

Generative AI & Creative Problem Solving

Léonard Boussioux

Joint work with Jacqueline N. Lane*, Miaomiao Zhang, Vladimir Jacimovic, Karim Lakhani

* equal authorship



Harvard
Business
School



DIGITAL
DATA
DESIGN
INSTITUTE
AT HARVARD

FOSTER
SCHOOL OF **BUSINESS**
UNIVERSITY OF WASHINGTON



AI and Innovation
 **Wharton**
UNIVERSITY OF PENNSYLVANIA

How to solve the United Nations' Sustainable Development

GOALS

in 5 minutes

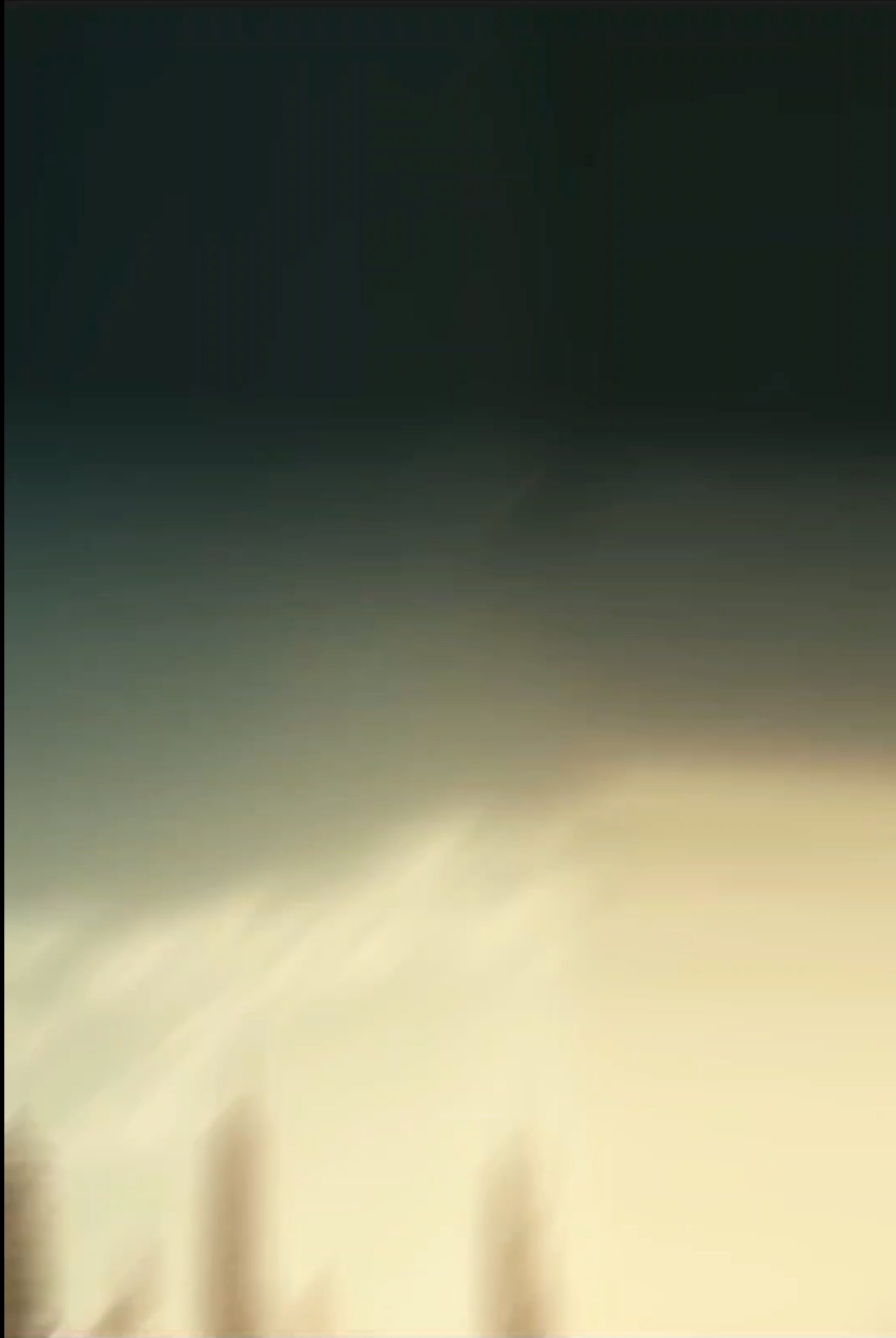


GOALS



**11 SUSTAINABLE CITIES
AND COMMUNITIES**





**11 SUSTAINABLE CITIES
AND COMMUNITIES**



Research Questions

- How can generative AI augment creative problem solving?
- What are the implications of generative AI for human crowdsourcing of innovations?



Setting

- Partnered with Continuum Lab, an AI company, and Freelancer.com to develop a crowdsourcing challenge about new business ideas on circular economy.
- Both human crowd solvers and human-AI (using GPT-4) submitted solutions.
- Human judges rated all solutions for their novelty and value.



Summary of Findings



Creating good business ideas
in **circular economy** in the
format of **Problem-Solution**.

Human crowd:
Open call to
human solvers



Extreme, highly novel
solutions



Novel solutions

**Human crowd
intelligence**

= Creativity*
(novelty x value)

