[1]PEKING UNIVERSITY

[2]BIGAI

CoRe

COGSCI2025
THEORIES OF THE PAST
THEORIES OF THE FUTURE
SAN FRANCISCO · JUL.30-AUG.2

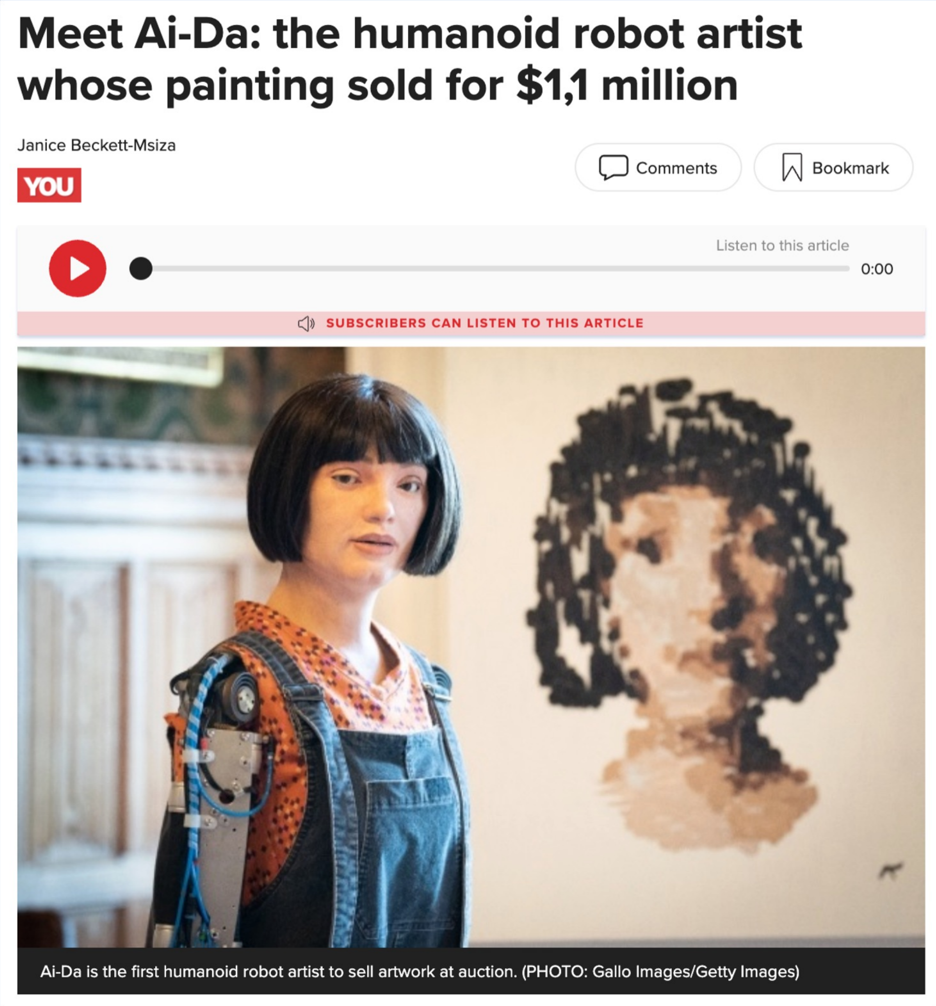
Probing and Inducing Combinational Creativity in Vision-Language Models

Yongqian Peng[*,1], Yuxi Ma[*,1], Mengmeng Wang[2], Yuxuan Wang[2], Yizhou Wang[1], Chi Zhang[2], Yixin Zhu[1,✉], Zilong Zheng[2,✉]

*equal contribution ✉corresponding authors

Background

https://www.bbc.com/news/videos/cn4vwq0v9v5o



Meet Ai-Da: the humanoid robot artist whose painting sold for $1,1 million

Janice Beckett-Msiza

YOU

Ai-Da is the first humanoid robot artist to sell artwork at auction. (PHOTO: Gallo Images/Getty Images)

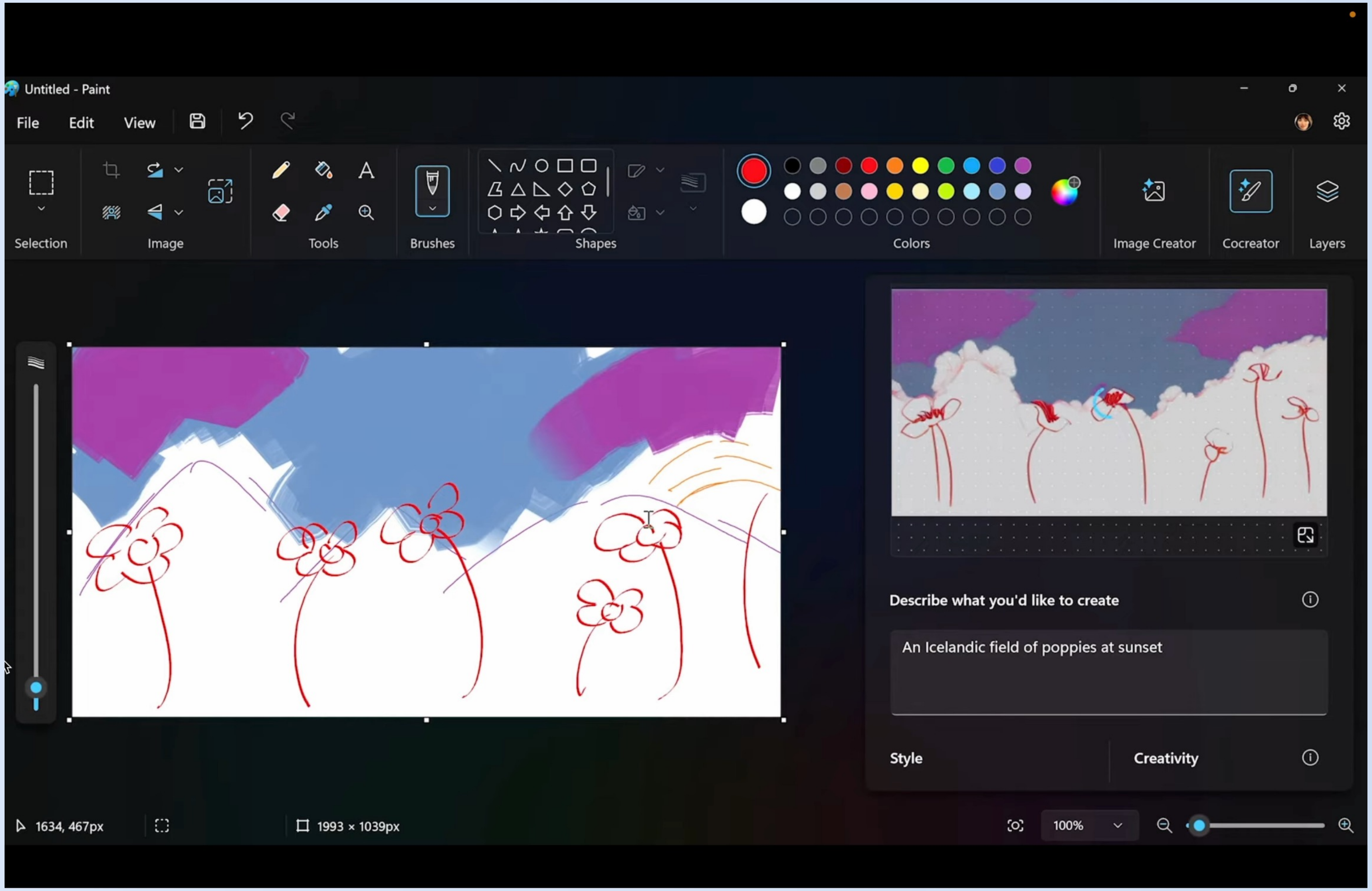
https://www.news24.com/you/news/international/meet-ai-da-the-humanoid-robot-artist-whose-painting-sold-for-11-million-20241112

Do these creative products generated by AI emerge from **genuine creative processes** or **sophisticated pattern matching**?

Why Study Machine Creativity?

“Creativity is the ability to produce work that is both novel and useful.”

