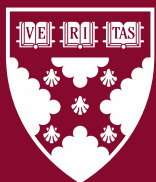


Working Paper 24-005

# The Crowdless Future? How Generative AI is Shaping the Future of Human Crowdsourcing

Léonard Boussioux  
Jacqueline N. Lane  
Miaomiao Zhang  
Vladimir Jacimovic  
Karim R. Lakhani



**Harvard  
Business  
School**

# The Crowdless Future?

## How Generative AI is Shaping the Future of Human Crowdsourcing

Léonard Boussioux

University of Washington

Massachusetts Institute of Technology

Jacqueline N. Lane

Harvard Business School

Miaomiao Zhang

Harvard Business School

Vladimir Jacimovic

Harvard Business School

Karim R. Lakhani

Harvard Business School

**Working Paper 24-005**

Copyright © 2023 by Léonard Boussioux, Jacqueline N. Lane, Miaomiao Zhang, Vladimir Jacimovic, and Karim R. Lakhani.

Working papers are in draft form. This working paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.

We are thankful for Harvard Business School Research Associate Justin Ho and Program for Research in Markets & Organizations (PRIMO) fellow Stella Jia, who supported the data analysis for this project.

Funding for this research was provided in part by Harvard Business School.

# The Crowdless Future?

## How Generative AI Is Shaping the Future of Human Crowdsourcing

Léonard Boussioux<sup>1</sup>, Jacqueline N. Lane<sup>2†</sup>, Miaomiao Zhang<sup>2</sup>,  
Vladimir Jacimovic<sup>2,3</sup>, & Karim R. Lakhani<sup>2</sup>

<sup>1</sup>University of Washington, Michael G. Foster School of Business & Massachusetts Institute of Technology, Operations Research Center

<sup>2</sup>Harvard Business School & Digital, Data, and Design (D<sup>3</sup>) Institute at Harvard

<sup>3</sup>ContinuumLab.AI

\*Corresponding author: [jnlane@hbs.edu](mailto:jnlane@hbs.edu)

†Léonard Boussioux and Jacqueline N. Lane share co-first authorship

### Abstract

This study investigates the capability of generative artificial intelligence (AI) in creating innovative business solutions compared to human crowdsourcing methods. We initiated a crowdsourcing challenge focused on sustainable, circular economy business opportunities. The challenge attracted a diverse range of solvers from a myriad of countries and industries. Simultaneously, we employed GPT-4 to generate AI solutions using three different prompt levels, each calibrated to simulate distinct human crowd and expert personas. 145 evaluators assessed a randomized selection of 10 out of 234 human and AI solutions, a total of 1,885 evaluator-solution pairs. Results showed comparable quality between human and AI-generated solutions. However, human ideas were perceived as more novel, whereas AI solutions delivered better environmental and financial value. We use natural language processing techniques on the rich solution text to show that although human solvers and GPT-4 cover a similar range of industries of application, human solutions exhibit greater semantic diversity. The connection between semantic diversity and novelty is stronger in human solutions, suggesting differences in how novelty is created by humans and AI or detected by human evaluators. This study illuminates the potential and limitations of both human and AI crowdsourcing to solve complex organizational problems and sets the groundwork for a possible integrative human-AI approach to problem-solving.

**Keywords:** Generative AI, ChatGPT, LLMs, innovation, crowdsourcing, idea generation, evaluation, novelty, value

We are thankful for Harvard Business School Research Associate Justin Ho and Program for Research in Markets & Organizations (PRIMO) fellow Stella Jia, who supported the data analysis for this project. All errors remain our own.

## 1. Introduction

In 2006, Netflix launched a competition offering a \$1 million prize to anyone who could improve the accuracy of its movie recommendation algorithm by 10%. The competition attracted thousands of teams from around the world, and three years later the grand prize was awarded to a team that combined several different machine learning methods to achieve this previously insurmountable goal (Bennett & Lanning, 2007).

The Netflix anecdote illustrates the competitive advantage of crowdsourcing as a strategic model. Organizations like Netflix, LEGO, Procter & Gamble, Starbucks, and Dell use crowdsourcing to tap into the collective intelligence of a large and diverse group, commonly known as the “crowd,” to solve problems and generate innovative ideas. By opening up challenges to a crowd, organizations can access a broader range of ideas and solutions than they might otherwise create internally (Chesbrough, 2003; Jeppesen & Lakhani, 2010). A key advantage of crowdsourcing is the diversity of perspectives it brings to the problem-solving process. This diversity can lead to more innovative and successful solutions, increasing the chances of uncovering novel, extreme-value ideas or unique combinations of existing ideas (Jeppesen & Lakhani, 2010; Terwiesch & Ulrich, 2009). Given the rise of generative AI, and its potential to generate ideas rapidly and efficiently, we ask: *Does generative AI have the potential to replace human crowdsourcing for innovative ideas to solve business problems?*

