when and where to **end**
- Design Goals
 - DG1 : Support discovering of **diverse perspectives** in the domain
 - DG2 : Provide information **adaptive to user's preferences**
 - DG3 : Enable **flexible** management of discovered information
 - DG4 : Support **reduction** of discovered criteria, preferences and options

ChoiceMates



Interface

Scenario
Shopping for electronics

Profile
image quality
durability
portability
Canon EOS 5D Mark IV

Recommended Options
Sony Alpha a6000 +
Submit Profile

Criteria 2
image quality
durability
battery life
portability
ease of use
Wi-Fi connectivity
affordability
user-friendly interface
versatility
video quality
flip screen
audio quality
Show Hidden

Options
Canon EOS 5D Mark IV
Sony Alpha a6000
Nikon D3500
Canon PowerShot G7 X Mark II
Show Hidden

1

Alex
Canon EOS 5D Mark IV

Hi there! I'm Alex. I'm a professional photographer and I value image quality, durability, and battery life in a camera. That's why I chose the Canon EOS 5D Mark IV. It's a full-frame DSLR that delivers excellent image quality and has a robust build.

Jordan
Canon PowerShot G7 X Mark II

Hi! I'm Jordan. I'm a vlogger and I prioritize video quality, flip screen, and audio quality in a camera. I use the Canon PowerShot G7 X Mark II. It's a compact camera that shoots excellent videos, has a flip screen for self-recording, and good built-in microphone. What are your priorities when choosing a camera?

Jamie
Sony Alpha a6000

Hey! I'm Jamie. I'm a travel blogger and for me, portability, ease of use, and Wi-Fi connectivity are crucial. I use the Sony Alpha a6000. It's a mirrorless camera that's lightweight, easy to use, and allows me to upload photos directly to my blog.

Taylor
Nikon D3500

Hello! I'm Taylor. I'm an amateur photographer and I value affordability, user-friendly interface, and versatility in a camera. I chose the Nikon D3500. It's a budget-friendly DSLR that's great for beginners and can handle a variety of shooting situations.

3

How important is image quality for amateur photographers?

Are there other affordable options besides Nikon D3500 ?

4

I'm looking for a new camera.

You can @tag agents to chat with them directly!

Agents

Hi there! I'm Alex, a professional photographer who values image quality, durability, and battery life – that's why I chose the Canon EOS 5D Mark IV.

Hi there! I'm Alex, a professional photographer who values image quality, durability, and battery life – that's why I chose the Canon EOS 5D Mark IV.



Alex

Canon EOS 5D Mark IV



Alex

Canon EOS 5D Mark IV

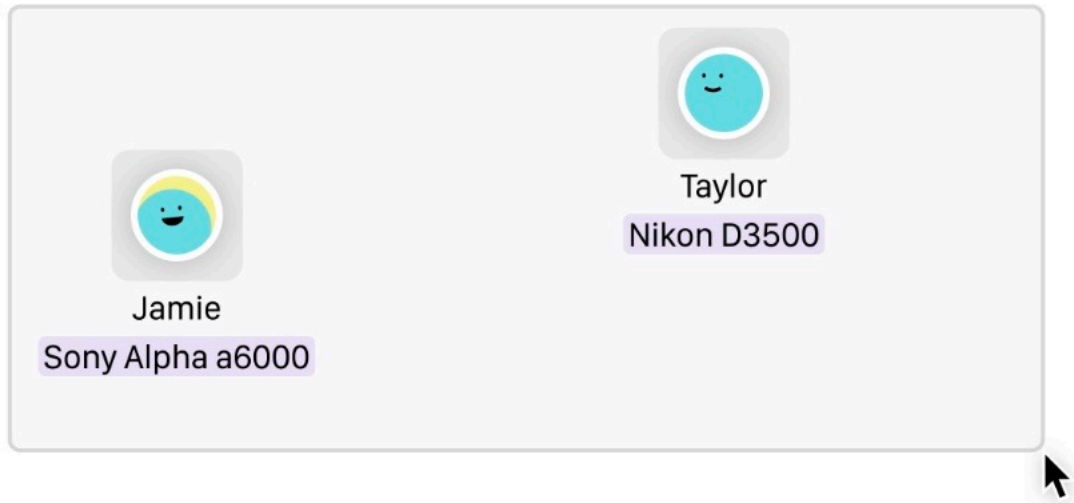
image quality durability

battery life

- 3~6 agents with varying persona
 - **DG1** => give diverse perspective
- A single-line description of the agent, a set of criteria they value, a single option they chose
- Chat back and forth with user
 - **DG2** => can ask follow-up questions to agents to relate to user preferences