- **Understanding Human Creativity Better**
- **Enhancing Human Creativity**



Describe what you'd like to create

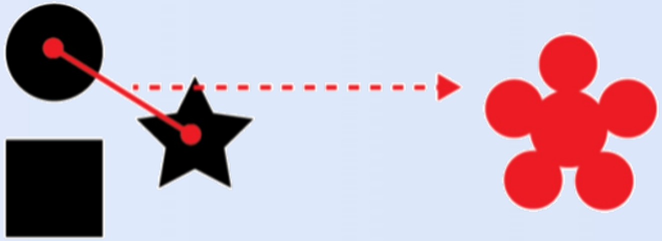
An Icelandic field of poppies at sunset

Style

Creativity

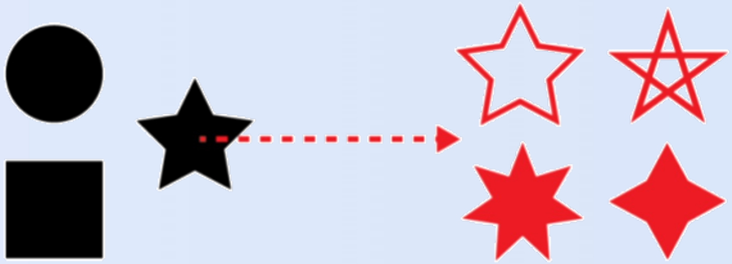
https://www.youtube.com/watch?v=VBRYaeK_cT0

Combinational Creativity in AI



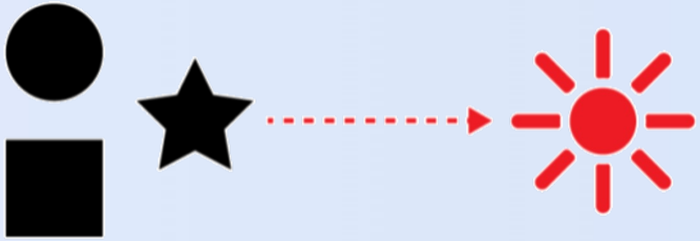
Combinational Creativity

Combining existing ideas and things into something new



Explorative Creativity

Exploring possibilities within a domain

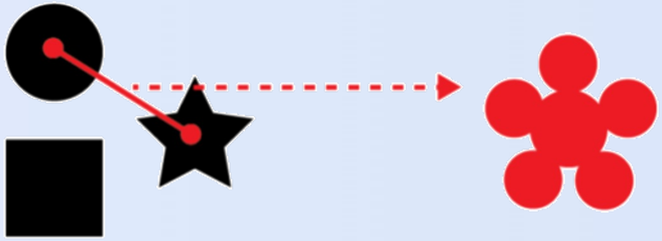


Transformational Creativity

Radically new ideas that redefine the domain and existing rules

Boden, M. A. (2007). *Creativity in a nutshell. Think, 5 (15), 83–96.*

Combinational Creativity in AI

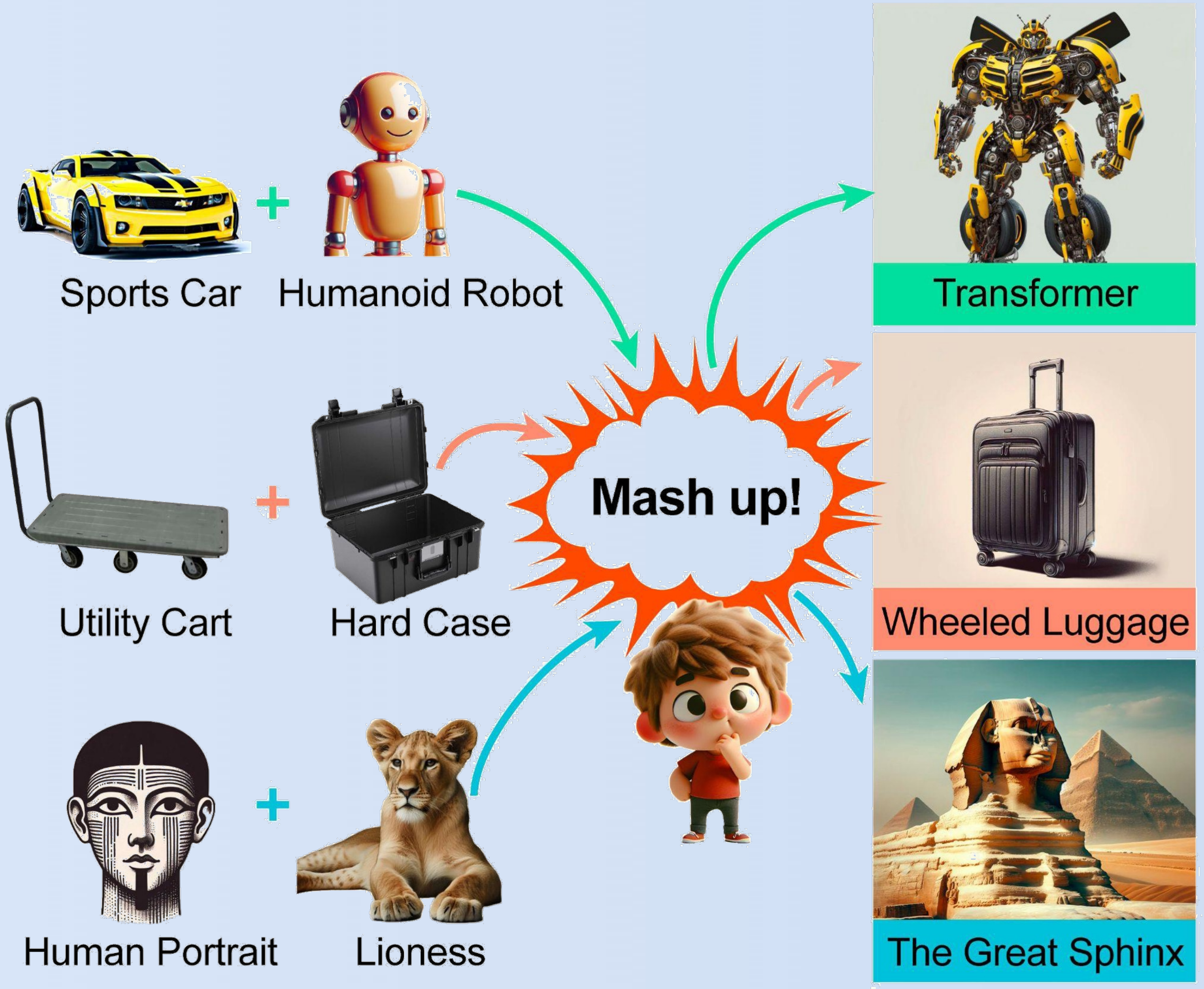


Boden, M. A. (2007). *Creativity in a nutshell. Think, 5 (15), 83–96.*

Combinational Creativity

Combining existing ideas and things into something new

- Well-defined
- Easy to implement
- Empirically dominant



Sports Car + Humanoid Robot

Utility Cart + Hard Case

Human Portrait + Lioness

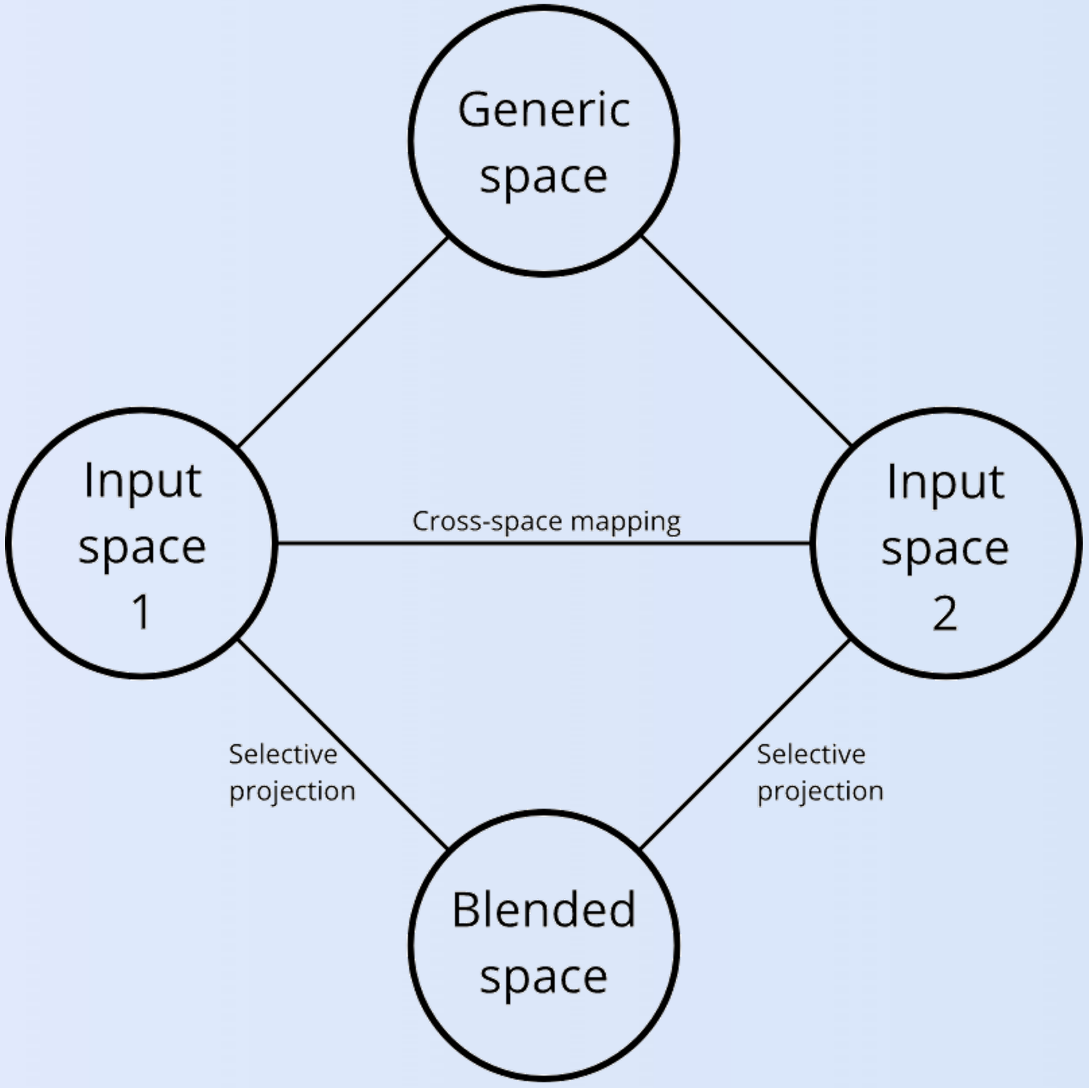
Mash up!