The literature on open innovation indicates that the crowd offers several benefits for generating problem- and solution-based knowledge. First, crowdsourcing brings “marginality” or diversity in knowledge and perspectives that may be amenable to solving the problem at hand (Jeppesen & Lakhani, 2010; Page, 2008). Although specialists with prior knowledge and deep expertise in the domain are often successful in developing solutions to known problems or tasks, they frequently struggle with tackling new, unprecedented problems (Allen, 1970, Lovett & Anderson, 1996, Sørensen & Stuart, 2000). Marengo et al. (2000) and Page (2008) highlight the effectiveness of leveraging diverse problem solvers through platforms like Innocentive.com or Kaggle. The insight behind crowdsourcing is that the selected perspective can significantly influence the perceived difficulty of a problem (Marengo et al., 2000; Page, 2008). Therefore, even in situations where a problem seems new, challenging, or even impossible to some, the unique perspectives of others peripheral to the knowledge domain of the problem could potentially bring a novel, breakthrough solution. Similarly, needs-based crowdsourcing can potentially increase an organization’s value by improving its ability to identify and address relevant problems (von Hippel, 1986; von Hippel & von Krogh, 2016). Although organizations may have access to innovative technologies, they may sometimes lack the specific knowledge required to apply these technologies to solve real-world problems (Catalini & Tucker, 2016; Gruber et al., 2008; Shane, 2000). Crowds can aid in highlighting innovative applications for such technologies (Dahlander et al., 2019).

Second, crowdsourcing increases the likelihood of generating exceptional ideas rather than just average or modal ones (Terwiesch & Ulrich, 2009; Girotra et al., 2010). These outliers align with extreme value theory, which focuses on the upper-end distribution of events, ideas, or inventions (Dahan & Mendelson, 2001; Terwiesch & Loch, 2004; Terwiesch & Ulrich, 2009). The theory proposes that extraordinary successes, despite their rarity, unpredictability, and difficulty to obtain, can produce greater value than numerous “good” but not outstanding outcomes (Dahan & Mendelson, 2001).

Third, broadcast search, using markets to solve complex and ill-structured problems (Fernandes and Simon, 1999), offers efficiency benefits beyond the value creation associated with it. Organizations have increasingly adopted crowdsourcing as an innovation strategy to seek knowledge beyond organizational boundaries (Boudreau & Lakhani, 2013; Baldwin & von Hippel, 2011; Faraj, et al., 2016; von Hippel, 2005). This is because crowdsourcing provides a cost-effective means for organizations to expand the range of viewpoints they can access. It enables the crowd to generate various potential solutions, from which organizations can then choose the most promising ones to implement (Boudreau et al., 2011). The extended potential scope of crowdsourcing, as well as the reduction in information and computational costs followed by the pervasiveness of the Internet, has drastically increased the number of approachable opportunities and enabled sophisticated problem-solving at the individual level. As such, crowdsourcing allows organizations to capitalize on the power of diversity and the extraordinary potential of outlier ideas.

Despite these advantages of crowdsourcing, many organizations fail to crowdsource successfully because crowds differ in how they are organized compared to traditional internal sourcing of ideas. Successful crowdsourcing of new products or services requires careful planning and management. This includes formulating and broadcasting the problem at hand at the optimal specificity level (Fernandes & Simon, 1999), creating a clear and compelling challenge that delineates the characteristics of the problem, the solution, the knowledge needed to solve the problem, and the expertise required to evaluate the solution (Dahlander et al., 2019). In addition, organizations need to consider the complications involved in the evaluation and selection process (Koput, 1997; Laursen & Salter, 2006), provide the appropriate incentives to motivate participation, and set up a platform for idea submission and collaboration, while ensuring fair treatment of contributors (Afuah & Tucci, 2012; Dahlander & Piezunka, 2020; Lifshitz-Assaf, 2018; Piezunka & Dahlander, 2015).

Generative AI has become a promising tool to augment or replace human crowdsourcing in certain contexts (Allen et al., 2023). Generative AI, a subset of artificial intelligence, is characterized by its ability to generate new data that closely mirrors the characteristics of the data used to train it (Brown et al., 2020). This can include creating text, images, music, and other forms of creative content. Generative AI models are particularly well-suited to propose diverse ideas due to their training on a vast amount of data and a vast capacity to generate a wide variety of potential solutions or concepts that humans may not think of. In situations where the barriers to designing a crowdsourcing challenge to attract a large and diverse crowd or where time or resources are limited (Dahlander & Piezunka, 2020), generative AI could potentially aid human crowdsourcing by creating a large number of ideas at scale both quickly and efficiently, either replacing or augmenting human participation. Hence, generative AI could be more scalable and consistent in developing high-quality output than human crowdsourcing.