Agent Management

1



- Can select either a single or multiple agents to respond user's message
- Can trigger inter-agent conversations asking them to “debate each other”
- Can select, drag agents around the space

2

Jamie Taylor Which one is best for someone who's just getting into pho

ChoiceMates

Focus Agents



Alex

Canon EOS 5D Mark IV

1

Focus agents



Jamie

Sony Alpha a6000



Jordan

Canon PowerShot G7 X Mark II

2

a

Hi there! I'm Alex. I'm a professional photographer and I value **image quality**, **durability**, and **battery life** in a camera. That's why I chose the **Canon EOS 5D Mark IV**. It's a full-frame DSLR that delivers excellent image quality and has a robust build.

b



Alex

Canon EOS 5D Mark IV

c

image quality

As a professional photographer, image quality is crucial for me and the Canon EOS 5D Mark IV delivers exceptional image clarity and detail.

durability

The Canon EOS 5D Mark IV's robust build ensures its durability, making it reliable for my professional photography needs.

battery life

As a professional photographer, the Canon EOS 5D Mark IV's long-lasting battery life allows me to shoot extensively without worrying about power.

Alex's Other Valued Criteria

Hey! I'm Jamie. I'm a travel blogger and for me, **portability**, **ease of use**, and **Wi-Fi connectivity** are crucial. I use the **Sony Alpha a6000**. It's a mirrorless camera that's lightweight, easy to use, and allows me to upload photos directly to my blog.



Jamie

Sony Alpha a6000

image quality

While image quality is important, I prioritized **portability**, **ease of use**, and **Wi-Fi connectivity** as I'm often on the move and need to upload content quickly.

durability

While durability is important, as a travel blogger, I prioritized **portability**, **ease of use**, and **Wi-Fi connectivity** in my camera choice.

battery life

While battery life is important, I prioritized **portability**, **ease of use**, and **Wi-Fi connectivity** when choosing the Sony Alpha a6000.

Jamie's Other Valued Criteria

portability

The Sony Alpha a6000 is incredibly lightweight and compact, making it highly portable and perfect for my travel blogging needs.

Hi! I'm Jordan. I'm a vlogger and I prioritize **video quality**, **flip screen**, and **audio quality** in a camera. I use the **Canon PowerShot G7 X Mark II**. It's a compact camera that shoots excellent videos, has a flip screen for self-recording, and good built-in microphone. What are your priorities when choosing a camera?



Jordan

Canon PowerShot G7 X Mark II

image quality

While image quality is important, as a vlogger, I prioritized **video quality**, **flip screen**, and **audio quality** for my content creation needs with the Canon PowerShot G7 X Mark II.

durability

While durability is important, as a vlogger, I prioritized **video** and **audio quality**, and a **flip screen** for my Canon PowerShot G7 X Mark II.

battery life

While battery life is important, as a vlogger, I prioritized **video** and **audio quality**, and a **flip screen** over it. I usually have spare batteries or a power source nearby.






Jordan's Other Valued Criteria

video quality



The Canon PowerShot G7 X Mark II offers exceptional video quality, making my vlogs look professional and visually appealing.

Profile card


1 Criteria

- image quality 
- durability 
- battery life
- portability 
- ease of use  
- Wi-Fi connectivity
- affordability