Transformer

Wheeled Luggage

The Great Sphinx

Conceptual Blending Theory



https://commons.wikimedia.org/wiki/File:Conceptual_blending_-_the_network_model.png

How It Works: The Four-Space Model

1. **Two Input Spaces**

   Each contains a separate scenario or concept.

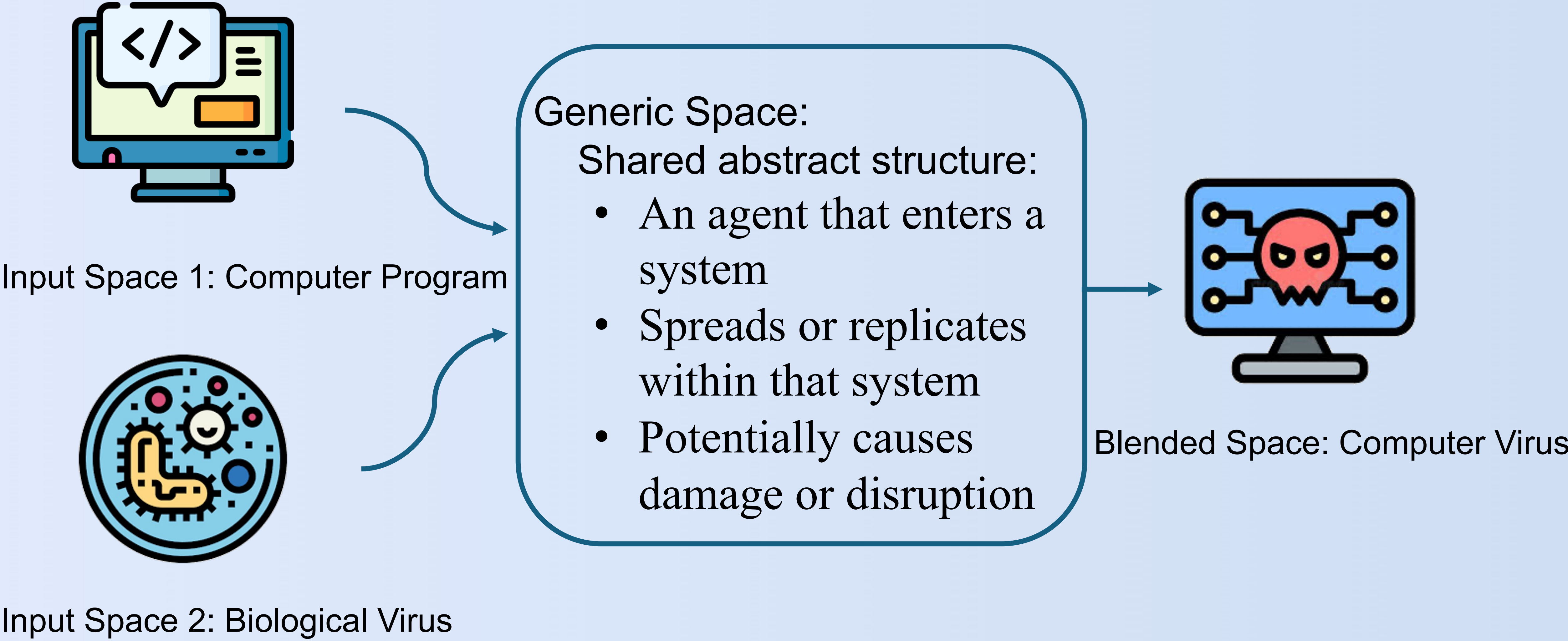
2. **Generic Space**

   Captures what is common across both inputs—shared structure or schema.

3. **Blend Space**

   Selectively projects parts from the inputs into a **new space**, yielding emergent meaning.

Conceptual Blending Theory – An Example



Input Space 1: Computer Program

Input Space 2: Biological Virus

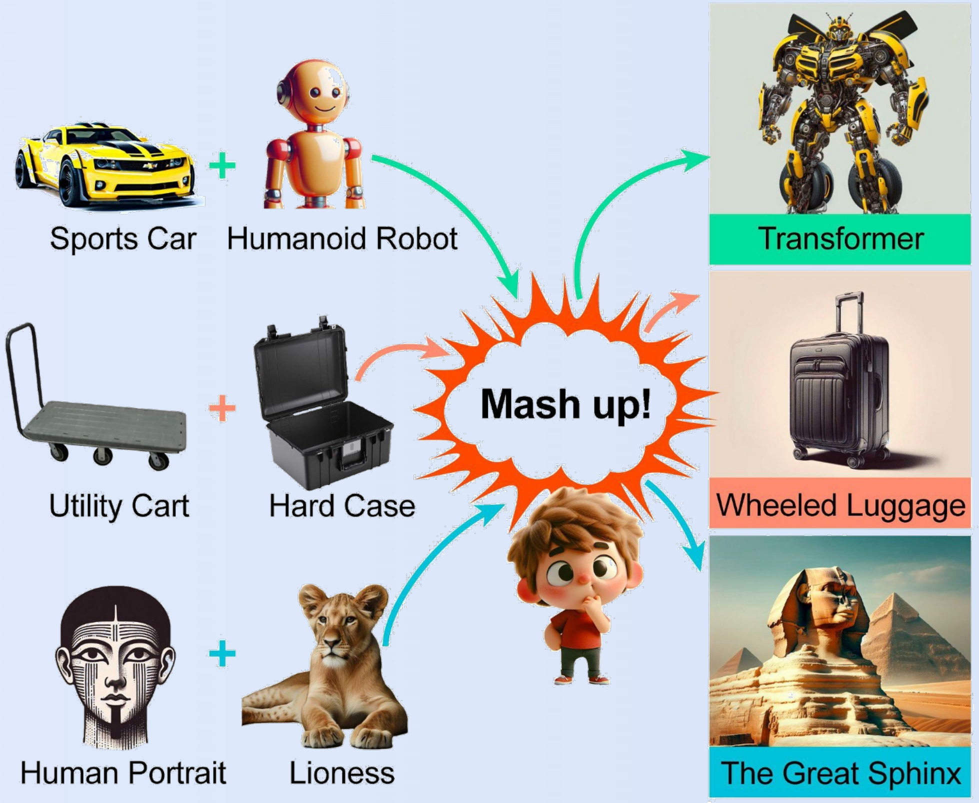
Generic Space:

Shared abstract structure:

- An agent that enters a system
- Spreads or replicates within that system
- Potentially causes damage or disruption