On the other hand, the capabilities of generative AI may fall short of emulating the multifaceted nature of human-based crowdsourcing. Not all problems addressed through crowdsourcing may be delegated to generative AI (Boudreau & Lakhani, 2013), particularly if the innovativeness and value of ideas depend on the problem's features and the crowdsourcing challenge's goals. These factors might lead to potential tradeoffs in the type of solutions that are produced by AI versus humans. Moreover, AI may lack the ability to understand social and cultural nuances, which can be crucial in generating truly innovative and valuable ideas (Amabile, 2018). Generative AI might grapple with the challenge of paralleling human creativity

and the capacity to undertake intuitive leaps encompassing serendipity, unpredictability, and profound domain expertise (Lane et al., 2021). Despite its potential, the application of generative AI to solve innovation-related problems and guide managerial decision-making remains in its infancy and relatively unexplored to the best of our knowledge (Cheng et al., 2022).

To reconcile these competing perspectives on the use of generative AI, together with ContinuumLab.AI, a Venture Studio focused on AI startup innovation, incubation, and investment, we developed a crowdsourcing challenge to develop new business ideas on the circular economy in order to directly compare the capabilities of human-generated and AI-generated solutions. We chose the circular economy because it is a broad problem that involves a wide range of disciplines, such as environmental science, economics, design, and engineering, and corresponds to a robust test of the ability of humans and AI to generate innovative and high-quality solutions that consider multiple perspectives and types of knowledge. Moreover, the circular economy is a global matter with real-world applicability and the potential to draw ideas from different cultures and contexts, but, importantly, it's also a problem that requires relevant knowledge in the topic area to make a novel yet applicable contribution.

Our results are three-fold. First, they underscore a compelling tradeoff between human- and AI-generated responses. Although our evaluations indicate no difference in the overall quality between the human- and AI-generated solutions, we find a significant difference between their novelty and value. Human ideas receive higher novelty ratings, while AI ideas are perceived as more valuable due to their potential for significant environmental and financial impact on businesses. Second, we compare the diversity of our human- and AI-generated solutions. Leveraging the semantic richness of our text data, we adopt natural language processing techniques to measure the diversity of solutions. The semantic dissimilarity scores indicate that humans utilize more extensive vocabulary than AI. The level of differentiation also strongly correlates with evaluators' novelty ratings for human solutions, whereas differentiation in semantics does not correlate as much with their perceptions of value. Third, we apply industry clustering from ChatGPT to gauge the scope of industry applications for the respective human and AI groups. We find that humans and AI generated a similar range of industry classifications.

Overall, our study makes several contributions. We contribute to the open innovation literature by offering a novel and more scalable way of enhancing human crowdsourcing with AI. Our findings indicate that current generative AI does not completely match human ingenuity but can potentially overcome some of the limitations of traditional crowdsourcing through effective prompt engineering. Moreover, our contributions extend to the broader innovation literature on the distinction between novelty and value (Berg, 2016; Dewar & Dutton, 1986; Ghosh & Wu, 2021; Rindova & Petkova, 2007). We suggest a similar tradeoff may exist between human- and AI-led idea generation. In particular, given the current technological capabilities and/or prompt engineering used in this study, AI crowdsourcing cannot achieve the same level of novelty as the human crowd but may offer additional promise for value creation compared to humans alone. Since creativity often hinges on generating novel *and* useful ideas (Amabile, 1996), our findings open up a new debate beyond AI as a substitute for human capabilities but as a complement for providing solutions that balance the simultaneous pursuits of novelty and value.

Lastly, we offer a new perspective to the emerging literature on generative AI, which mostly focuses on the productivity implications in the post-Gen AI era (e.g., Byrnjolfsson et al.

2023; Noy & Zhang, 2023; Gilardi et al., 2023; Horton, 2023). Responding to the call for more study on generative AI as a complementary asset in management research (Berg et al., 2023), one recent article discusses the clear advantages of LLMs over humans (Girotra et al., 2023); others offer some limitations regarding LLMs' creativity (Haase et al., 2023). Still, to the best of our knowledge, we are the first to examine the applications of generative AI to organizational strategy pertinent to idea generation at scale, with more sophisticated prompt engineering and evaluation processes.