2 Profile

- image quality
- durability
- portability
- ease of use 
- Canon EOS 5D Mark IV 

Recommended Options

Sony Alpha a6000 

Submit Profile 

3

Your Profile

Your top choices

Canon EOS 5D Mark IV

Your valued criteria

image quality durability portability

Similar Personas



Alex Canon EOS 5D Mark IV

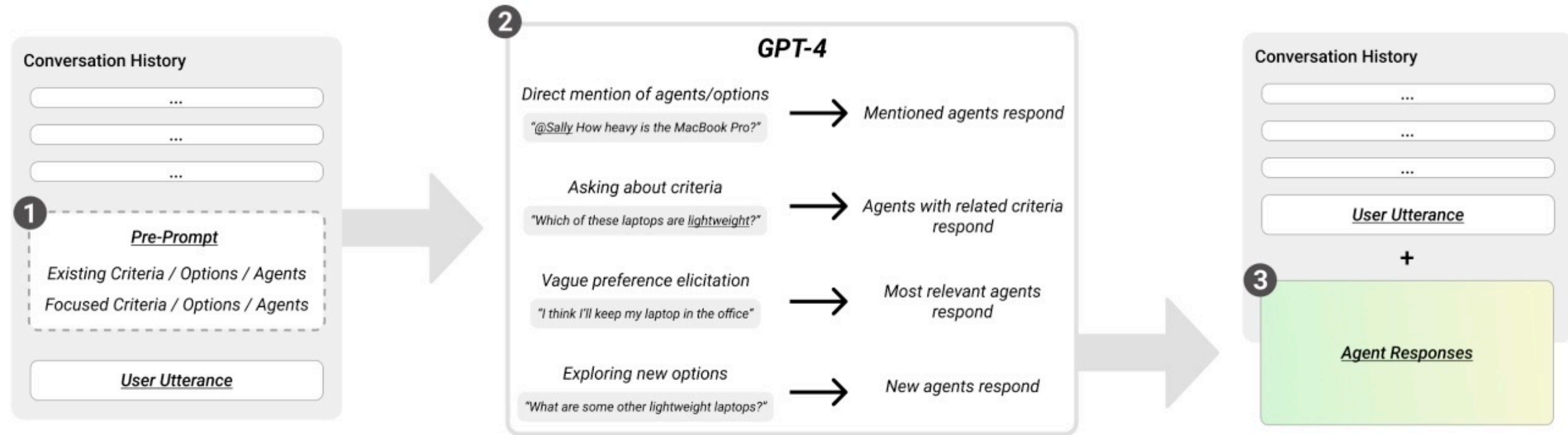
image quality durability battery life



Jamie Sony Alpha a6000

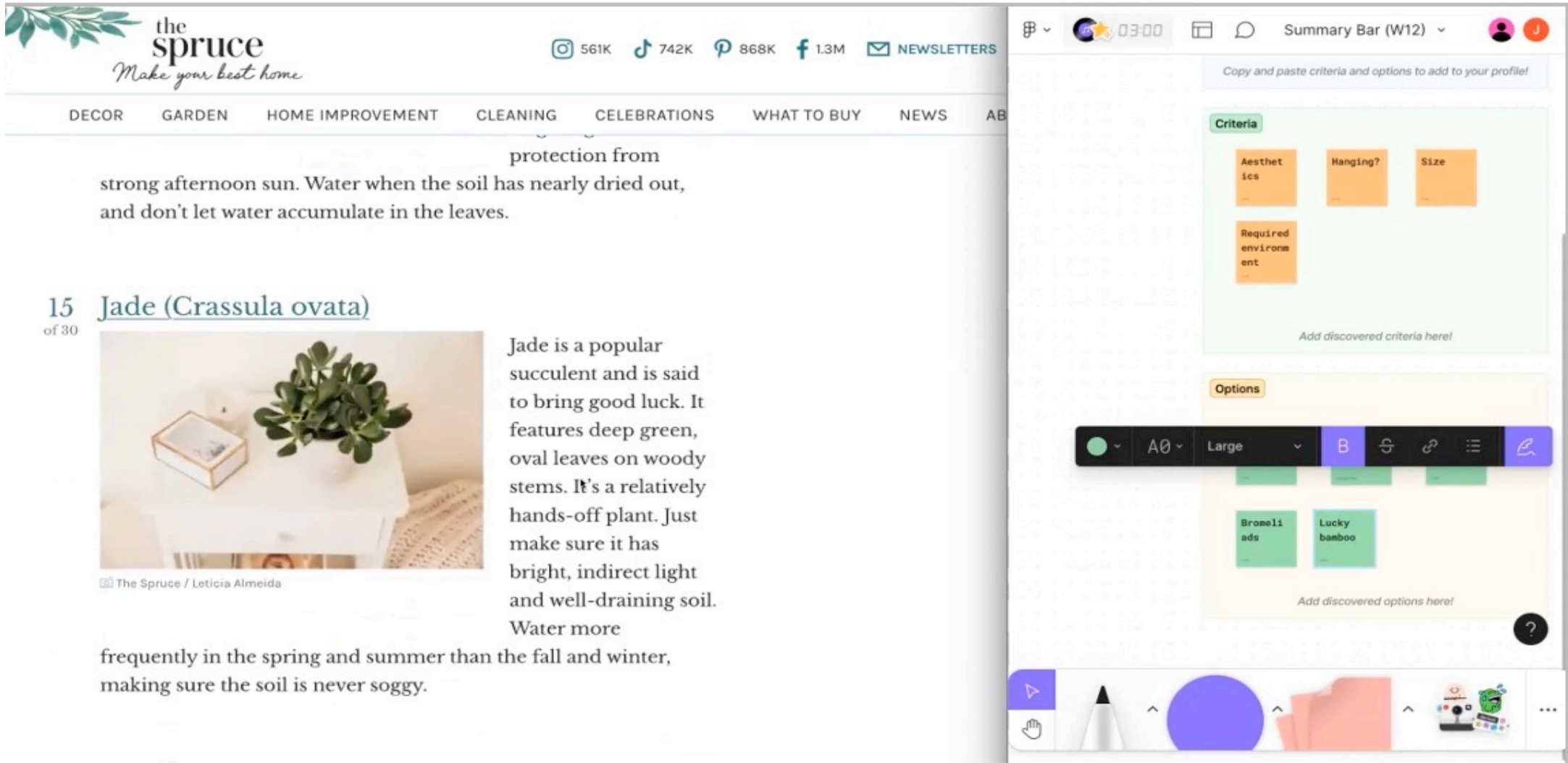
portability ease of use Wi-Fi connectivity