Blended Space: Computer Virus

Research Questions

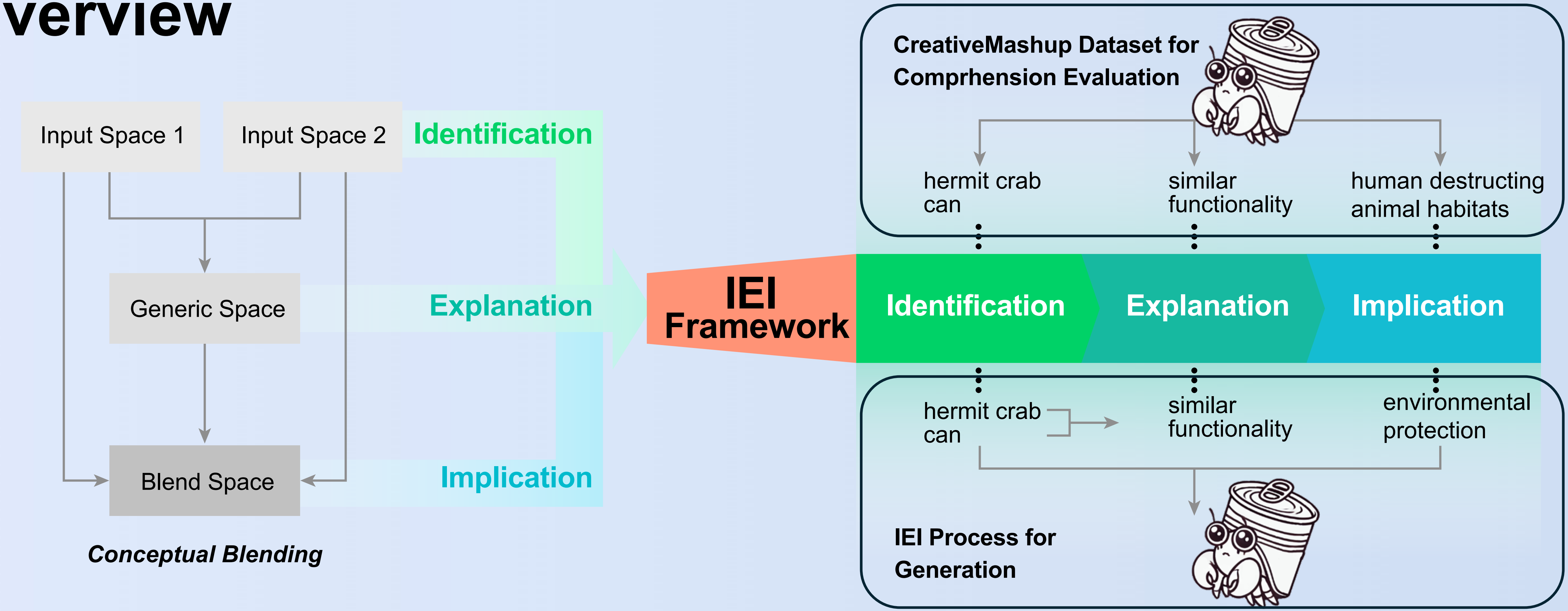


Peng, Y., Ma, Y., Wang, M., Wang, Y., Wang, Y., Zhang, C., ... & Zheng, Z. (2025). Probing and inducing combinational creativity in vision-language models. *arXiv preprint arXiv:2504.13120*.

https://openai.com/index/dall-e/

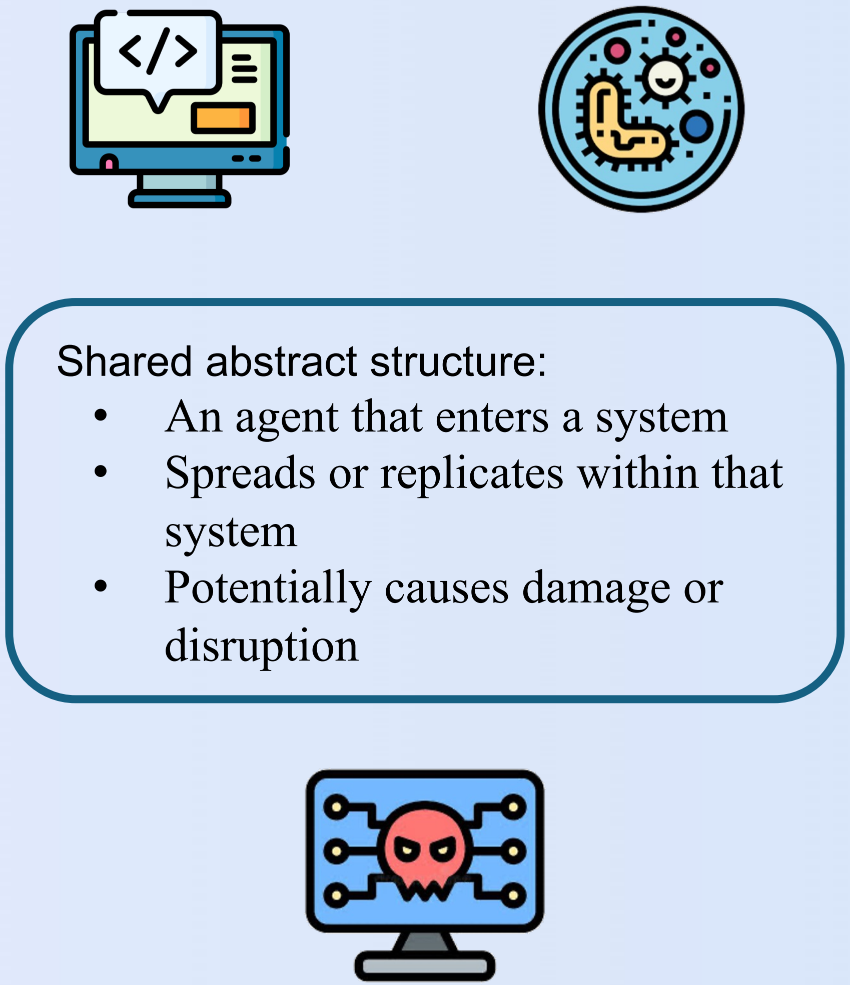
- To what extent can VLMs comprehend combinational creativity?
- Can the explicit integration of the combination process enhance models' ability to generate more creative products?

Overview



- We built a three-level framework of combinational creativity and a novel dataset containing mashup images for comprehensive analysis of how VLMs understand combinational creativity.
- We study whether explicitly incorporating this three-level framework into the generation process can enhance text-to-image models’ ability to generate creative mashup images.

Three Levels of Combinational Creativity



Shared abstract structure:

- An agent that enters a system
- Spreads or replicates within that system
- Potentially causes damage or disruption

Identification-level — Input Space

Identify the objects used in the combination from the final product, answering: What objects are used for a combination?

Explanation-level — Generic Space

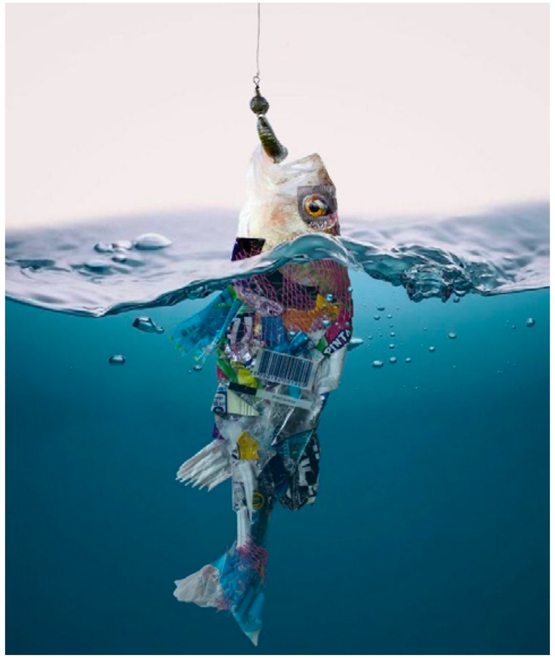
Explain the principles behind the combination, delving into relationships between objects, and answering: How does the combination work?

Implication-level — Blended Space

Examine the underlying meaning behind the combinational creativity product, answering: What is the meaning of the combination?

CreativeMashup Dataset

(a) Understanding Task



Identification

? Identify the primary objects in the image.