Human-AI:
Prompt GPT-4

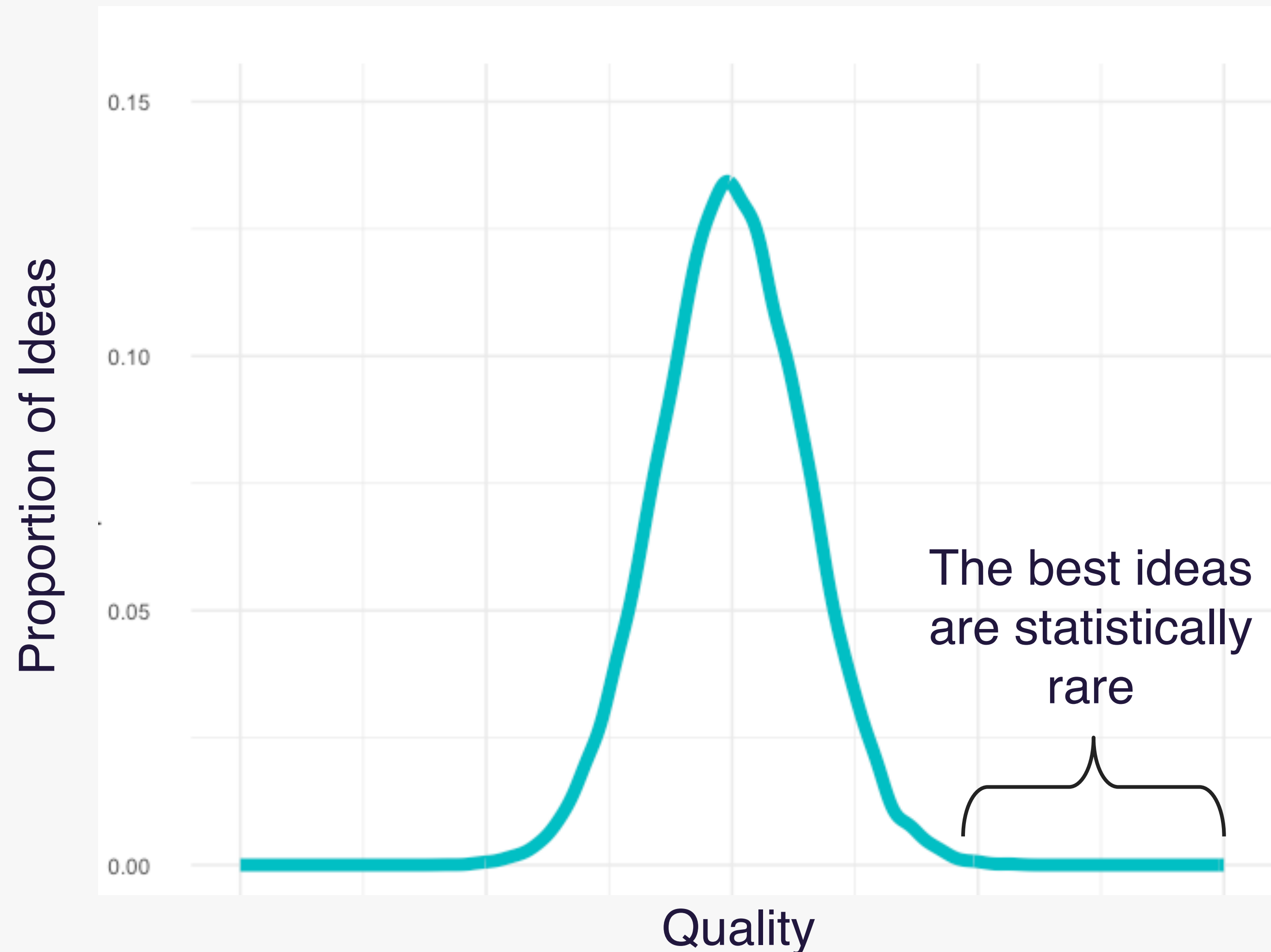


Valuable solutions

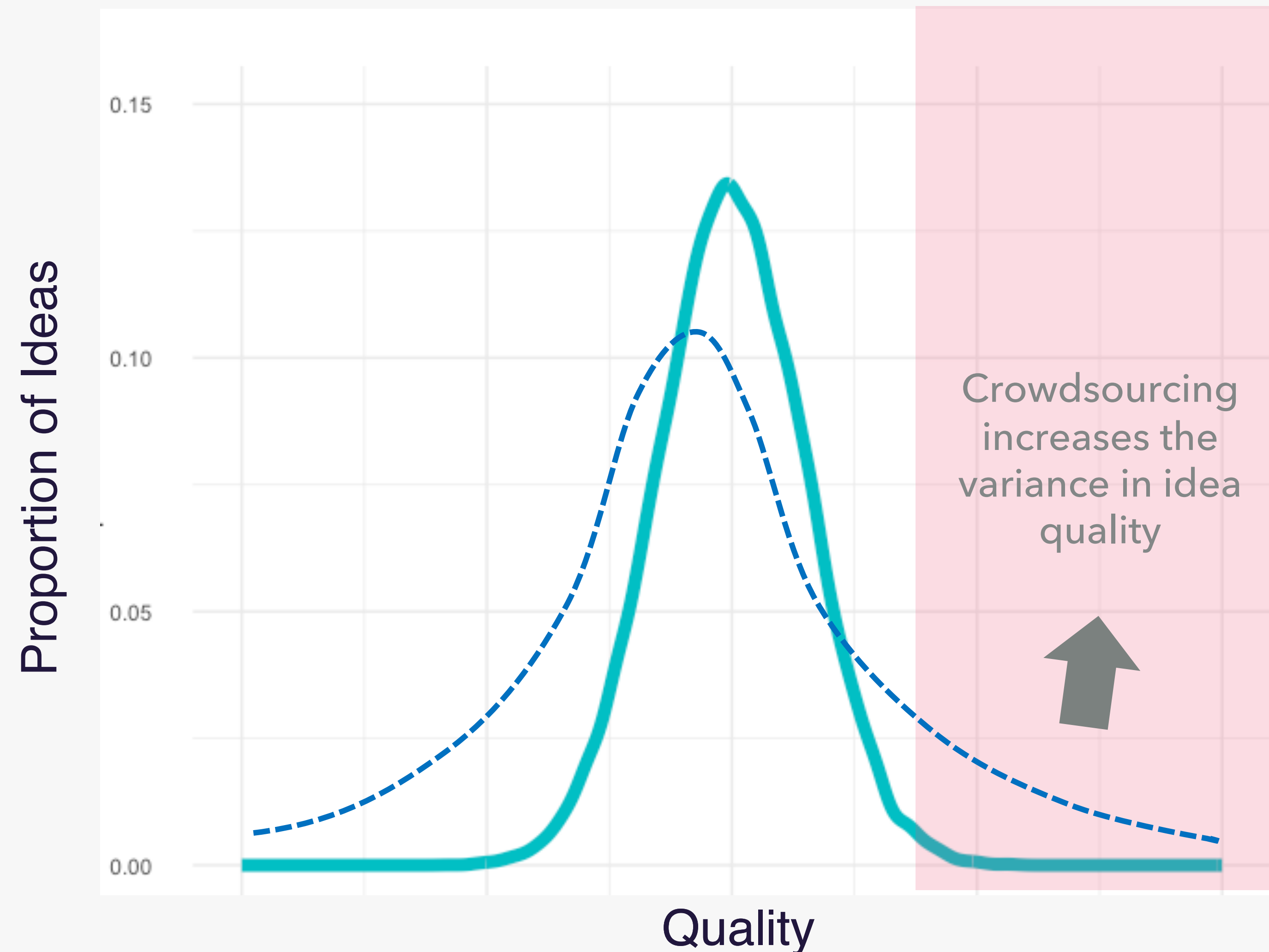
**Human-AI
intelligence**

*Poetz and Schreier, 2012

Motivation: A Statistical View of Innovation



Crowdsourcing can affect idea quality



Benefits of crowdsourcing



Access to Diverse Knowledge & Perspectives

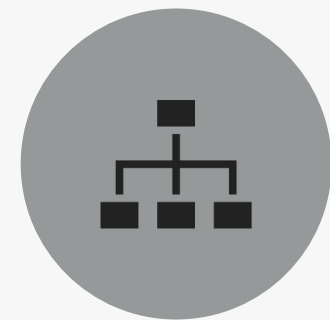


Higher Likelihood of Capturing 'Extreme Outcomes'



Cost Effective and Efficient vs. In-house

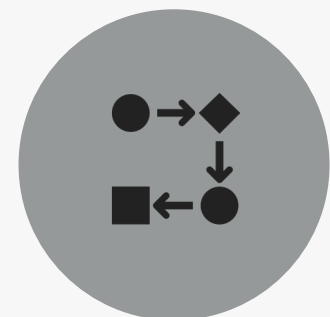
Challenges of Crowdsourcing



Complexity in Organization



Problem Formulation and Decomposition



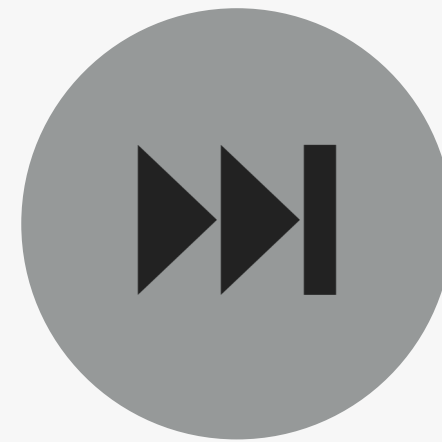
Filtering out low-quality ideas



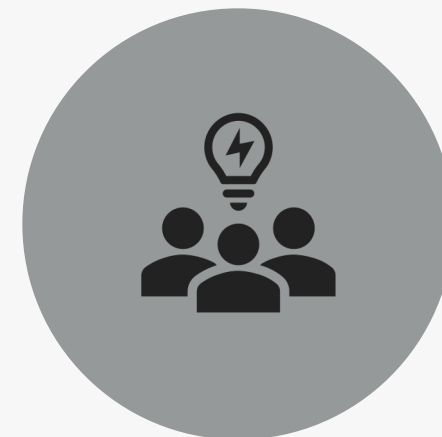
Incentives for Participation



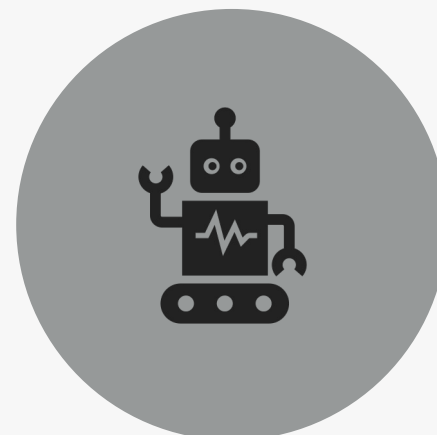
The Potential of Large Language Models (LLMs) for Idea Generation



Scalability: LLMs can produce many ideas fast, efficiently, and cost-effectively.

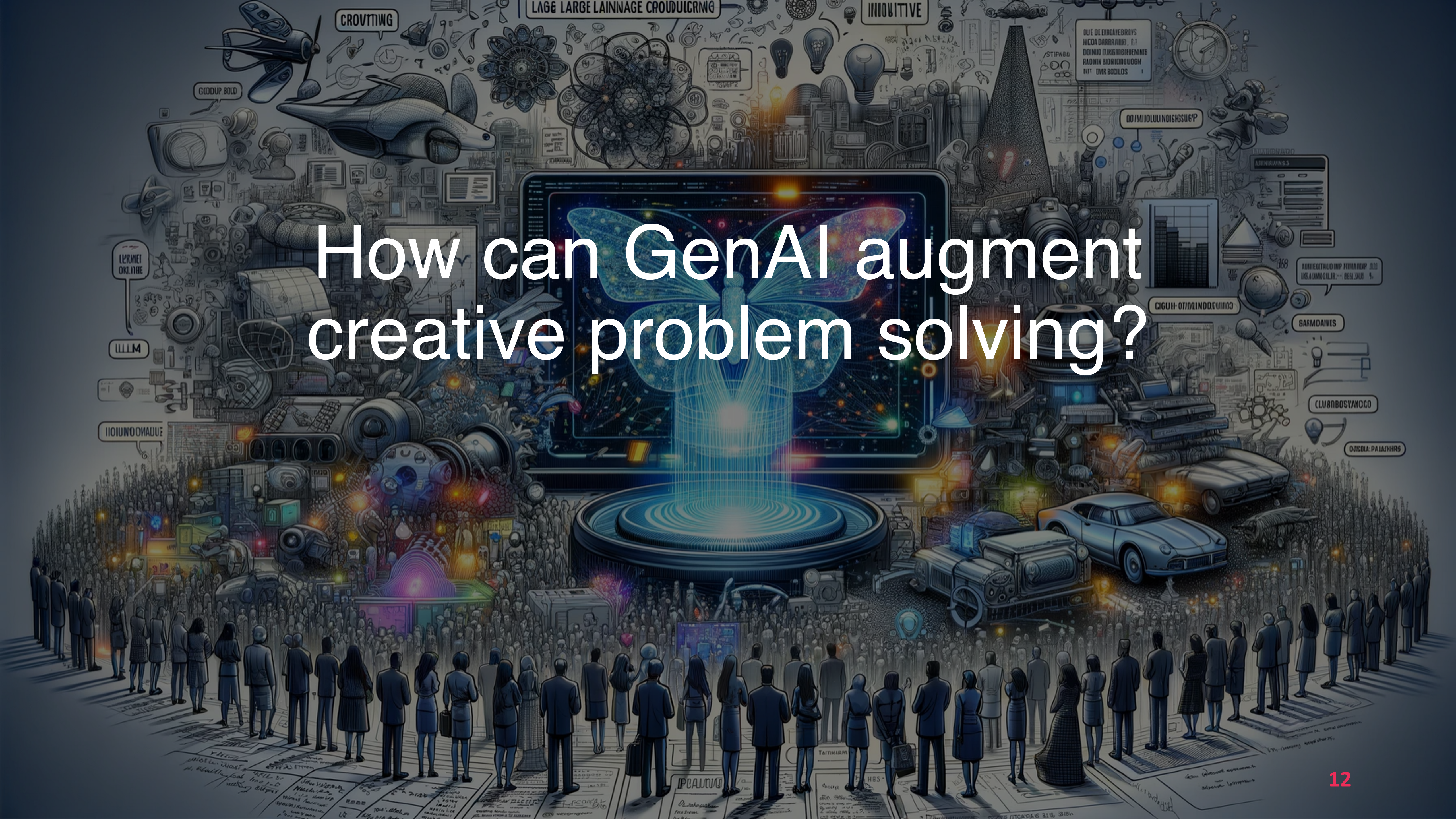


Diversity: LLMs are trained on a vast amount of data and can generate various potential solutions or concepts via novel “recombinations”.