## **2. Research Design, Recruitment of Evaluators, and Methods**

### **2.1. Generating Human Solutions: Circular Economy Crowdsourcing Challenge**

We partnered with Freelancer.com and ContinuumLab.AI, a US-based AI venture studio, to launch a crowdsourcing challenge to come up with new business ideas on “Creating a Sustainable Future: Unlock the Potential of Circular Economy” from January 30, 2023, to May 15, 2023. Participants were encouraged to submit real-life use cases on how companies can implement the circular economy in their businesses. New ideas were welcome, even if they were ‘moonshots’. As part of the contest details, participants were told that their ideas would be evaluated using four criteria: Novelty, Environmental Impact, Financial Impact, as well as Feasibility and Scalability of Implementation.

All participants submitted their solutions using a Google Form. Accompanying their submissions, we also collected demographic information, including their job title, geography, industry of application for their solution (a dropdown of 23 industries), and solution maturity (ideation, R&D, proof of concept, market testing, or full commercial). The contest received a total of 310 submissions. Each participant received \$10 for submitting a solution, and the best overall solution received a \$1,000 prize. Of these 310 submissions, our industry partner manually filtered out off-topic, incomplete, and blank entries, resulting in a total of 125 submissions that were deemed eligible to provide a solution to the crowdsourcing challenge.

### **2.2. Generating AI Solutions: Prompt Engineering the Wisdom of the Crowd**

We use GPT-4 (a large language model or LLM developed by OpenAI) to generate a range of solutions in response to the same crowdsourcing challenge described in section 2.1.

Because prompt engineering (i.e., the art and science of crafting effective input prompts to generate desired responses from the AI) can substantially impact the output from AI models, including the relevance or quality of the output, we generate solutions using three levels of prompting. As shown in Table 1, level 1 matches the initial problem description given to the human solvers to draw a baseline comparison between the quality of the human- and AI-generated solutions. In addition to the baseline problem description, level 2 adds individual-level solver characteristics for each of the 125 human solvers who submitted a solution (i.e., job title, industry, geography, and solution maturity). Thus, level 2 enables us to reproduce the characteristics of the human crowd that initially opted into the crowdsourcing challenge. In doing so, the intention is to simulate the context in which the human crowd was operating, possibly enabling the AI to generate solutions more similar to what human solvers would produce.

Lastly, we introduce *expert persona prompting* in level 3. By tapping into GPT-4's extensive training data, we enable the model to mimic a diverse crowd's perspectives and generate varied, context-specific outputs. In a separate instance, we first instruct the model to provide five distinct real-world “expert” personas for each of the 23 industries from the original

contest. The aim is to represent industry-specific expertise through these personas. Next, we ask GPT-4 to generate unique solutions, each reflecting its assigned expert persona’s perspective and knowledge.

We use two alternative configurations for each level reported in Table 1 to generate AI responses. The first configuration uses separate instances of GPT-4 to generate each solution. The second configuration uses a single instance of GPT-4 to generate multiple solutions successively, one at a time. The first configuration mimics the concept of independent crowd solvers more closely, as each instance of GPT-4 operates independently, starting from a different initialization and possibly exploring different areas of the problem and solution space. The second configuration may allow for more diverse, creative outputs, as all solutions are generated by the same model configuration and state, in which the output of previously generated solutions informs the input of the next and helps explore different aspects. Appendix A provides details of the prompt engineering used to produce the outputs associated with the approaches described in Table 1.

We created a total of 730 AI solutions. For levels 1 and 2, we generated 250 pairs each, divided equally between single and multiple instances (125 each). For level 3, we produced 230 pairs, also divided between single and multiple instances (115 each). These level 3 pairs correspond to the 23 industries and the five expert personas generated within each industry. Each solution was generated in 27.2 seconds on average (min = 5.9s, max = 80.8s, std = 8.4s) from a Google Colab notebook and costed \$0.037 on average. Appendix B.1 provides examples of human-generated and AI-generated solutions for each level.

Table 1. Prompt Engineering Configurations Used to Generate AI Solutions

Prompt Engineering Configurations	Description	Rationale
Level 1	The AI model receives the same problem description given to human participants or solvers.	This baseline allows for a direct comparison between AI and human responses, as both parties receive identical initial conditions.
Level 2	In addition to the original problem description, the AI model is given individual solver characteristics reflecting the human crowd (i.e., job title, geographical location, industry, solution maturity).	By adding individual-level characteristics, the AI model’s context becomes more similar to the human crowd’s, potentially generating solutions closer to those produced by humans.
Level 3	The AI model receives the original problem description and individual-level personas of experts from 23 industries relevant to the circular economy.	Incorporating expert personas from various industries encourages the AI model to emulate the diverse perspectives of a knowledgeable crowd. Additionally, it encourages the generation of creative industry-specific solutions.