🤖 Fish
Garbage

Explanation

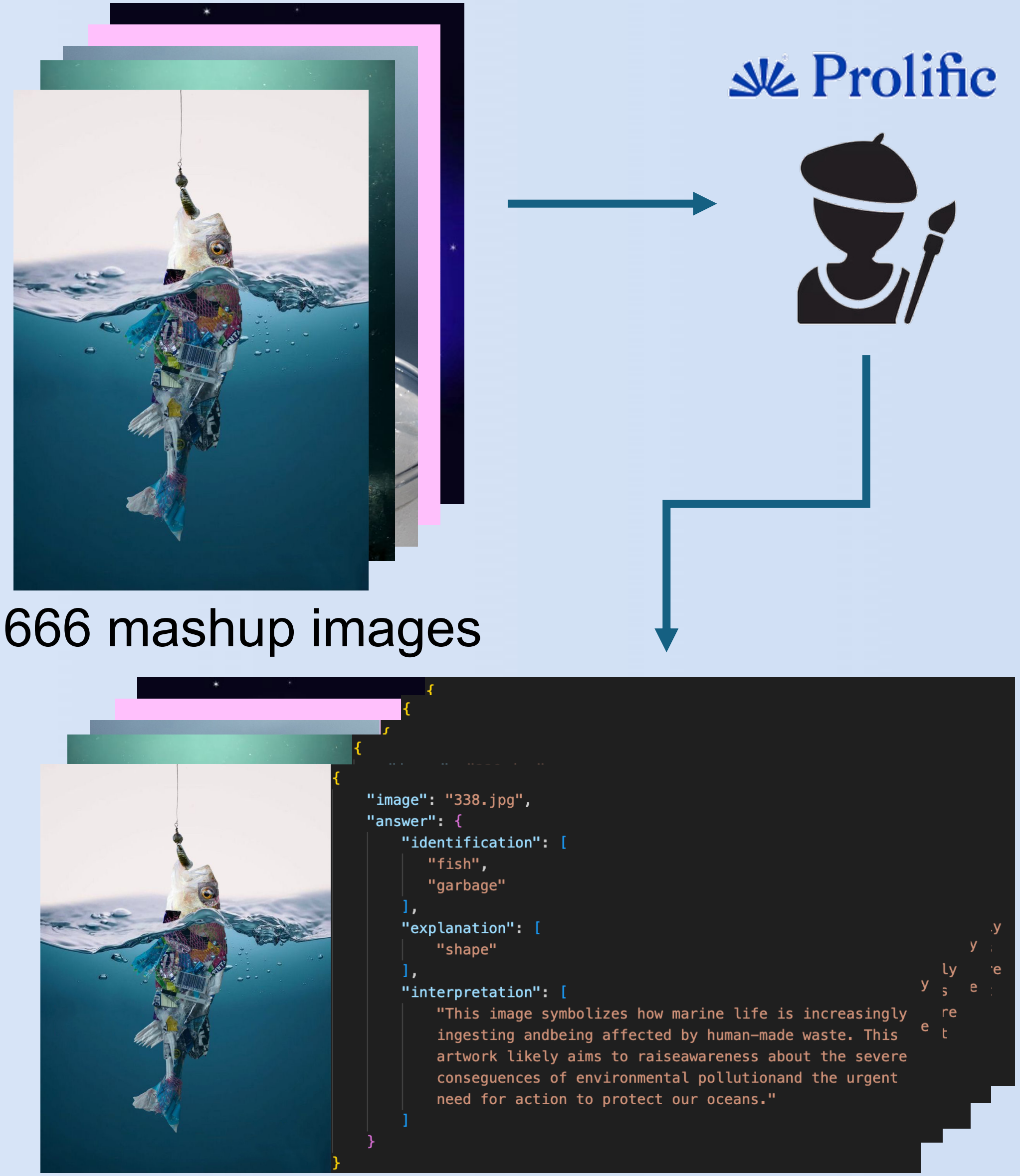
? Choose the relevant attributes that make the combination feasible.

🤖 Shape

Implication

? Interpret what the combination of objects might be intended to convey.

🤖 This image symbolizes how marine life is increasingly ingesting and being affected by human-made waste. This artwork likely aims to raise awareness about the severe consequences of environmental pollution and the urgent need for action to protect our oceans.

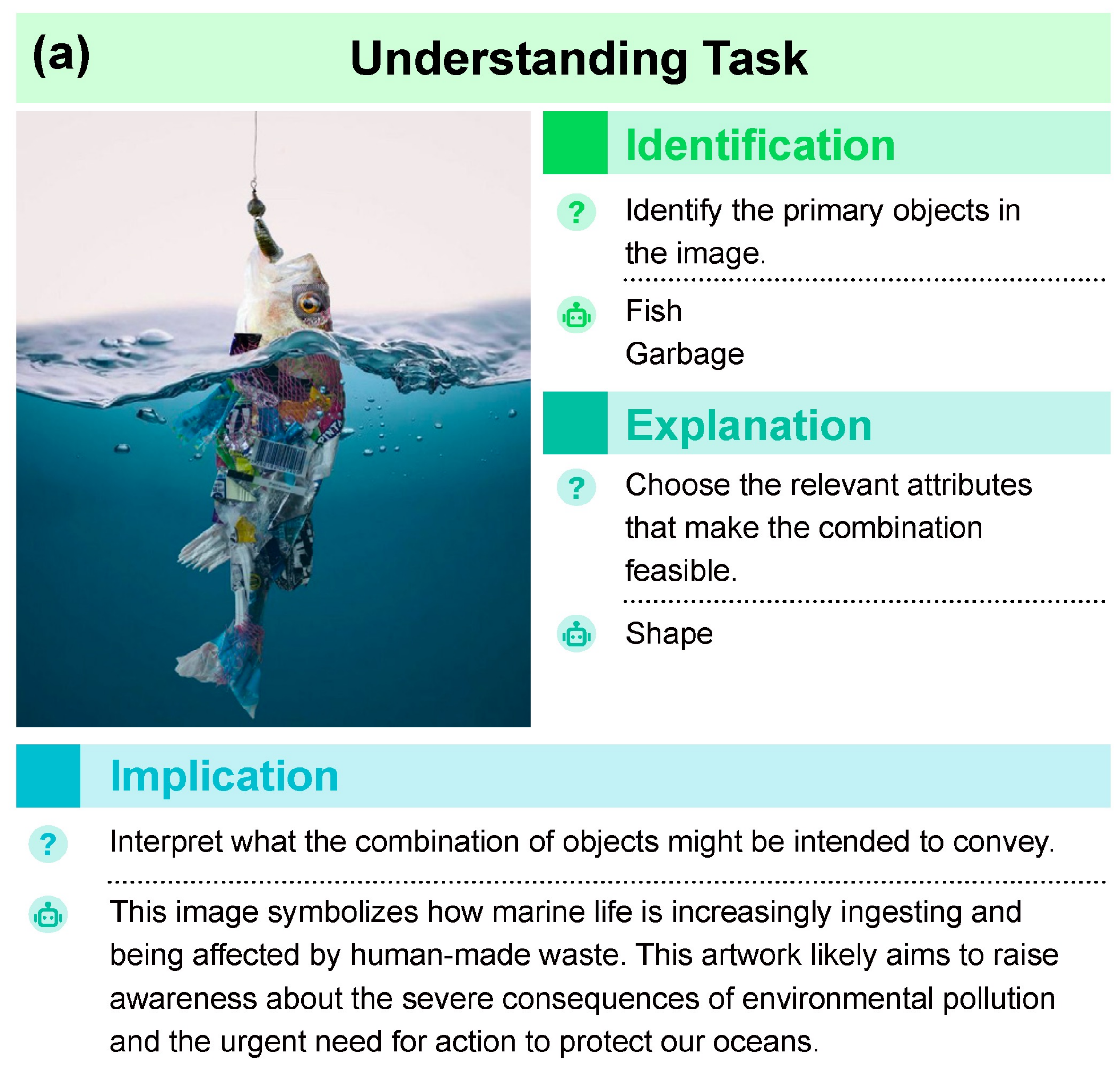


666 mashup images

```
{
    "image": "338.jpg",
    "answer": {
        "identification": [
            "fish",
            "garbage"
        ],
        "explanation": [
            "shape"
        ],
        "interpretation": [
            "This image symbolizes how marine life is increasingly
            ingesting andbeing affected by human-made waste. This
            artwork likely aims to raiseawareness about the severe
            consequences of environmental pollutionand the urgent
            need for action to protect our oceans."
        ]
    }
}
```

Annotated data

Understanding Task



(a) Understanding Task

Identification

? Identify the primary objects in the image.

🤖 Fish
Garbage

Explanation

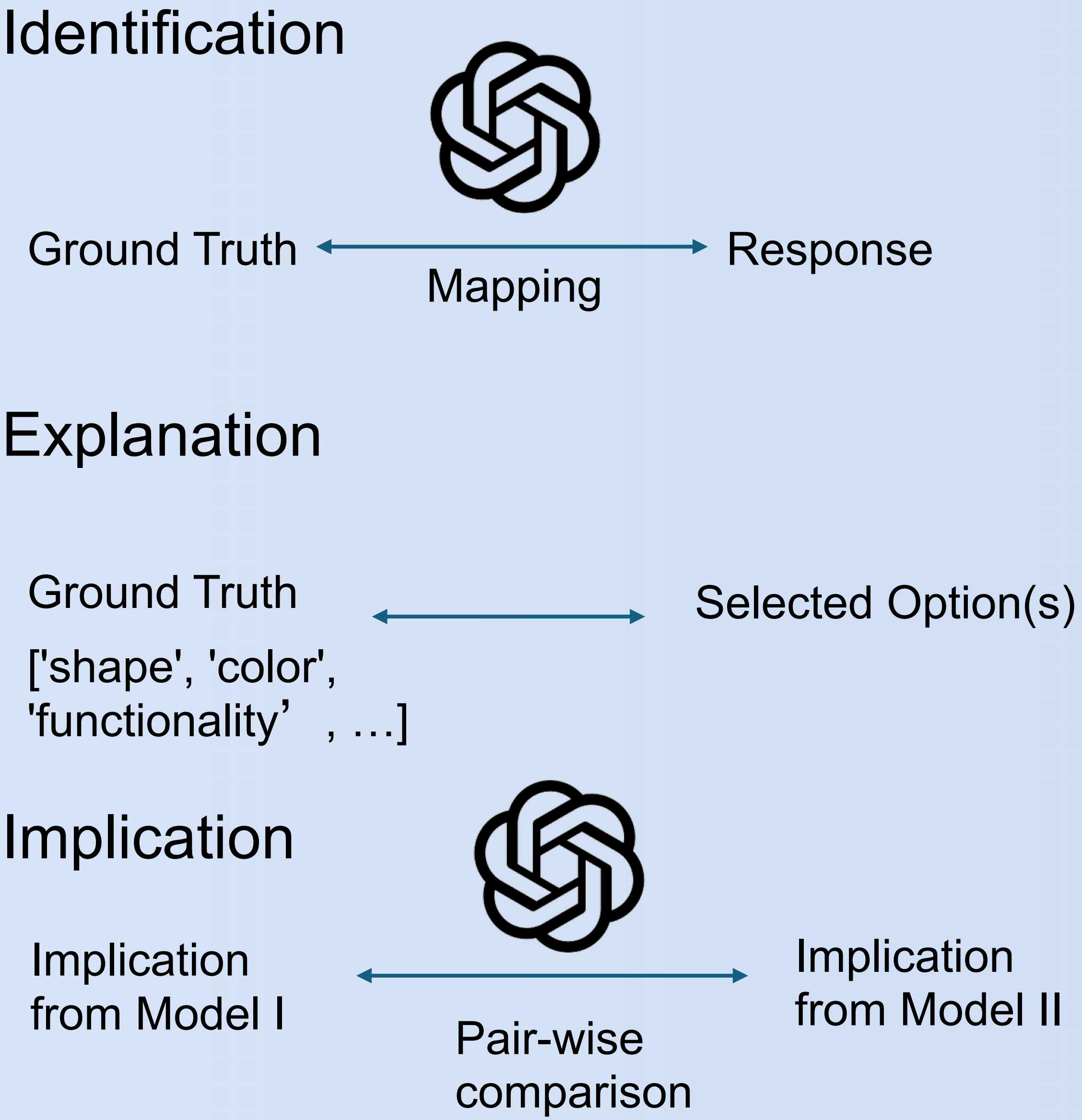
? Choose the relevant attributes that make the combination feasible.