Augmentation: LLMs, with strategic prompt engineering, can improve human productivity and creativity for well-defined organizational tasks.

How can GenAI augment creative problem solving?



Setting of the study



Crowdsourcing Process

- **Humans:**

- Launched crowdsourcing challenge on Freelancer.com from Jan – May 2023.
- **Received 125 eligible human solutions (after filtering).**

- **LLMs:**

- Used **GPT-4 to generate solutions using strategic prompt engineering approaches** (few-shot prompting, Chain-of-Thought, role-play, prompt chaining).

Evaluation Process

On Prolific: Out of 1000 individuals screened, 300 passed.

Each Problem-Solution pair evaluated 16 times on average.



How It Works

Browse Jobs

Find Jobs

Hire Freelancers

Get Ideas

About

Resources

Freelancer > Contest > Academic Writing > Join Harvard in Creating a Sustainable Future: Unlock

Join Harvard in Creating a Sustainable Future: Unlock the Potential of a Circular Economy - 30/01/2023 18:12 EST

Status: Closed Prize: \$1000 Entries Received: 310 Winner: [rashasafwat](#)

Contest Brief

THE CONTEST IS CLOSED AND WE ARE AIMING TO AWARD LAST SELECTED WINNERS BY 15TH OF MAY

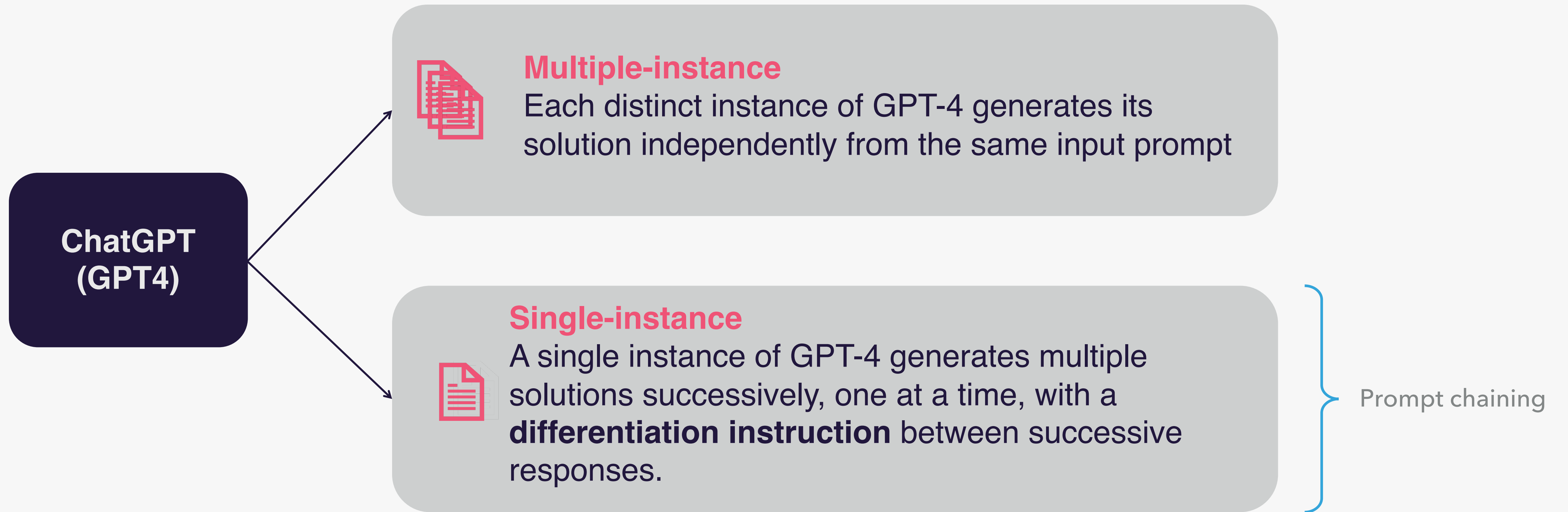
We are excited to announce an opportunity for freelancers to collaborate with researchers at the Digital, Data, and Design Institute (D³) (<https://d3.harvard.edu/>) at Harvard to source the most innovative and cutting-edge circular economy solutions for the business world.

Circular Economy is a simple idea.

Basically it involves sharing, leasing, reusing, repairing, refurbishing and recycling existing materials and products as long as possible.

We would like you to submit your circular economy idea, which can be a unique new idea or an existent idea that is used in the industry.

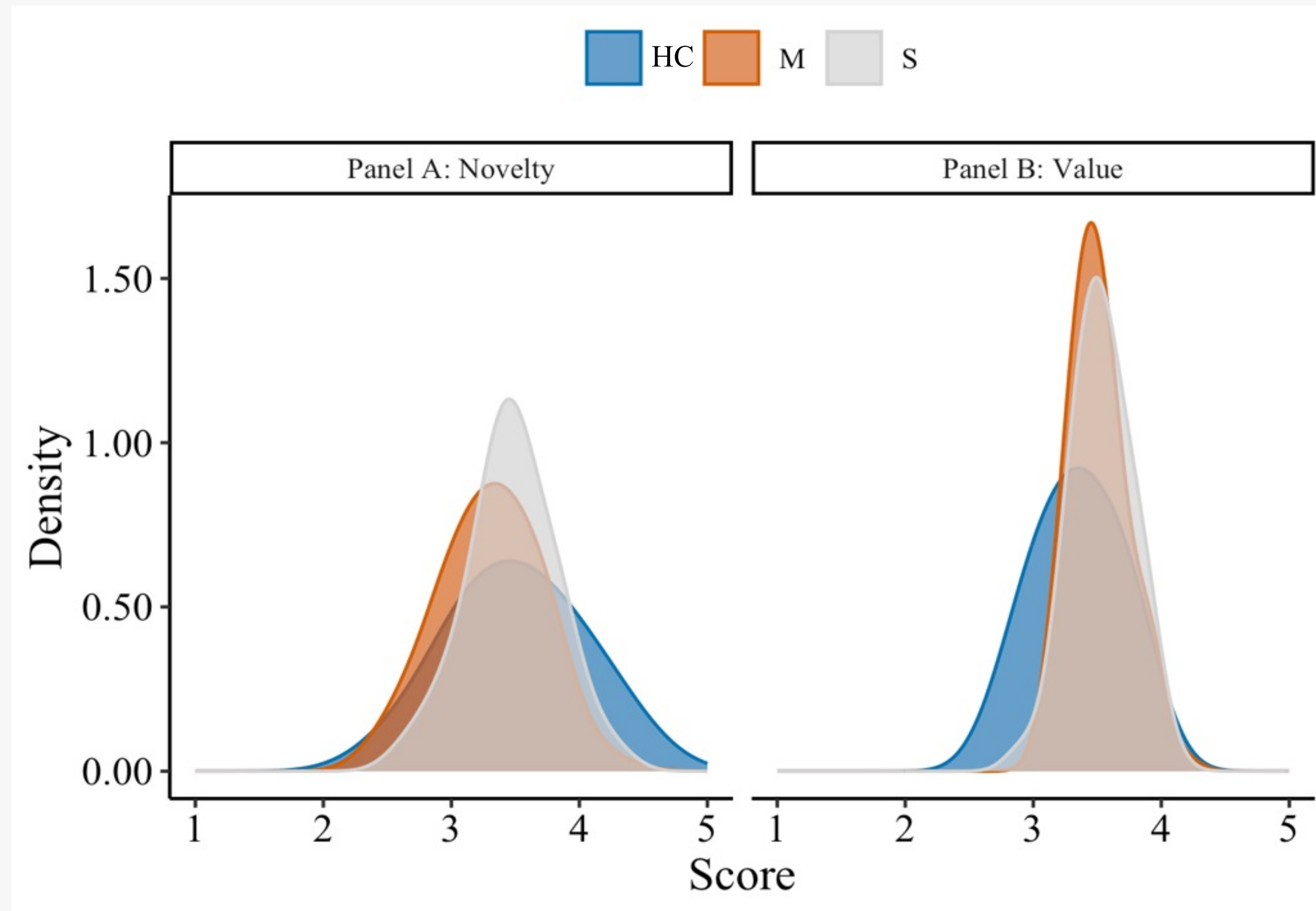
ChatGPT Prompting Instances



Distribution of Mean Novelty and Value Ratings by AI Instance

Figure 2. Density plot distributions of the mean novelty (Panel A) and value (Panel B) evaluator ratings by solution source and prompting configuration

- 1 Human-AI solutions 7% less likely to achieve top novelty rating than Human Crowd solutions. ($p < 0.001$)
- 2 No difference between multiple and single instance Human-AI and Human-Crowd top novelty rating. ($p < 0.001$)
- 3 Human-AI Single instance outperforms the average creativity of Human-Crowd outputs ($p < 0.1$).