### 2.3. Evaluator Recruitment and Procedures

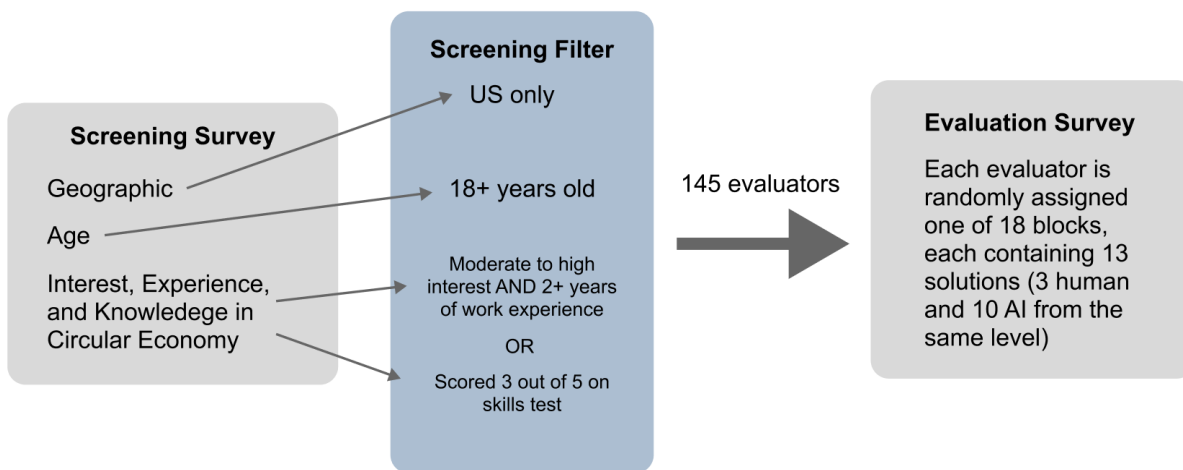


Our study uses human evaluators to judge the quality (average and likelihood of an extreme value outcome) of the human and GPT-4 solutions. First, as shown in Figure 1, we recruited potential evaluators on Prolific.org. We used a screening survey to screen potential evaluators for geographic location (US only) and age (18 years old or older), as well as for their level of interest, work experience, and knowledge of the circular economy through a multiple-choice skills test. Individuals who passed the screening filter (indicating a moderate level of interest or higher) and had either two or more years of work experience or scored 3 out of 5 (or 60%) or more on the skills test were invited to participate in the evaluation survey (see Appendix B.2 and B.3 for survey instruments). Overall, our recruited sample indicated an average level of interest in the circular economy of 3.84 out of 5 (s.d. = 0.80, min = 3, max = 5), an average experience of 3.82 years in related industries (s.d. = 3.24, min = 0, max = 14.5), and an average score of 2.63 out of 5 on the skills test (s.d. = 1.11, min = 1, max = 5).

Then we randomly selected 180 AI-generated (i.e., 60 pairs for each level)<sup>1</sup> and 54 human-generated solutions, totaling 234 pairs, for human evaluation. As shown in Figure 1, we developed an evaluation survey employing a randomized block design. This design randomized the AI- and human-generated solutions into distinct blocks. Each block included 10 AI-generated solutions from the same level and three human-generated pairs, totaling 13 pairs per block and 18 unique blocks overall.

As a motivation to exert effort and to ensure thoroughness, we offered each evaluator \$12 for completing the survey, with a bonus of \$1 for each solution where they matched the consensus or mode quality rating among all evaluators assigned to the same solution. Each evaluator received a mean bonus of \$6.35 (s.d. = \$2.28, min = \$1, max = \$11). The total compensation per evaluator ranged from \$12 to \$25.

Figure 1. Flow of Evaluator Screening and Survey Procedures



Each evaluator rated the solutions on the four criteria given to the solvers. (Novelty: *How different is it from existing solutions?* Feasibility and Scalability of Implementation: *How likely is it to succeed and how scalable is it?* Environmental Impact: *How much does it benefit the planet?*)

<sup>1</sup> For each level, we randomly selected 30 solutions from the multiple-instance configuration, and 30 solutions from the single-instance configuration.

Financial Impact: *What financial value can it create for businesses?* And an overall Quality rating: *Based on the four criteria above, what is the overall quality of the solution?*) Following their evaluations, the evaluators reported their demographics, levels of expertise, and understandings of the circular economy.

We recruited 145 evaluators to evaluate 13 responses each, for a total of 1,885 observations and a mean of 8.06 ratings (s.d. = 0.91, min = 6, max = 9) for each solution.<sup>2</sup> Table 2 reports summary statistics on our evaluator sample.