🤖 Shape

Implication

? Interpret what the combination of objects might be intended to convey.

🤖 This image symbolizes how marine life is increasingly ingesting and being affected by human-made waste. This artwork likely aims to raise awareness about the severe consequences of environmental pollution and the urgent need for action to protect our oceans.



Identification

Ground Truth ↔ Response

Mapping

Explanation

Ground Truth ↔ Selected Option(s)

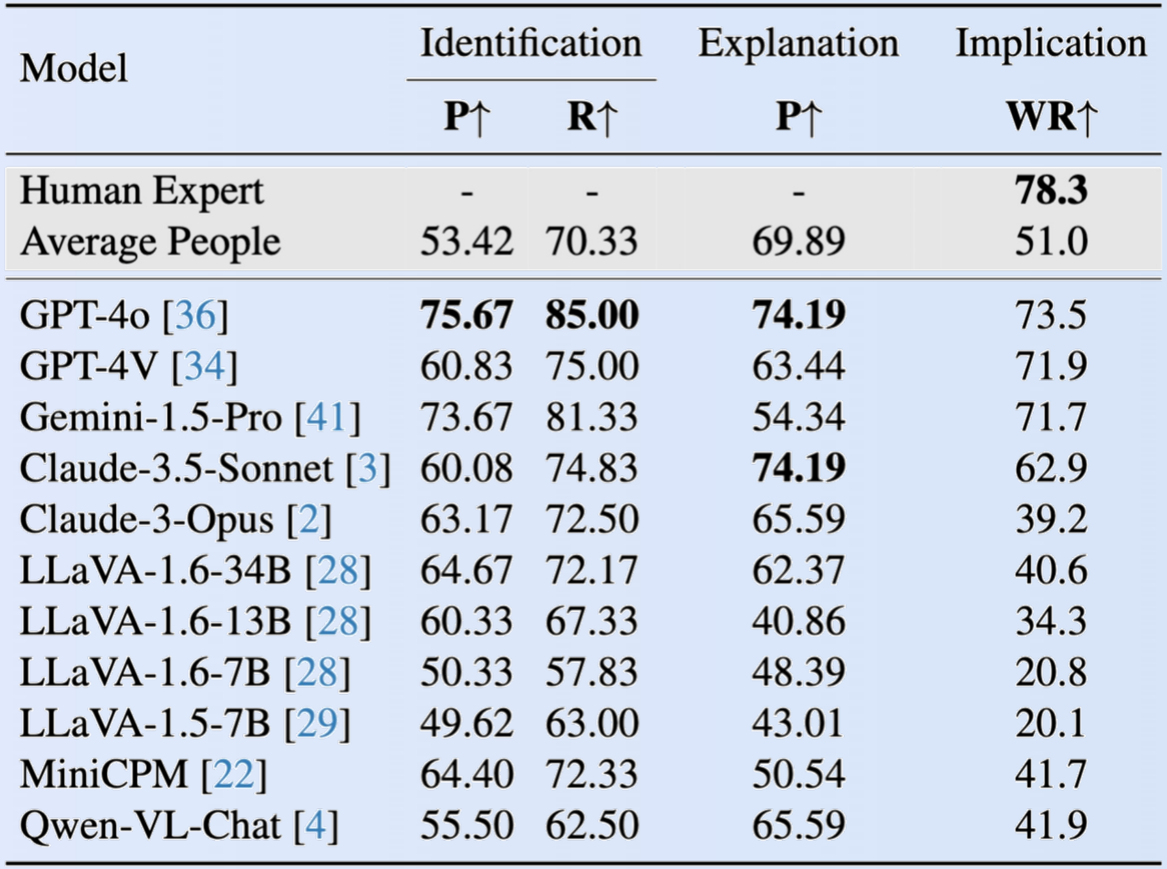
['shape', 'color', 'functionality’ , …]

Implication

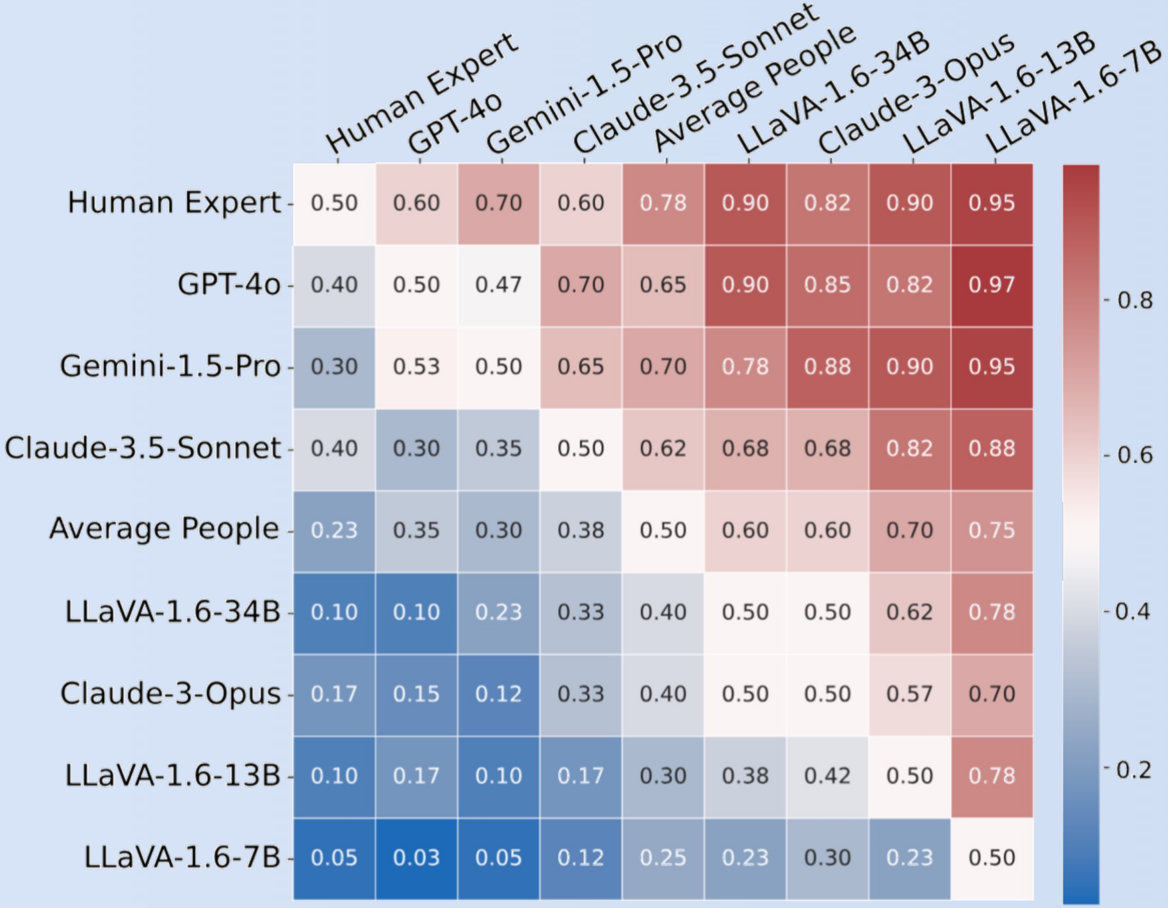
Implication from Model I ↔ Implication from Model II

Pair-wise comparison

Do VLMs Understand Combinational Creativity?