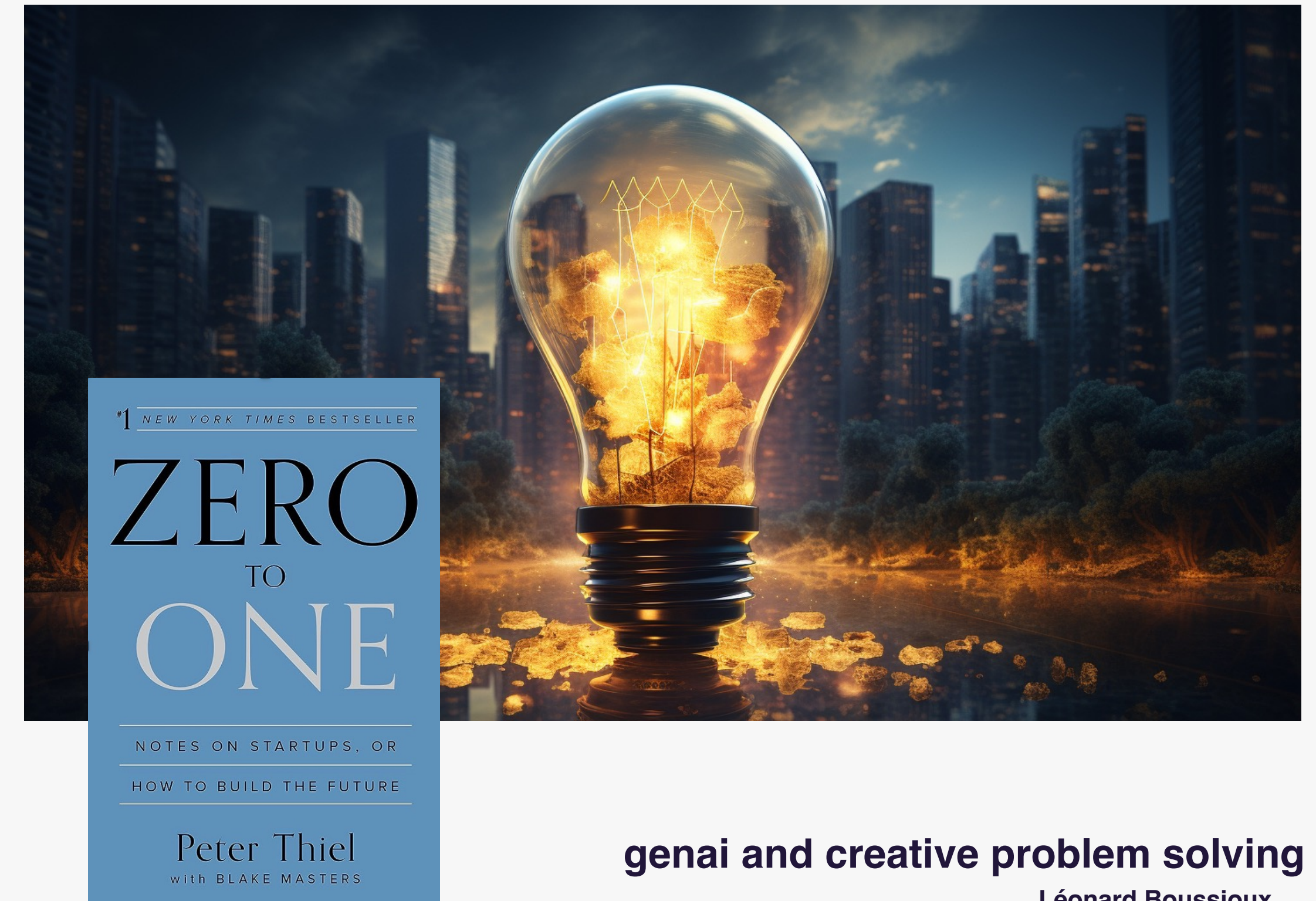
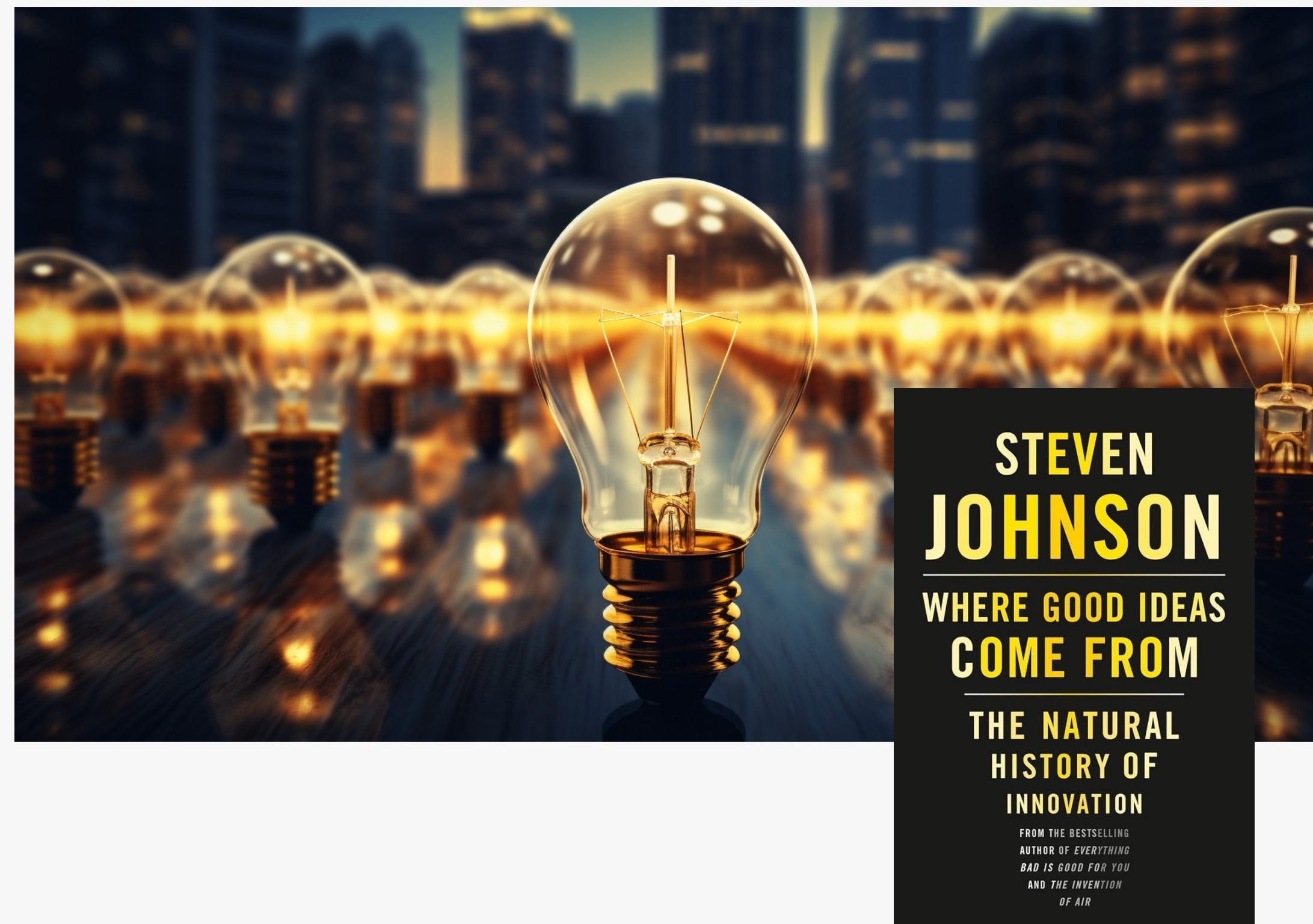


What would it take for AI to generate novel ideas?

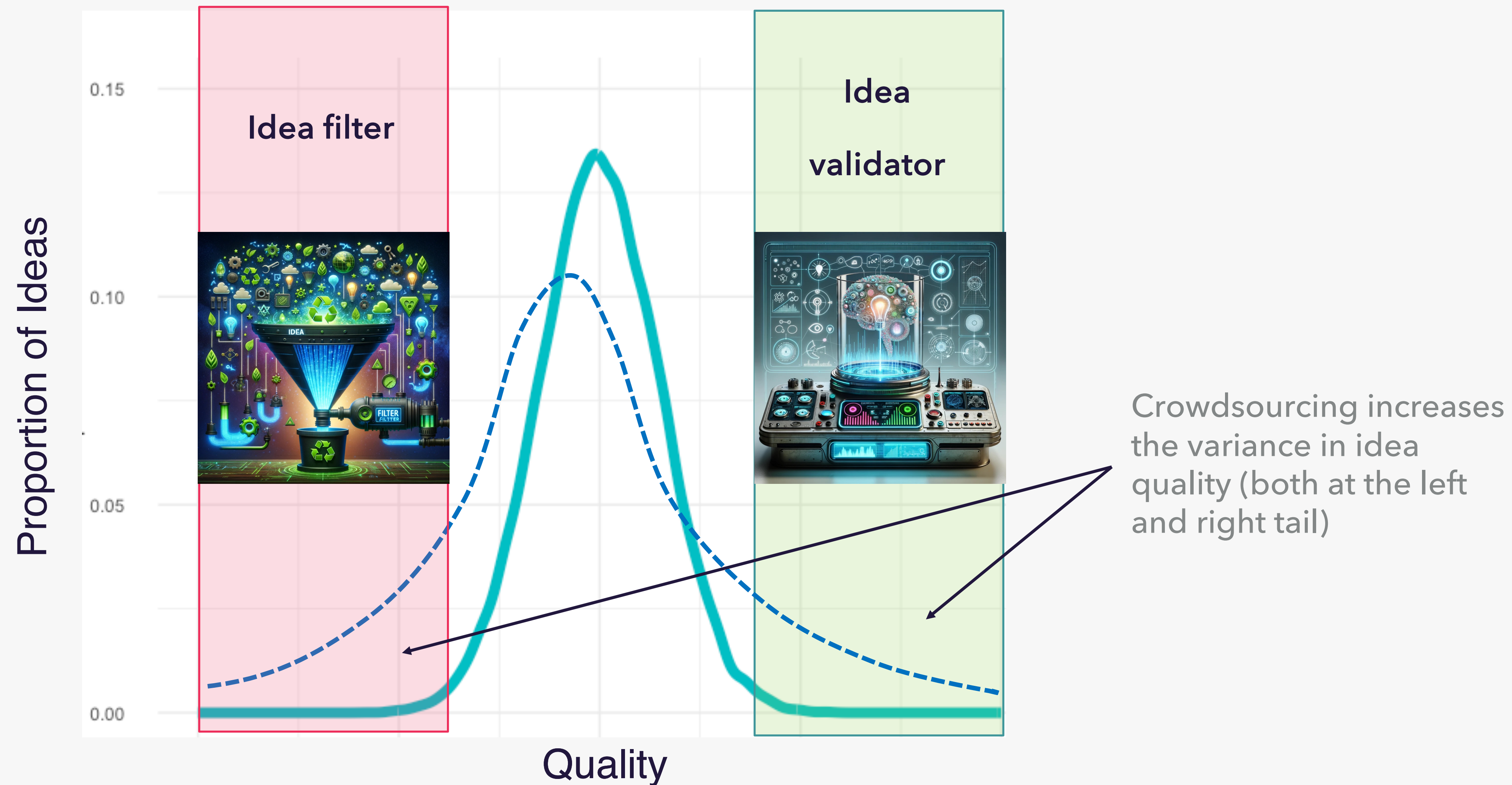
- Human crowd outputs are more innovative, including top novelty, but AI capabilities are advancing (e.g., GPT5, multi-modal & multi-agent LLMs, Retrieval-Augmented Generation (RAG) systems)

Innovation as recombination of ideas for AI?