**Table 2.** Evaluator Summary Statistics (N = 145 evaluators)

Female	40.69%
Age	41.04 (12.06)
Employed or self-employed	81.38%
Student	2.07%
Bachelor's degree or more	64.83%
Expertise	2.59 (0.98)
Understanding	3.36 (0.79)

Note: standard deviations reported in parentheses.

## 2.4. Variables

### 2.4.1. Dependent Variables

We use four dependent variables, corresponding to the evaluator's *Novelty rating*, *Feasibility rating*, *Value rating*, and *Quality rating* of each solution. We computed the *Value rating* by taking the average of the evaluator's assigned environmental and financial impact ratings. To examine extreme value ratings, we create binary variables for *Top novelty rating*, *Top feasibility rating*, *Top value rating*, and *Top quality rating* equal to 1 if an evaluator assigned the highest rating (i.e., 5) to a solution, and 0 otherwise.

### 2.4.2. Independent Variables

Our main independent variable, *AI level*, is a categorical variable corresponding to whether the solution is human-generated (baseline), AI level 1, AI level 2, or AI level 3. For robustness, we also use alternative independent variables, such as *AI-generated*, which is a dummy variable that indicates whether the solution was AI-generated (equal to 1) or human-generated (equal to 0) and accounts for the GPT-4 configuration (multiple or single).

### 2.4.3. Other Variables

Our statistical analyses rely on the random assignment of evaluators to solutions. That being said, we also use evaluator dummies to facilitate within-evaluator comparisons and add problem and solution adjective fixed effects to account for the length and style of the AI-generated solutions (see Appendix A for details).

## 2.5. Estimation Approach

<sup>2</sup> This study was approved by the Harvard University IRB (IRB23-0770).

We use ordinary least squares (OLS) regression models to regress the evaluator  $i$ 's rating for solution  $j$  on whether the solution was AI- or human-generated, as shown in equation (1):

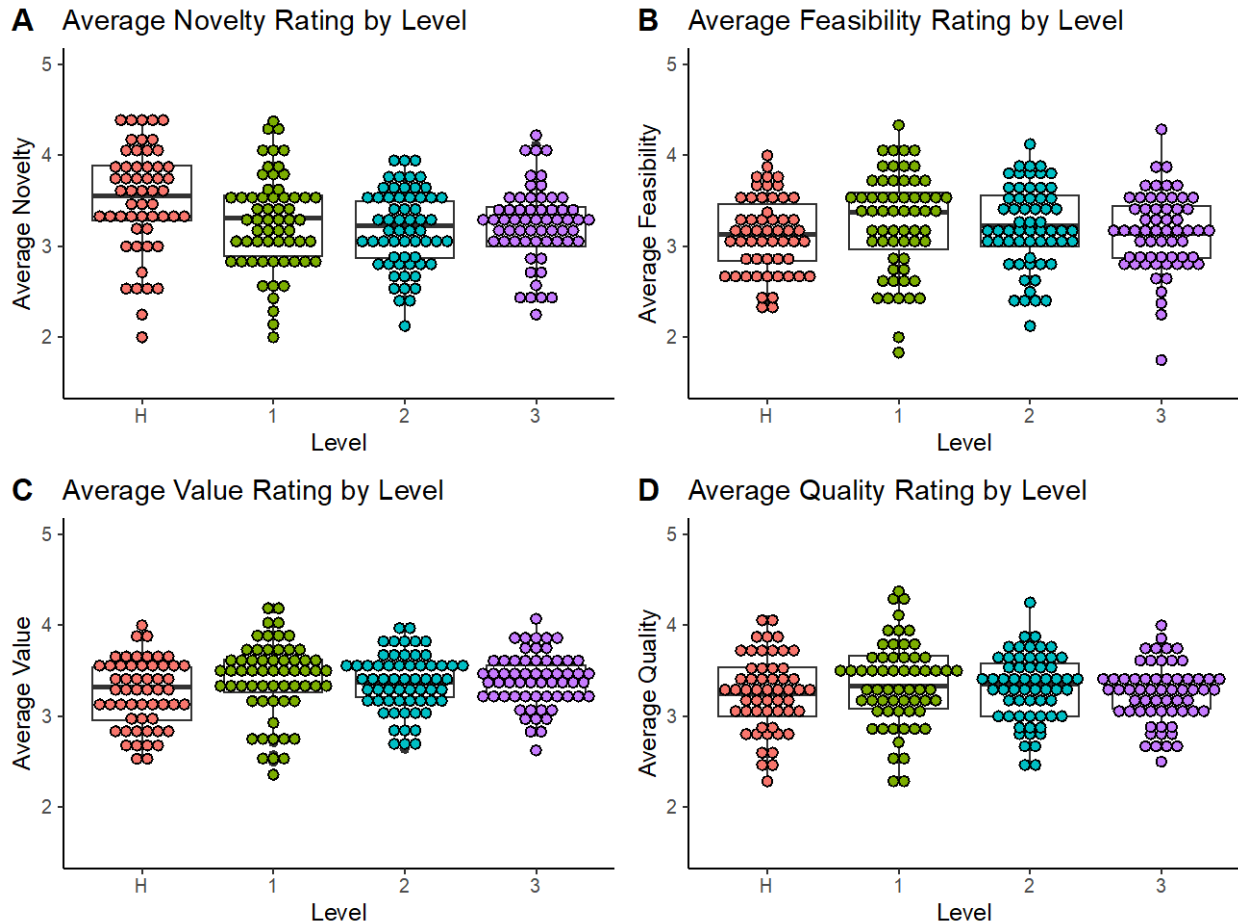
$$\text{Evaluator rating}_{ij} = \beta_0 + \beta_1(\text{AI level}_j) + \gamma_i + p_j + s_j + \varepsilon_{ij}. \quad (1)$$