Model	Identification		Explanation	Implication
	P↑	R↑	P↑	WR↑
Human Expert	-	-	-	**78.3**
Average People	53.42	70.33	69.89	51.0
GPT-4o [36]	**75.67**	**85.00**	**74.19**	73.5
GPT-4V [34]	60.83	75.00	63.44	71.9
Gemini-1.5-Pro [41]	73.67	81.33	54.34	71.7
Claude-3.5-Sonnet [3]	60.08	74.83	**74.19**	62.9
Claude-3-Opus [2]	63.17	72.50	65.59	39.2
LLaVA-1.6-34B [28]	64.67	72.17	62.37	40.6
LLaVA-1.6-13B [28]	60.33	67.33	40.86	34.3
LLaVA-1.6-7B [28]	50.33	57.83	48.39	20.8
LLaVA-1.5-7B [29]	49.62	63.00	43.01	20.1
MiniCPM [22]	64.40	72.33	50.54	41.7
Qwen-VL-Chat [4]	55.50	62.50	65.59	41.9

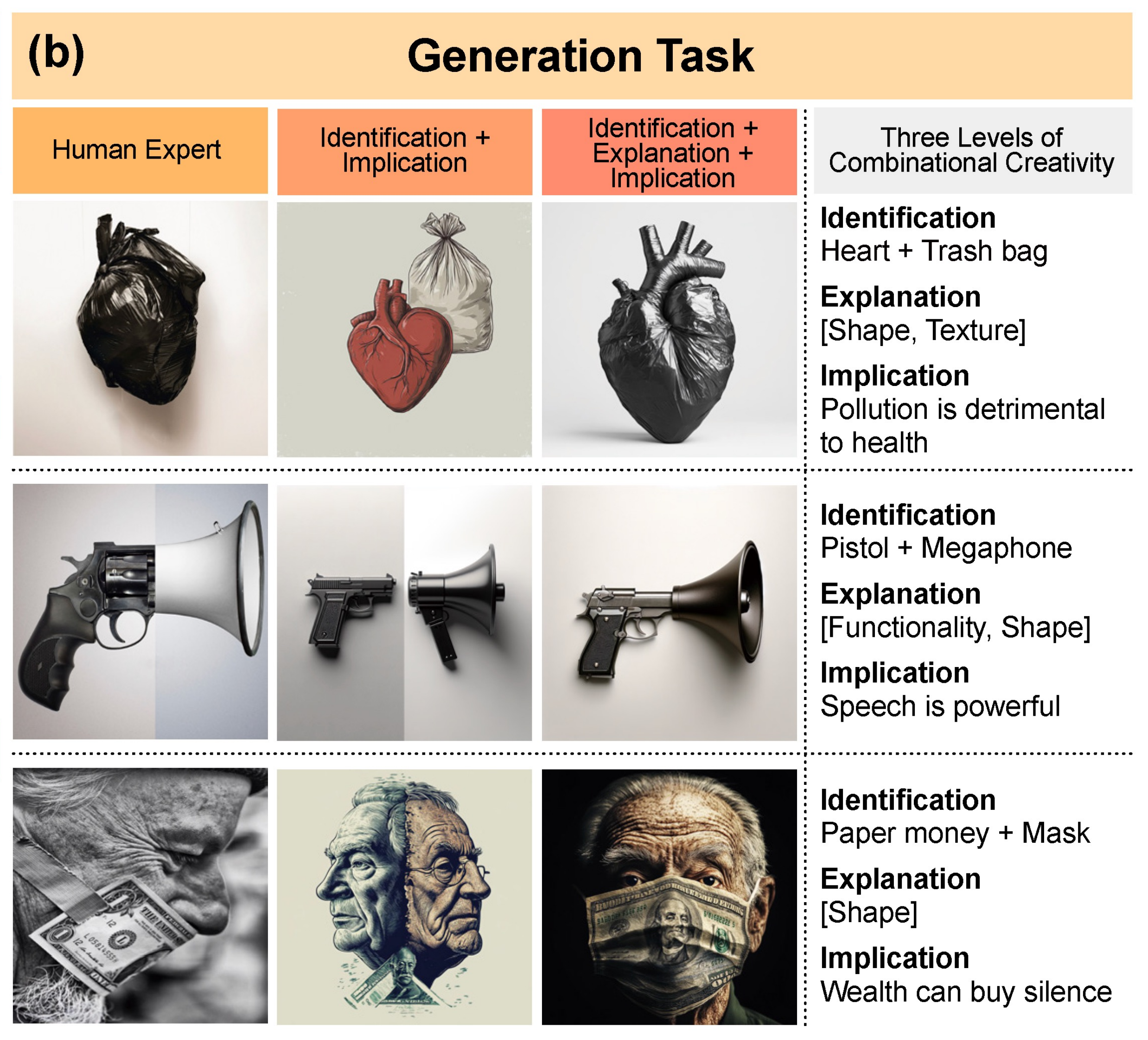


	Human Expert	GPT-4o	Gemini-1.5-Pro	Claude-3.5-Sonnet	Average People	LLaVA-1.6-34B	Claude-3-Opus	LLaVA-1.6-13B	LLaVA-1.6-7B
Human Expert	0.50	0.60	0.70	0.60	0.78	0.90	0.82	0.90	0.95
GPT-4o	0.40	0.50	0.47	0.70	0.65	0.90	0.85	0.82	0.97
Gemini-1.5-Pro	0.30	0.53	0.50	0.65	0.70	0.78	0.88	0.90	0.95
Claude-3.5-Sonnet	0.40	0.30	0.35	0.50	0.62	0.68	0.68	0.82	0.88
Average People	0.23	0.35	0.30	0.38	0.50	0.60	0.60	0.70	0.75
LLaVA-1.6-34B	0.10	0.10	0.23	0.33	0.40	0.50	0.50	0.62	0.78
Claude-3-Opus	0.17	0.15	0.12	0.33	0.40	0.50	0.50	0.57	0.70
LLaVA-1.6-13B	0.10	0.17	0.10	0.17	0.30	0.38	0.42	0.50	0.78
LLaVA-1.6-7B	0.05	0.03	0.05	0.12	0.25	0.23	0.30	0.23	0.50

State-of-the-art models have achieved human-level understanding in combinational creativity.

Human experts still surpass models in the realm of combinational creativity.

Generation Task



(b) Generation Task

Human Expert

Identification + Implication

Identification + Explanation + Implication

Three Levels of Combinational Creativity

Identification
Heart + Trash bag

Explanation
[Shape, Texture]

Implication
Pollution is detrimental to health

Identification
Pistol + Megaphone

Explanation
[Functionality, Shape]

Implication
Speech is powerful

Identification
Paper money + Mask

Explanation
[Shape]

Implication
Wealth can buy silence

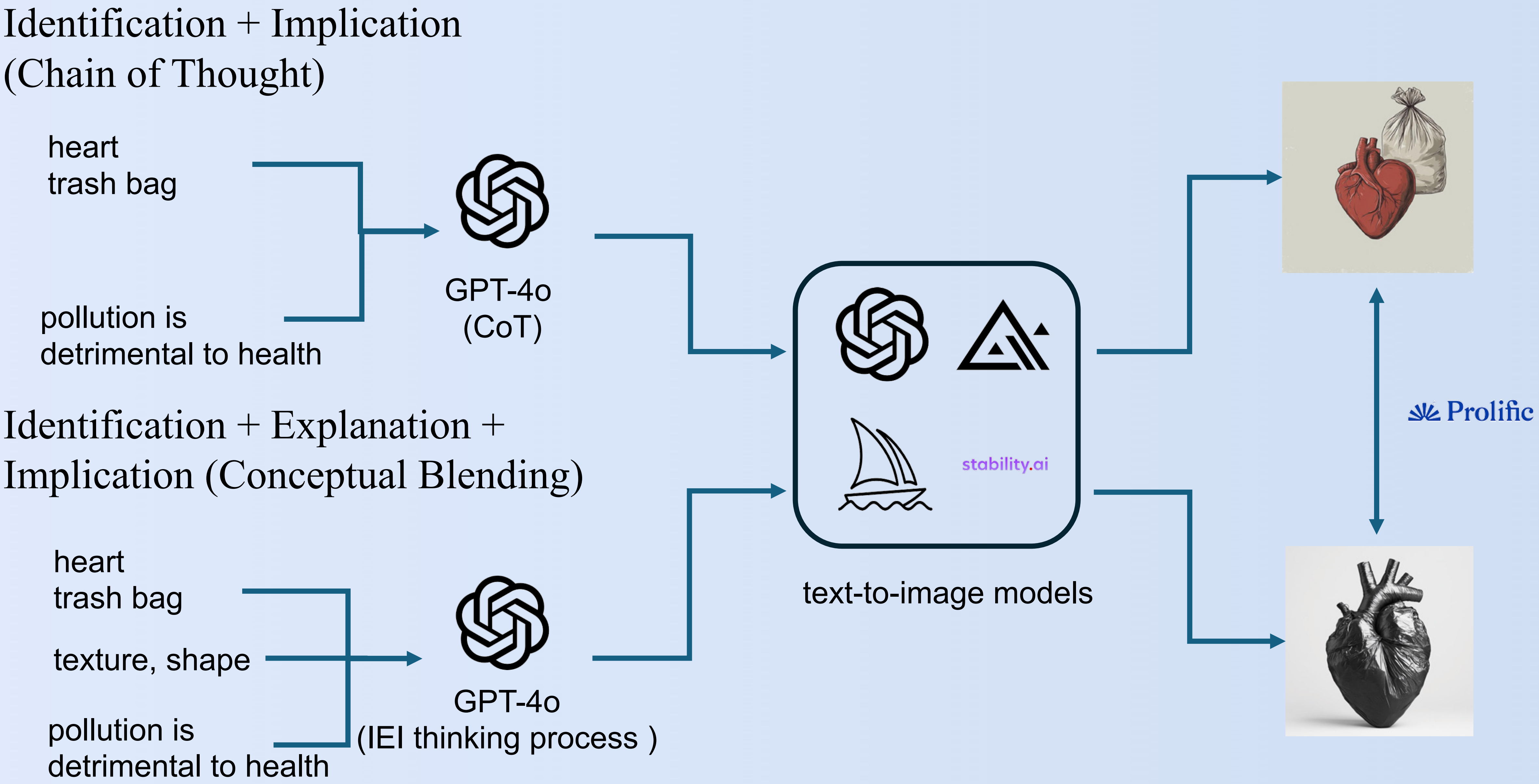
RQ: Can the explicit integration of the combination process enhance models' ability to generate more creative products?