What about moonshot ideas?



How to select and enhance the best ideas?



Conclusions

- Human-guided AI, with **prompt engineering**, can produce **creative outputs comparable** to human solvers alone.
- Human-AI outputs are **more valuable**, a reflection of LLMs' training, fine-tuning, and alignment.
- Human crowd outputs are **more innovative**, including top novelty, but AI capabilities are advancing.



Conclusions

- Human-guided AI, with **prompt engineering**, can produce **creative outputs comparable** to human solvers alone.
- Human-AI outputs are **more valuable**, a reflection of LLMs' training, fine-tuning, and alignment.
- Human crowd outputs are **more innovative**, including top novelty, but AI capabilities are advancing.



The Future of Creative Problem Solving

- Human-AI guided outputs can be a **cost-effective and scalable** approach to create multiple parallel paths, freeing up human resources for idea evaluation, selection, and implementation.
- Towards a **Human-AI synergistic future**:
 - Intellectual property rights and environmental impact (Rafner et al., 2023)
 - Human purpose vs. excessive dependence on LLMs, homogenization of outputs (e.g., Dell'Acqua et al. 2023; Doshi & Hauser, 2023)
 - Human-AI framework entails integrating AI-in-the-loop responsibly to augment, not replace creative thought.



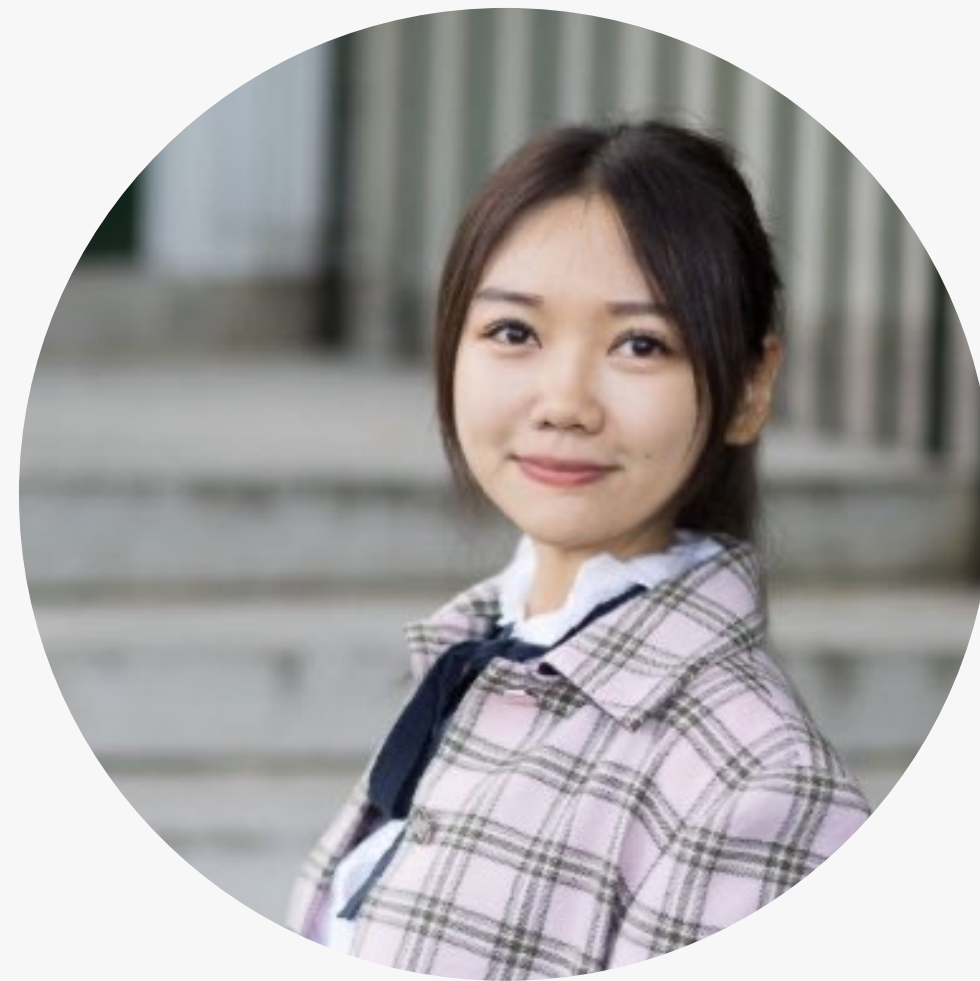
Thank you!

The amazing team

Jackie Lane



Miaomiao Zhang



Karim Lakhani



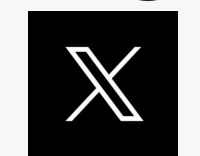
Vladimir Jacimovic



leobix@uw.edu



www.leobix.us



leo_bix

Summary of Findings



Creating good business ideas
in **circular economy** in the
format of **Problem-Solution**.

 leobix@uw.edu
 www.leobix.us
 leo_bix

Human crowd:
Open call to
human solvers



Extreme, highly novel
solutions



Novel solutions

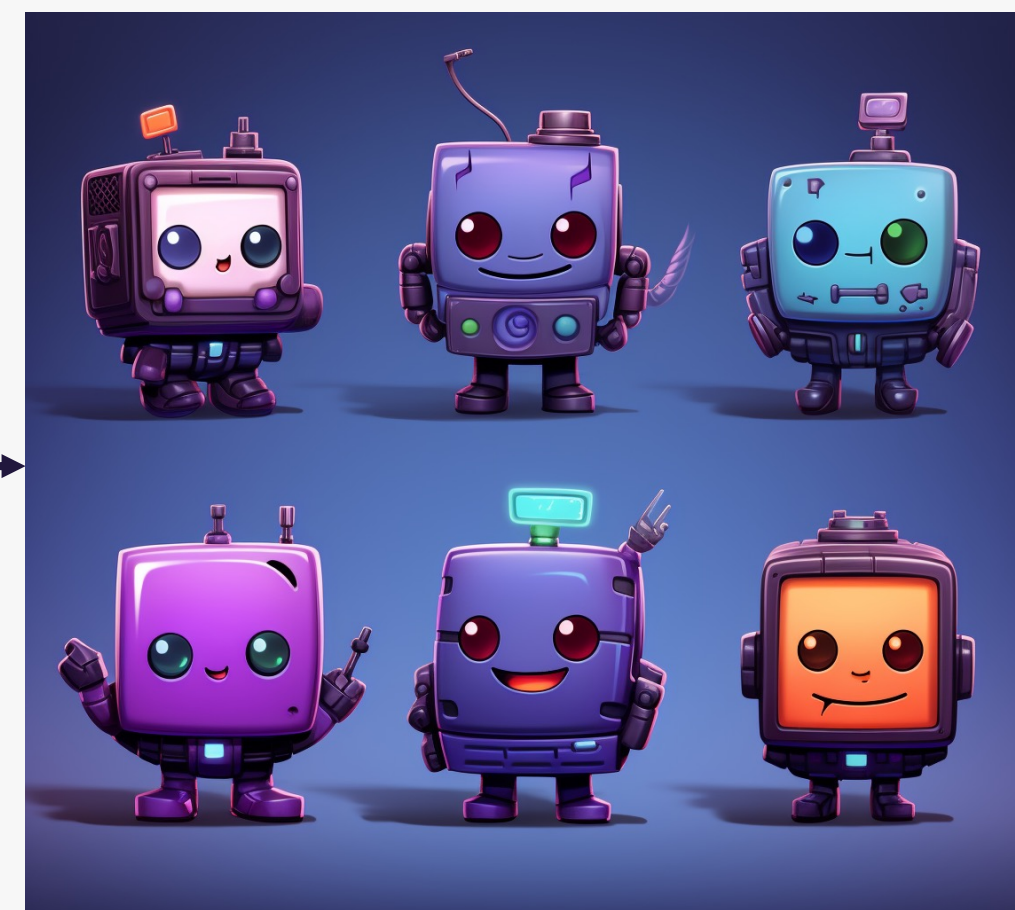
**Human crowd
intelligence**

= Creativity
(novelty x value)