In (1),  $\gamma_i$  corresponds to evaluator fixed effects,  $p_j$  corresponds to problem adjective fixed effects, and  $s_j$  corresponds to solution adjective fixed effects.

### 3. Results from Evaluator Ratings of Human and AI Solutions

Figure 2 illustrates the distribution of ratings by evaluation criteria for the human and AI solutions for the different levels of solutions (see Table 1). Overall, we find a significant difference by AI level for novelty ( $F(3,1881) = 9.71, p < 0.001$ ) and value ( $F(3,1881) = 3.05, p = 0.03$ ) but no significant difference for feasibility ( $F(3,1881) = 2.53, p = 0.07$ ) or overall quality ( $F(3,1881) = 1.58, p = 0.19$ ).

Figure 2. Comparison of Average Evaluator Ratings by Level (N = 234 solutions)



Note: Each dot represents the mean rating assigned to a solution. We use box plots to show the median, 25th and 75th percentiles, and the interquartile ranges of the distributions of ratings by level.

Next, Table 3 reports the OLS regression results of the evaluator ratings on *AI level*. We note that all models include evaluator fixed effects, as well as problem and solution adjective fixed

effects. As shown in Model 1, we find that, compared to human solutions, both level 1 (Model 1: -0.381,  $p < 0.05$ ) and level 3 (Model 1: -0.568,  $p < 0.01$ ) AI solutions were rated as less novel. While there was no difference in the feasibility ratings among the human and AI levels, we find that, compared to the human solutions, the level 1 (Model 3: 0.235,  $p < 0.10$ ) and level 3 (Model 3: 0.269,  $p < 0.05$ ) solutions were rated as being higher in value. Lastly, we observe no quality differences between the human- and AI-generated solutions. In Appendix C, we note that the reported results are robust to alternative specifications of the independent variable, *AI level*.

Table 3. OLS Regression Models of Evaluator Ratings on AI Levels

VARIABLES	Novelty Model 1	Feasibility Model 2	Value Model 3	Quality Model 4
<i>Baseline</i> = Human				
Level 1	-0.381** (0.192)	-0.0220 (0.168)	0.235* (0.125)	0.0480 (0.160)
Level 2	-0.221 (0.181)	0.0636 (0.167)	0.166 (0.124)	0.170 (0.155)
Level 3	-0.568*** (0.170)	-0.0661 (0.197)	0.269** (0.129)	-0.00780 (0.170)
Constant	3.538*** (0.0481)	3.147*** (0.0446)	3.266*** (0.0351)	3.274*** (0.0460)
Evaluator FE	Y	Y	Y	Y
Problem adjective FE	Y	Y	Y	Y
Solution adjective FE	Y	Y	Y	Y
Observations	1,885	1,885	1,885	1,885
R-squared	0.036	0.019	0.017	0.011
Number of evaluators	145	145	145	145

Robust standard errors in parentheses.

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table 4 examines the relationships between the top evaluator rating (i.e., evaluation score = 5) on *AI level*. Here, we observe that AI level 1 (Model 1: -0.288,  $p < 0.01$ ), level 2 (Model 2: -0.215,  $p < 0.01$ ), and level 3 (-0.290,  $p < 0.01$ ) solutions are less likely to receive top novelty ratings compared to human-generated solutions. That being said, there is no significant difference between the human and AI levels in the likelihood of receiving a top evaluation rating for feasibility or value, and a marginally significant difference for *Top quality* between AI level 3 and human problem-solutions pairs (Model 4: -0.049,  $p < 0.10$ ).