Human Expert

Identification + Implication (Chain of Thought)

Identification + Explanation + Implication (Conceptual Blending)

Generation Task



Identification + Implication (Chain of Thought)

heart
trash bag

pollution is detrimental to health

GPT-4o (CoT)

Identification + Explanation + Implication (Conceptual Blending)

heart
trash bag

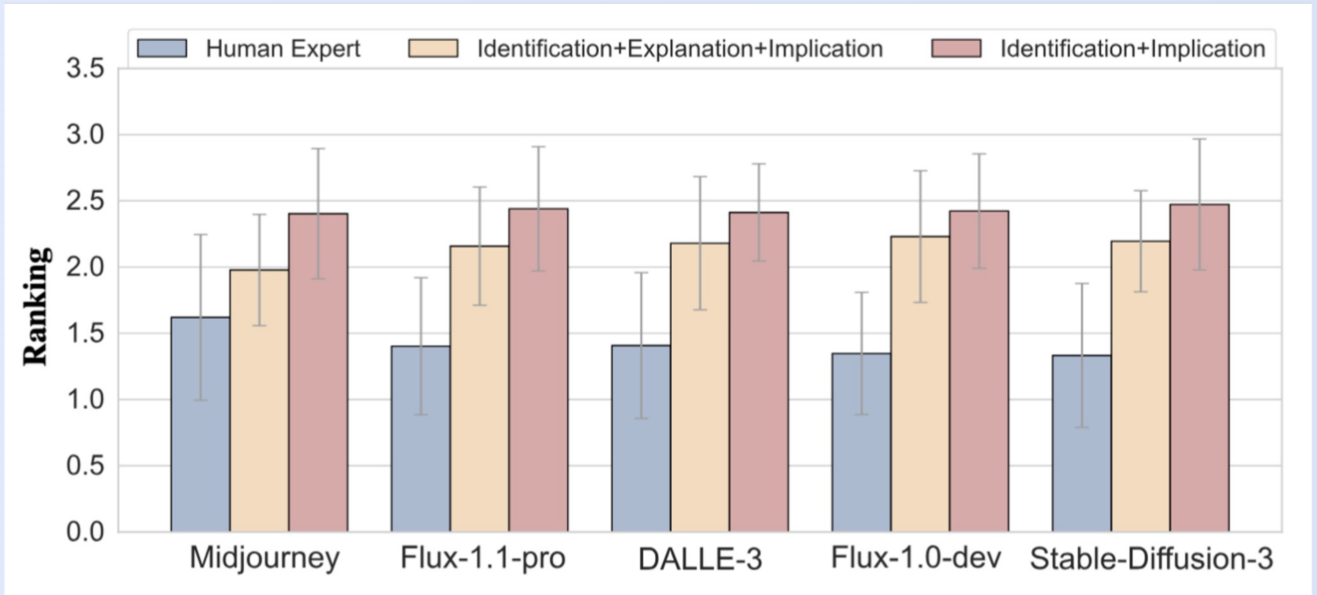
texture, shape

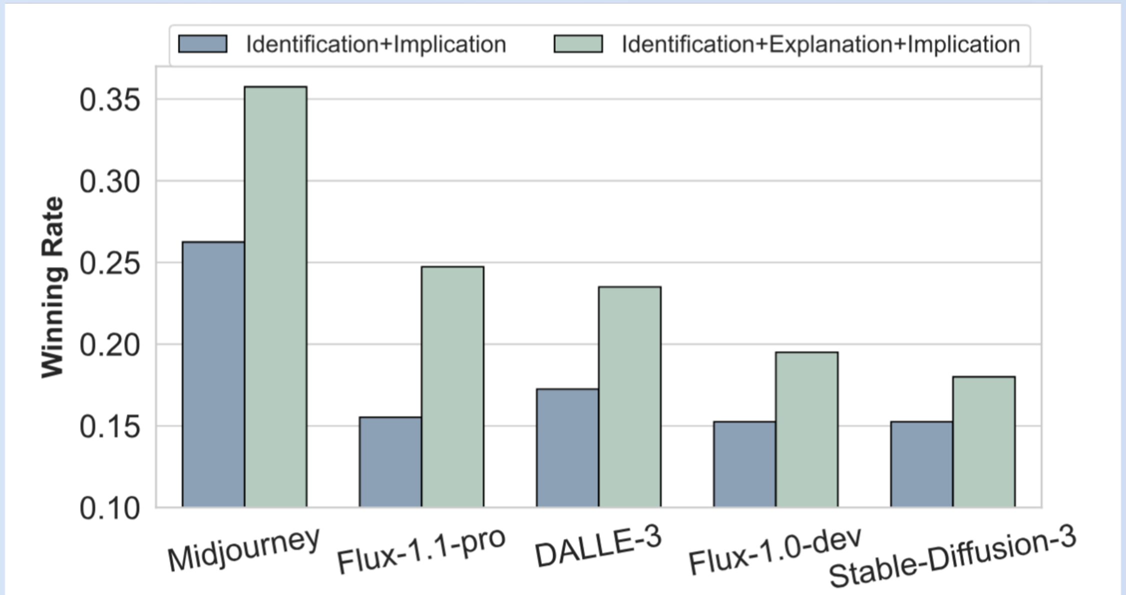
pollution is detrimental to health

GPT-4o (IEI thinking process)

text-to-image models

Results of Generation Experiment





The explanation level of combinational creativity can be leveraged to enhance creativity without making the prompts significantly longer.

II (M = 487.95)
IEI (M = 514.53)
T-test stat: 0.84 p-value: 0.40

Text-to-image models are currently the bottleneck in generating visual combinational creativity.

Text: 90% (36/40)
Text-to-image: 27.8% (10/36)

Take-away Message

- State-of-the-art models **have achieved human-level understanding** in combinational creativity, but still lag behind human experts.

- The explanation level of combinational creativity can be leveraged to **enhance creativity without making the prompts significantly longer**.

- **Text-to-image models** are currently the bottleneck in generating visual combinational creativity.

References

Holyoak, K. J., & Thagard, P. (1996). *Mental leaps: Analogy in creative thought*. MIT press.

Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial intelligence*, *103*(1-2), 347-356.

Fauconnier, G., & Turner, M. (2003). Conceptual blending, form and meaning. *Recherches en communication*, *19*, 57-86.

Boden, M. (2009). Creativity: How does it work. *The idea of creativity*, *28*, 237-50.

Kaufman, J. C., & Sternberg, R. J. (Eds.). (2010). *The Cambridge handbook of creativity*. Cambridge University Press.

Park, S., Nie, B. X., & Zhu, S. C. (2017). Attribute and-or grammar for joint parsing of human pose, parts and attributes. *IEEE transactions on pattern analysis and machine intelligence*, *40*(7), 1555-1569.

Richardson, E., Goldberg, K., Alaluf, Y., & Cohen-Or, D. (2024). Conceptlab: Creative concept generation using vlm-guided diffusion prior constraints. *ACM Transactions on Graphics*, *43*(3), 1-14.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, 35, 24824-24837.