**Human-AI
intelligence**



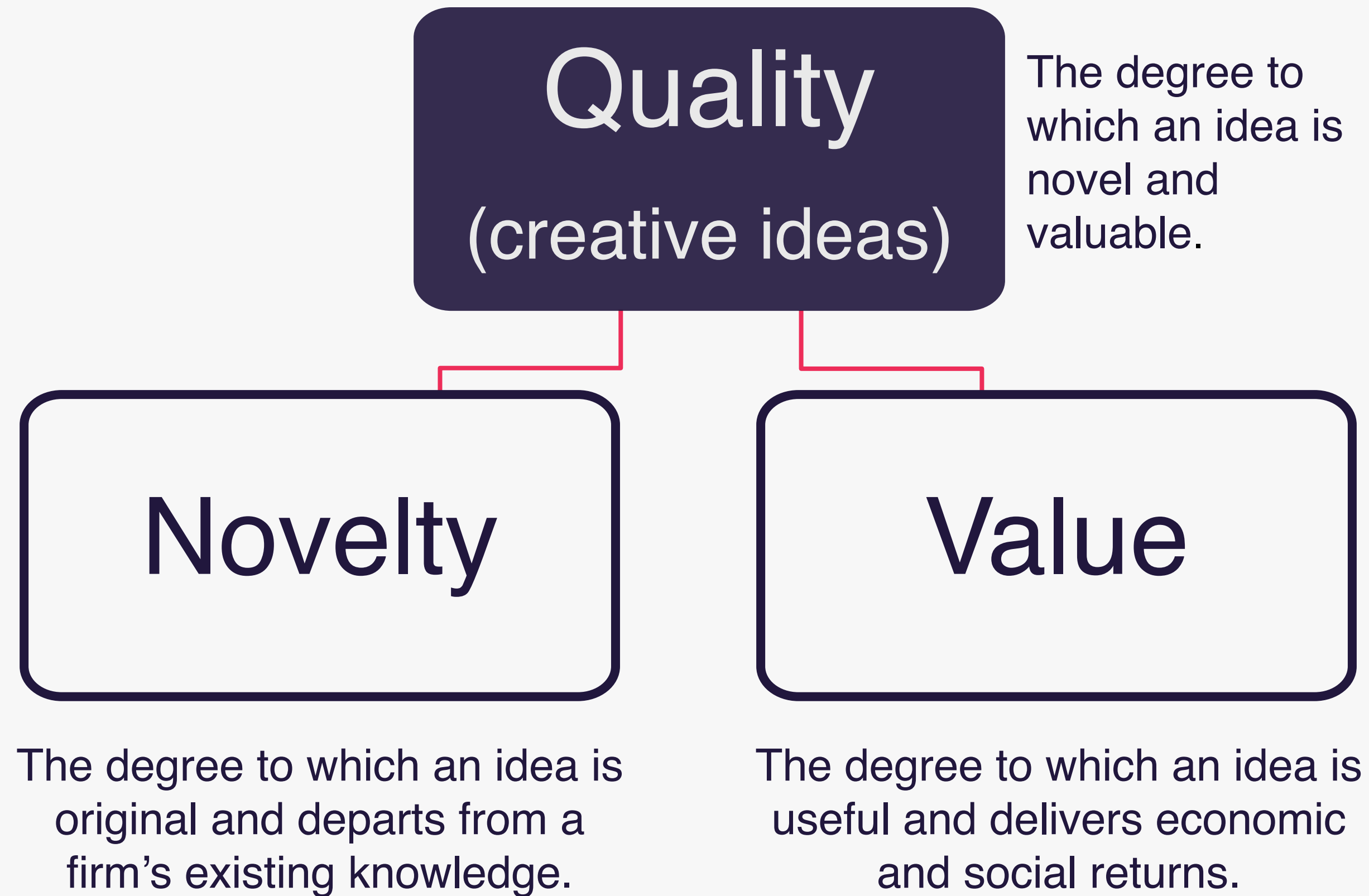
Valuable solutions



Human-AI:
Prompt GPT-4

Appendix

High quality, creative ideas balance novelty and value



The Possible Drawbacks of LLMs for Creative Problem-Solving



Confabulations



“Trapped” in data



Devoid of context

Context: Crowdsourcing Challenge on Circular Economy

Real-world challenge through a partnership with Continuum Lab and Freelancer.com around Circular Economy.

Reason for Choice:

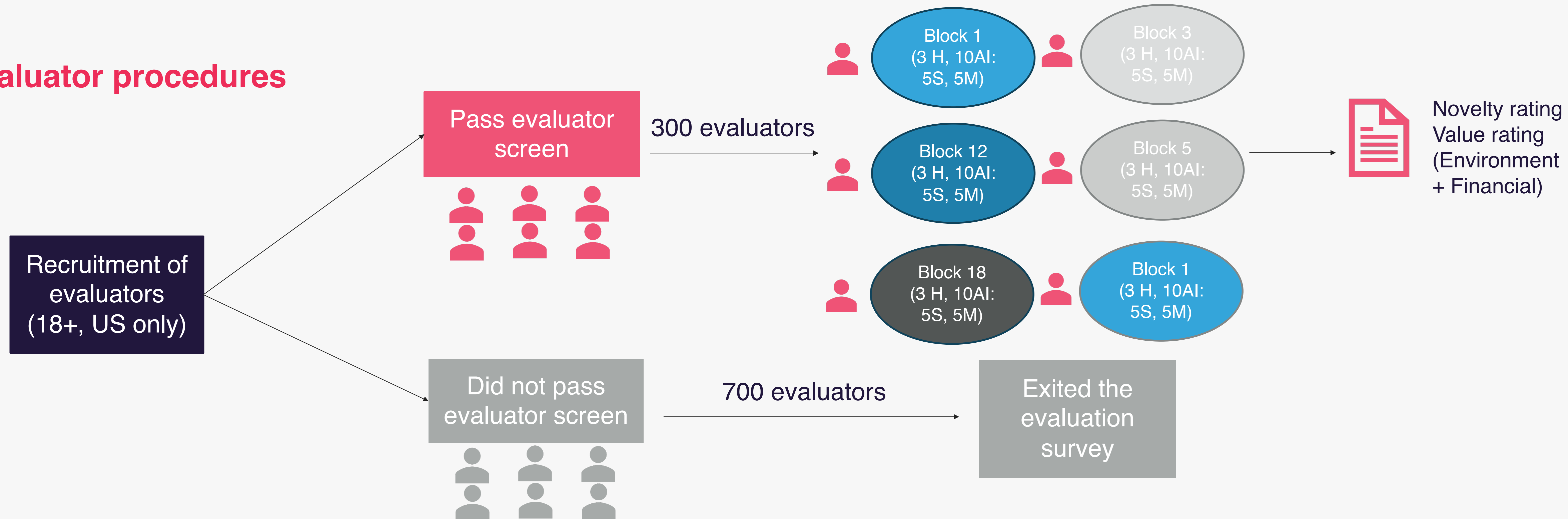
1. Open and multi-disciplinary problem
2. Requires domain knowledge
3. Real-world global implications with cultural nuances



Three ChatGPT Prompting Levels



Evaluator procedures



Recruited evaluators (18+, U.S. only) on Prolific

Each potential evaluator completed an initial screen for their interest, work experience, and knowledge in the circular economy. Evaluators with moderate or high interest, and either 2+ years of work experience, or 60% or more on the knowledge skills test passed the screen. Out of 1000 individuals screened, 300 passed.

Each of the 300 evaluators passing the circular economy screen was then randomly assigned one out of 18 blocks of circular economy solutions. Each block contained 13 solutions: 3 human (H) and 10 AI solutions. Of the AI solutions, 5 were generated using a single (S) instance of GPT-4, and 5 were generated using multiple (M) instances of GPT-4.

Each evaluator rated the novelty, environmental, and financial value of each assigned solution. The value rating is an average of the environmental and financial ratings.

Table 1. Nested Mixed Effects Models of Evaluator Ratings of Solution Creativity on Solution Source

Mixed Effects Models of Solution Creativity on Solution Source

	Dependent Variable: Solution Creativity	
	(1)	(2)
HAI Solution	0.066 (0.171)	
HAI Multiple Instance		-0.232 (0.190)
HAI Single Instance		0.363+ (0.190)
Intercept	12.256*** (0.244)	12.256*** (0.244)
N	3900	3900
# blocks	18	18
# evaluators	300	300
Screening criteria	N	N
Other controls	N	N
Log-Likelihood	-11721.69 df = 5	-11716.04 df = 6

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes. This table presents mixed-model (hierarchical linear modeling) results from evaluator ratings of solution creativity (solution novelty x solution value), with 300 evaluators nested in eighteen solution blocks. Models 2-3 and 5-6 include the screening criteria: Work Experience, Level of Interest, and Knowledge Test Score. Models 3 and 6 include the following covariates: Gender, Highest Level of Education, Major, Employment Status, Cohort Session, and Solution Word Count. Standard errors are in parentheses.

1 No difference in average creativity by solution source

2 H-AI Single instance slightly outperforms human crowd outputs

Mixed Effects Models of Top Solution Creativity on Solution Source

Table 2. Nested Mixed Effects Models of Evaluator Ratings of Top Solution Creativity on Solution Source

	Dependent Variable: Top Solution Creativity	
	(1)	(2)
HAI Solution	① -0.004 (0.006)	
HAI Multiple Instance		② -0.003 (0.007)
HAI Single Instance		-0.005 (0.007)
Intercept	0.038*** (0.007)	0.038*** (0.007)
N	3900	3900
# blocks	18	18
# evaluators	300	300
Screening criteria	N	N
Other controls	N	N
Log-Likelihood	1236.12 df = 5	1231.99 df = 6