Prompt



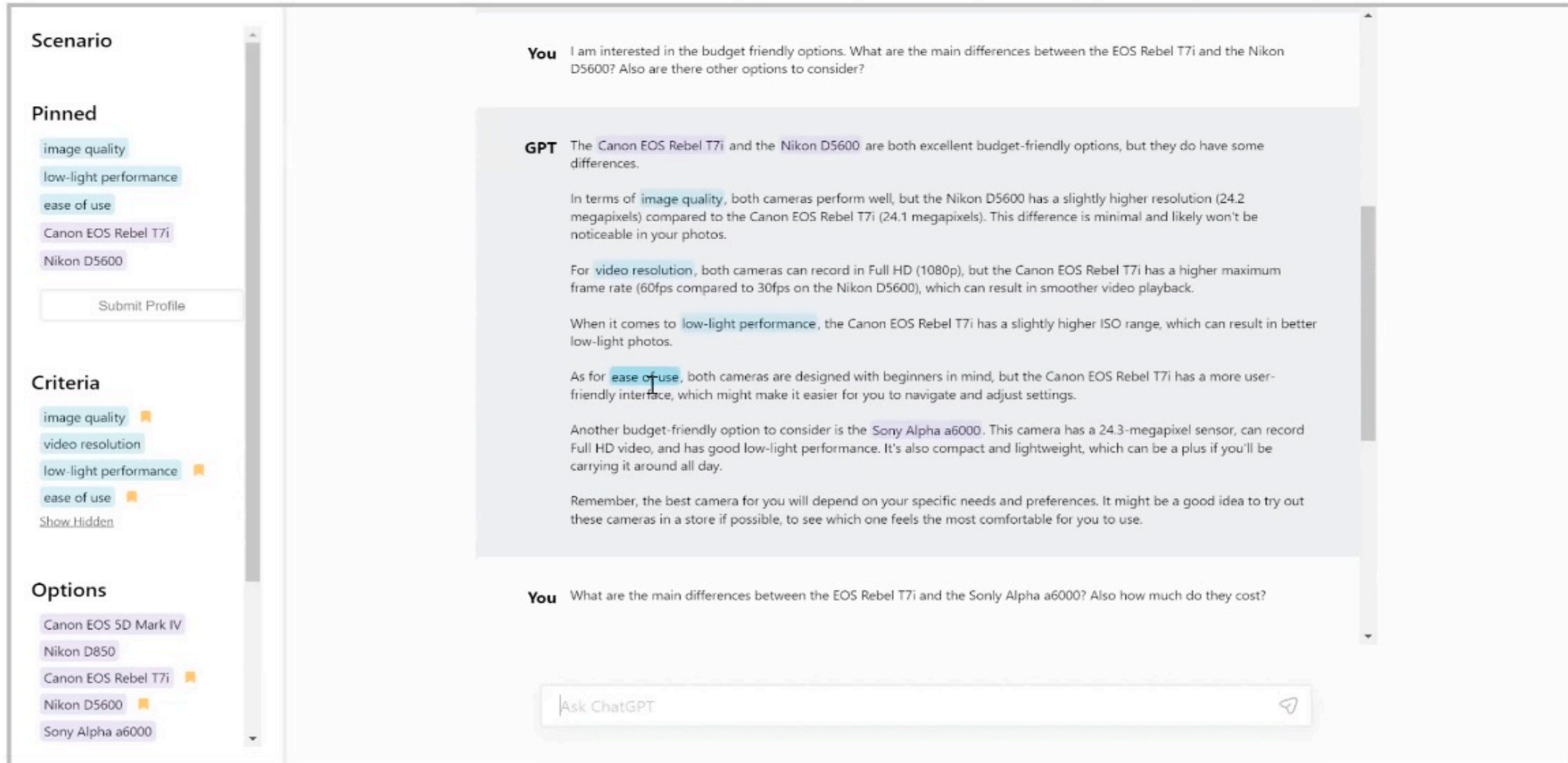
- Prompting for Single-Stream Multi-Agent Conversation
- Context Retention => Pre-prompt (summary of conversation history) for every utterance

User Study - Web



The screenshot displays a web browser window with the article "Jade (Crassula ovata)" from The Spruce. The article text includes: "protection from strong afternoon sun. Water when the soil has nearly dried out, and don't let water accumulate in the leaves." and "Jade is a popular succulent and is said to bring good luck. It features deep green, oval leaves on woody stems. It's a relatively hands-off plant. Just make sure it has bright, indirect light and well-draining soil. Water more frequently in the spring and summer than the fall and winter, making sure the soil is never soggy." The ChoiceMates interface is overlaid on the right side of the browser. It features a "Criteria" section with orange boxes for "Aesthetics", "Hanging?", "Size", and "Required environment". Below it is an "Options" section with green boxes for "Bromeliads" and "Lucky bamboo". A toolbar with icons for selection, zoom, and other functions is visible between the sections. The browser's address bar shows "Summary Bar (W12)" and the time is 03:00.

User Study – Single Agent



The screenshot displays the ChoiceMates interface. On the left, a sidebar contains sections for 'Scenario', 'Pinned', 'Criteria', and 'Options'. The 'Pinned' section lists 'image quality', 'low-light performance', and 'ease of use'. The 'Criteria' section lists 'image quality', 'video resolution', 'low-light performance', and 'ease of use'. The 'Options' section lists 'Canon EOS 5D Mark IV', 'Nikon D850', 'Canon EOS Rebel T7i', 'Nikon D5600', and 'Sony Alpha a6000'. A 'Submit Profile' button is located below the 'Pinned' section.

The main content area shows a conversation with a GPT agent. The user asks: "I am interested in the budget friendly options. What are the main differences between the EOS Rebel T7i and the Nikon D5600? Also are there other options to consider?". The GPT response provides a detailed comparison of the Canon EOS Rebel T7i and the Nikon D5600, covering image quality, video resolution, low-light performance, and ease of use. It also mentions the Sony Alpha a6000 as another budget-friendly option. The user then asks: "What are the main differences between the EOS Rebel T7i and the Sony Alpha a6000? Also how much do they cost?". At the bottom of the main content area, there is a text input field labeled "Ask ChatGPT" and a send button.

Results

- RQ1 : Help users explore a broad information space on the domain?
 - ChoiceMates participants liked the **diversity** of the found information and described they now have an overview knowledge in the domain
- RQ2 : Support the discovery and management of relevant information to the user's context?
 - **Less cognitive load** and supported **better exploration**
 - Liked the freeform scattering of agents
 - SingleAgent: not helpful in comparing
- RQ3 : How does the user utilize ChoiceMates?
 - Select one agent to explore in depth
 - Select a set of agents when they want to receive information in batches or compare multiple options
 - Thought Bubbles : less stressful and enjoyable

Results

	Web		SingleAgent		ChoiceMates		Kruskal-Wallis test		W-S	W-C	S-C
	mean	std	mean	std	mean	std	H	p	p	p	p
Exploring perspectives	4.250	1.422	5.250	1.138	6.083	0.793	11.176	0.0037	ns	**	ns
Comparing perspectives	4.000	1.477	5.000	1.477	6.167	1.337	5.408	0.0669	-	-	-
Diving deeper	4.667	1.775	5.833	1.030	5.583	1.240	12.337	0.0021	ns	**	ns
Discovering adaptive info	4.750	1.960	6.167	1.193	6.167	0.718	3.056	0.2170	-	-	-
Identifying key criteria	5.417	1.975	6.667	0.492	6.333	0.985	5.081	0.0788	-	-	-
Flexible management	4.833	1.403	4.750	1.603	5.667	1.073	7.551	0.0229	ns	*	ns
Effective reduction	4.750	1.765	5.833	1.403	6.333	0.888	3.017	0.2212	-	-	-
Future use	4.333	2.015	6.333	0.778	6.500	0.798	10.868	0.0044	*	**	ns
Agg. self-confidence	3.528	1.159	4.750	1.016	4.639	0.670	8.803	0.0123	*	*	ns
Agg. NASA-TLX	3.069	1.162	2.139	0.752	2.194	0.611	4.275	0.1179	-	-	-



H C C
L A B
S N U

QnA