① No difference in top creativity by solution source

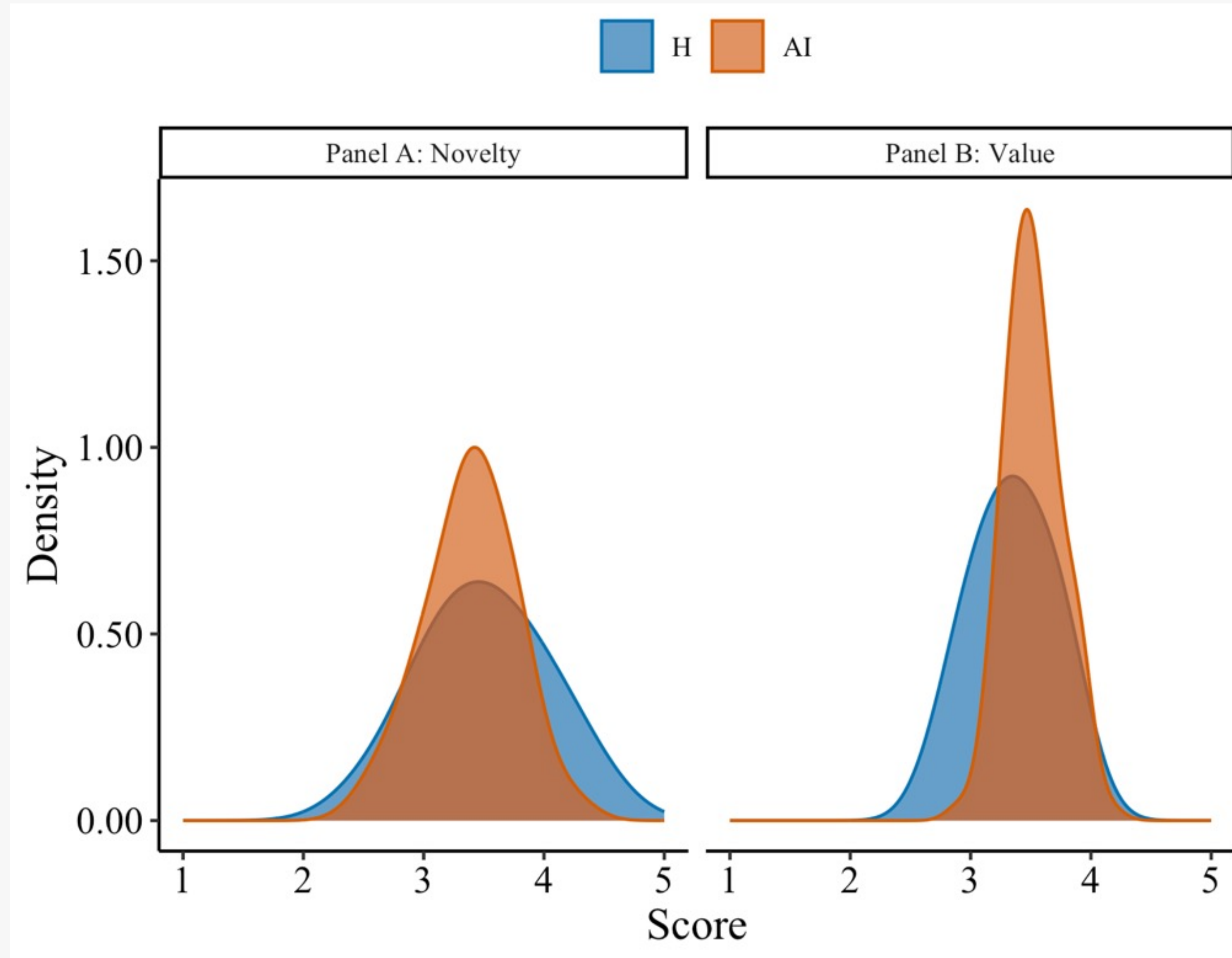
② No difference in top creativity by HAI single or multiple instance and HC

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes. This table presents mixed-model (hierarchical linear modeling) results from evaluator ratings of top solution creativity, with 300 evaluators nested in eighteen solution blocks. Models 2-3 and 5-6 include the screening criteria: Work Experience, Level of Interest, and Knowledge Test Score. Models 3 and 6 include the following covariates: Gender, Highest Level of Education, Major, Employment Status, Cohort Session, and Solution Word Count. Standard errors are in parentheses.

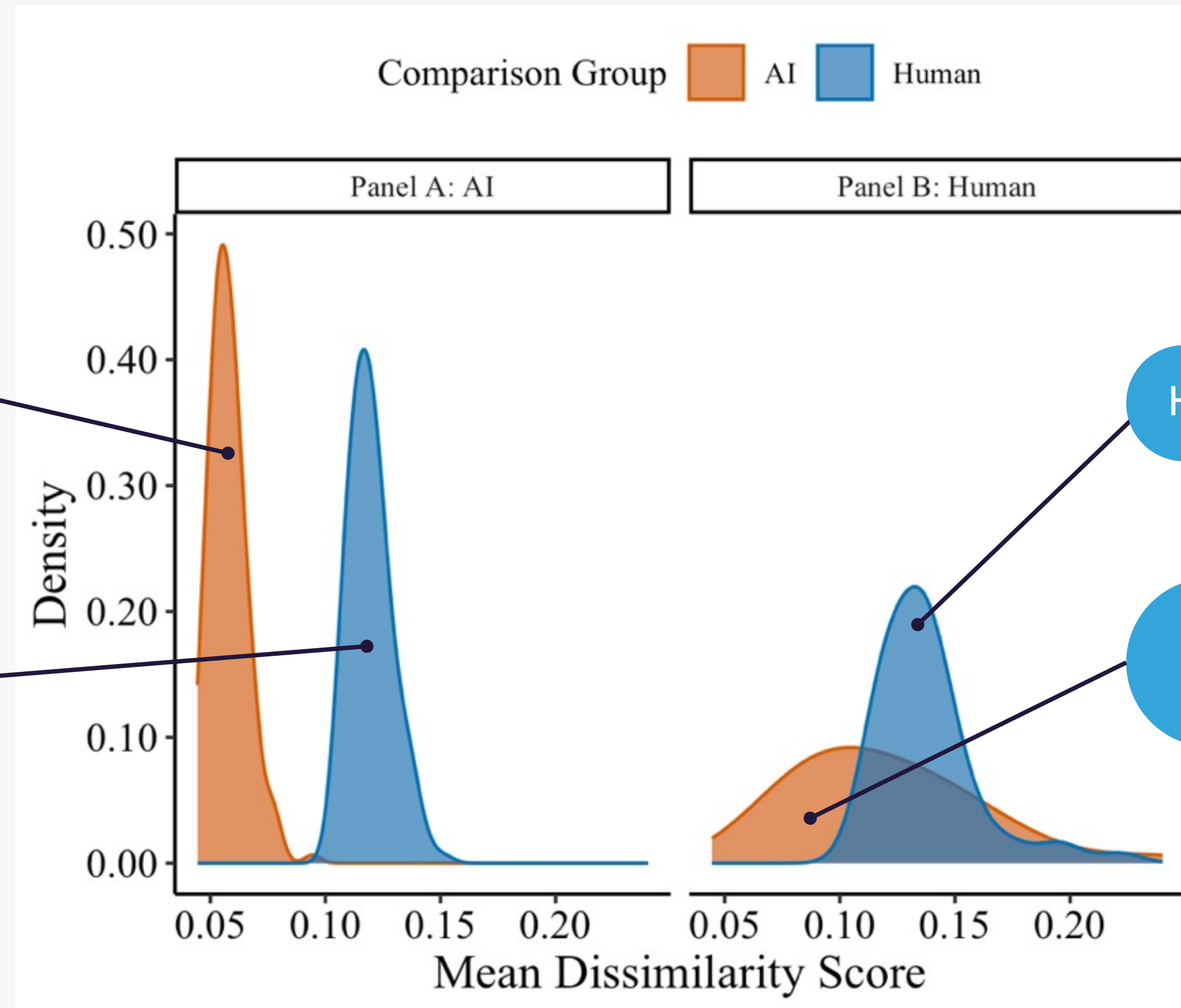
Distribution of Mean Novelty and Value Ratings

Figure 1. Density plot distributions of the mean novelty (Panel A) and value (Panel B) evaluator ratings by solution source.



Solution Semantic Level Dissimilarity between AI and Human Solutions

Figure 3. Density plot distributions of the mean dissimilarity score by solution source. Comparison made within- and across-AI (Panel A) or human (Panel B) sources.



AI solutions are similar to each other.

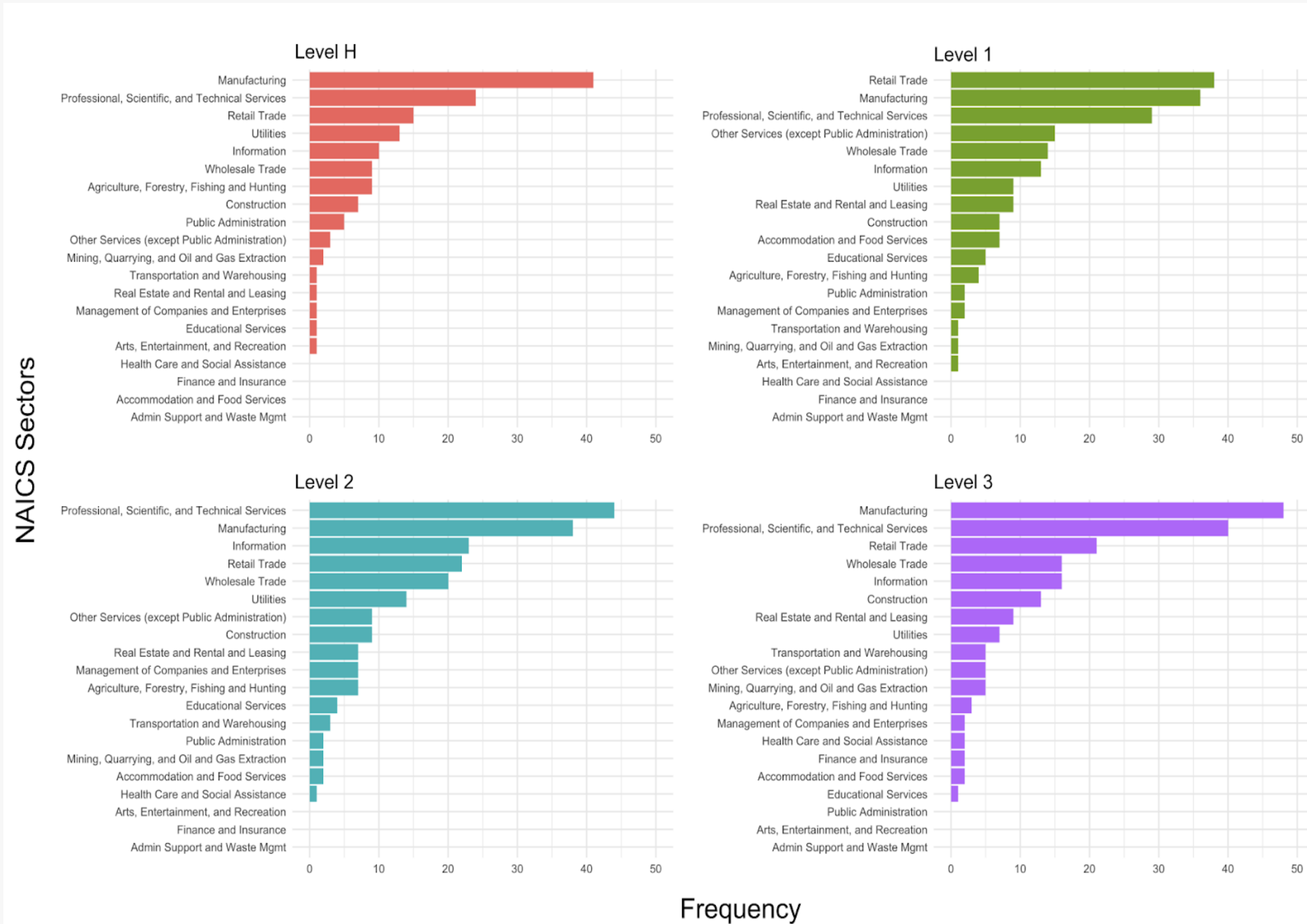
AI solutions are dissimilar to the typical human solutions.

Human solutions are diverse.

Some human solutions are similar to the typical AI solutions.

Note: The mean dissimilarity score is calculated using the average pairwise cosine distances using BERT embeddings.

Similar Breadth of Industry Coverage Across Human and AI Solutions



AI EarthHack

AI Idea Filters and Validators



Camila Lin



Pei-Hsin Wang



Ying Hao Chen

Rankings

🏆 Best Overall

Rank	Name
1	Aladdin Use Gambling
2	Aladdin Use Gambling
3	Reusing Space Junk
4	Advanced Recycling
5	Use Circular Economy

♻️ Eco Friendly

Rank	Name
1	Use Circular Economy
2	Showcase Virtual Fashion
3	Economy Fashion Leasing
4	Circular Economy Initiative
5	Greenar Bulletin Board

💰 Business Plan

Rank	Name
1	Smart Menstrual Cups
2	Idea Foldable Tablet
3	Notebooks Designed Environmental
4	Drinking Tea Clay
5	Fashion Sustainability

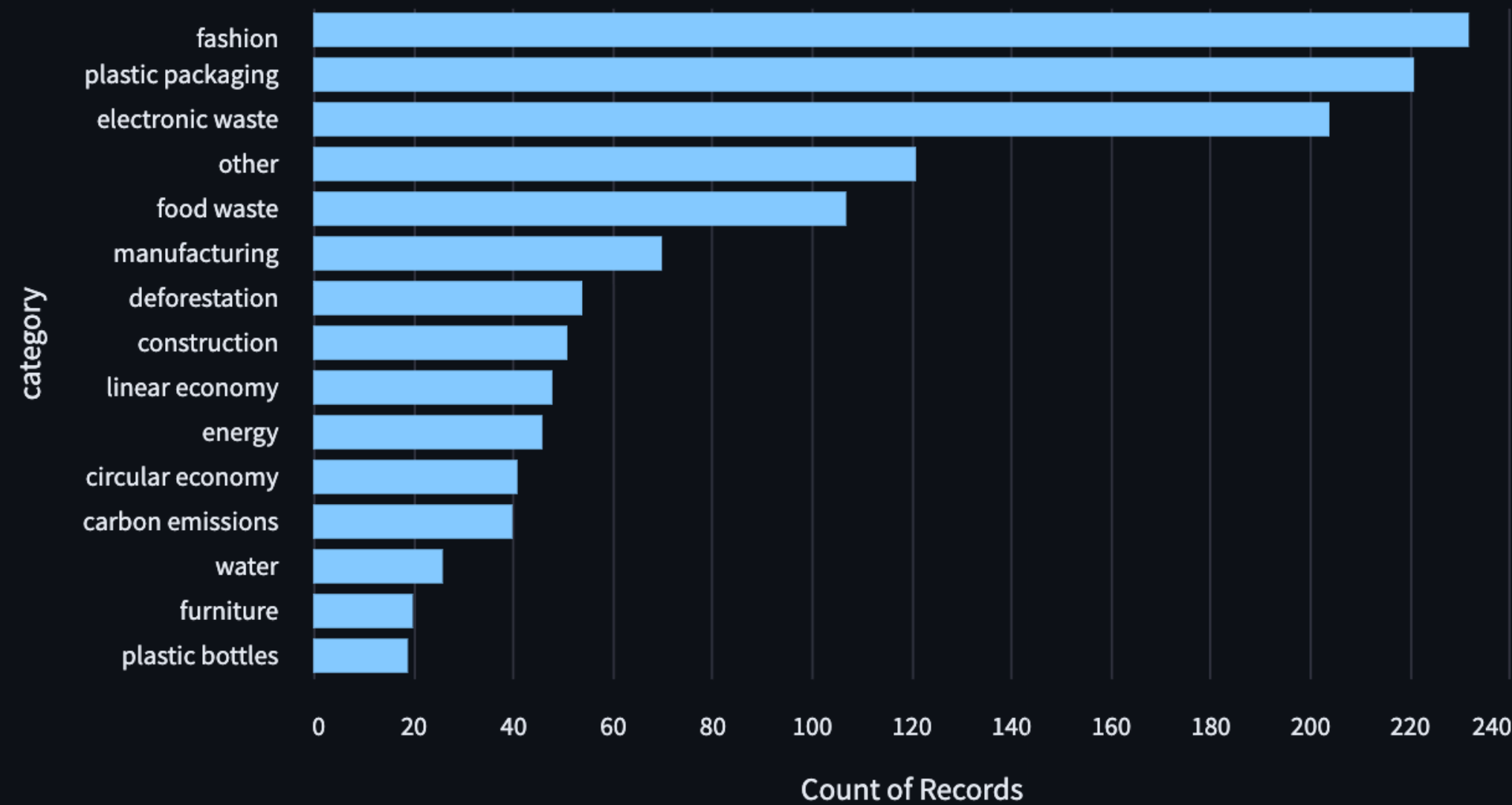
💡 Innovative

Rank	Name
1	Ink Residue Feel
2	
3	
4	Home Sliding Earthquake
5	Assessment Methodology Scoring

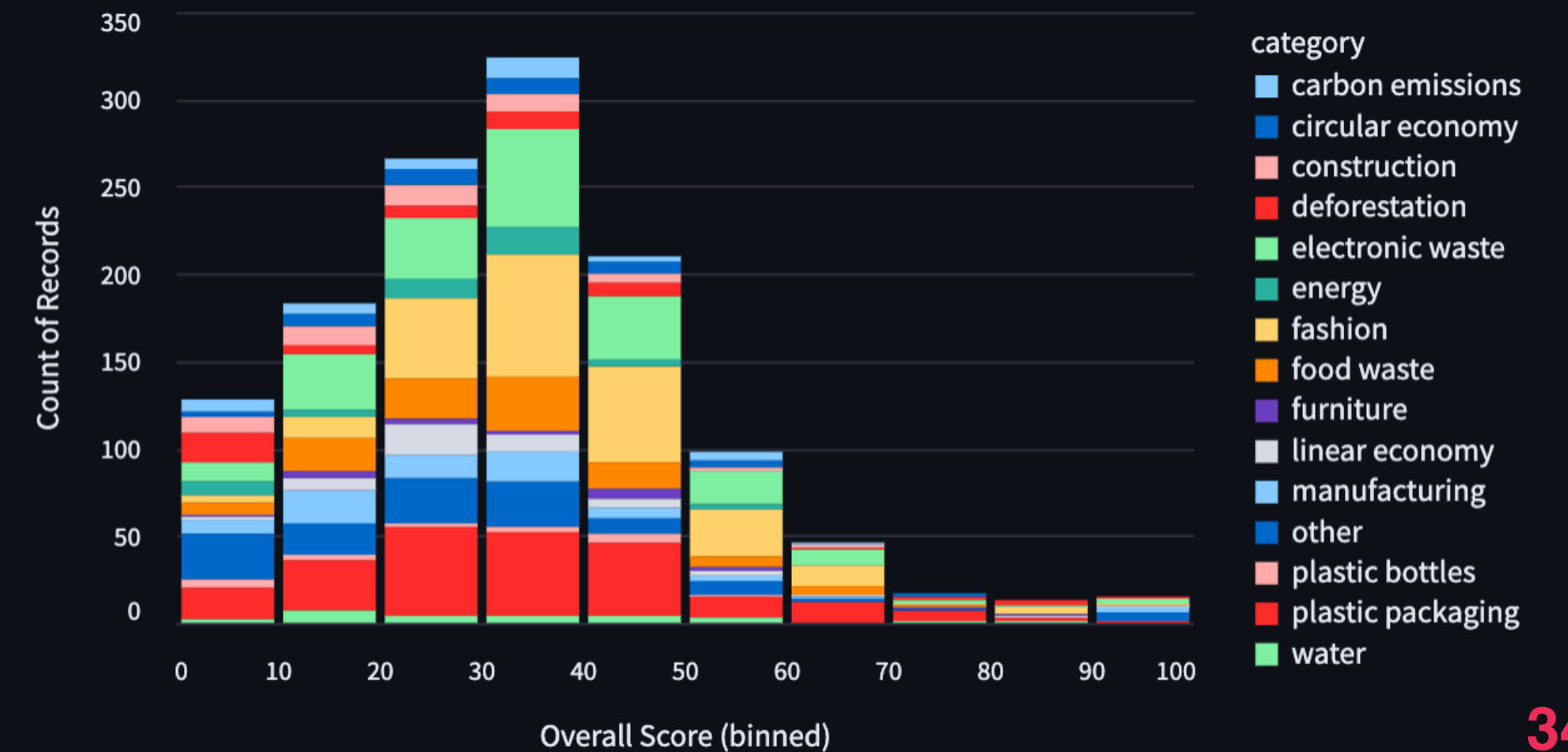
Overview

Link: <https://sprout.yhcapp.net/>

Number of Ideas in each Category



Overall Score Distribution



🔄 Circular Economic Impact

- ✔ Make better use of finite resources
- ✔ Reduce emissions
- ✔ Protect human health and biodiversity
- ✔ Boost economies
- ✔ Create more and better jobs

👛 Business

- ✔ Value Proposition
- ✔ Market Analysis
- ✔ Target Audience
- ✔ Revenue Streams
- ✔ Go-to Market Strategy

📄 Product and Resource

- ✔ Technology Development Maturity
- ✔ Technology Scalability
- ✔ Intellectual Property Potential
- ✔ Financial Plan
- ✔ Team Structure

Competition Landscape

Novelty Rank

351

There are 5 similar ideas