Table 4. OLS Regression Models of Top Evaluator Rating on AI Levels

VARIABLES	Top Novelty Model 1	Top Feasibility Model 2	Top Value Model 3	Top Quality Model 4
<i>Baseline</i> = Human				
Level 1	-0.288***	0.0374	0.0147	-0.0210

	(0.0617)	(0.0464)	(0.0345)	(0.0503)
Level 2	-0.215***	0.0445	0.0161	0.0122
	(0.0667)	(0.0471)	(0.0295)	(0.0463)
Level 3	-0.290***	-0.0807	-0.00527	-0.0949*
	(0.0658)	(0.0496)	(0.0272)	(0.0540)
Constant	0.218***	0.0989***	0.0345***	0.115***
	(0.0184)	(0.0115)	(0.00734)	(0.0146)
Evaluator FE	Y	Y	Y	Y
Problem adjective FE	Y	Y	Y	Y
Solution adjective FE	Y	Y	Y	Y
Observations	1,885	1,885	1,885	1,885
R-squared	0.041	0.025	0.015	0.015
Number of evaluators	145	145	145	145

Robust standard errors in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

### 3.1. Implications of Prompt Engineering Levels on Evaluation Scores

A closer examination of the three prompt engineering AI levels reveals distinct outcomes that shed light on the nuanced relationship between prompt complexity and solution quality. The baseline approach of level 1, while providing a direct comparison with human solutions, produced less novel but higher-value responses on average. This result underscores the AI's ability to generate valuable solutions but suggests a limitation in its capacity for novelty without additional context.

By mimicking the human crowd characteristics in level 2, we find no detectable differences between the human and AI solutions. The emphasis on individual solver characteristics might enhance the AI's ability to generate novel solutions, constraining the AI to a defined persona and thereby facilitating more focus on novelty.

In level 3, even as the integration of expert personas enhanced the contextual richness of the AI prompts, it did not translate into an increase in novelty compared to level 1 or 2. However, its higher-value ratings indicate the effectiveness of infusing industry-specific knowledge through expert persona prompts that create more targeted and potentially useful solutions. Level 3 also adds the opportunity to specify a field of interest for a more targeted solution. Notably, no significant difference across all levels regarding feasibility suggests AI's consistent ability to propose implementable solutions, irrespective of prompt complexity. These findings highlight the subtle interplay between prompt sophistication, novelty, and value in AI-generated solutions.

## 4. Natural Language Processing of Human and AI Solutions

Another advantage of human crowdsourcing and innovation contests is the ability to draw upon diverse perspectives. We use two approaches to measure the diversity of human and AI responses. First, we employ natural language processing to investigate the dissimilarity between the text responses in the human solutions, as well as the AI solutions by level. Second, we use GPT-4 to classify each solution into their industries of application to compare the diversity in industries captured by the human solvers and AI.

#### 4.1. Dissimilarity of Human and AI Solutions

We measure the dissimilarity of the solution text using pairwise cosine distances (Carlson, 2023; Park et al., 2023). The dissimilarity between solutions  $i$  and  $j$  is measured as  $d_{ij} = 1 - \frac{\vec{v}_i \cdot \vec{v}_j}{\|\vec{v}_i\| \|\vec{v}_j\|}$ , where  $\vec{v}_i$  and  $\vec{v}_j$  are the pre-processed TF-IDF (term frequency-inverse document frequency) vectorized solutions with  $d_{ij} \in [0,1]$ . We then capture a solution  $i$ 's dissimilarity by averaging the dissimilarities between all possible combinations of solutions:

$$\text{Solution Dissimilarity}_i = \frac{1}{N-1} \sum_{j=1}^{N-1} d_{ij}, \quad (2)$$

where  $N$  corresponds to the total number of solutions.

#### 4.2. Results on Dissimilarity of Human and AI Solutions

In Figure 3, we plot the distribution of dissimilarity scores by human and AI levels. Figure 3 indicates greater diversity among the human solutions than the AI-generated responses. In other words, there is greater distance between the vocabulary (i.e., words) being used among the human solvers than those of GPT-4. However, it is important to note that our dissimilarity measure captures variation in the vectorized representations of occurrences of unique words. Therefore, although human responses may be semantically more diverse from each other, it is also possible that the choice of words used to express similar ideas is different due to the diverse representation of countries among the human solvers. GPT-4, on the other hand, is trained on a curated dataset, which can have a standardizing effect on how GPT-4 describes the solutions.

In Figure 4, we examine the relationship between solution dissimilarity and novelty. Here, we observe that there is a positive correlation between the dissimilarity scores (see Section 4.1) and the evaluators' ratings on novelty, particularly for the human solutions, which have the highest correlation ( $\rho = 0.33$ ). This means that the TF-IDF dissimilarity metric picks up on the unique words or combinations of words that the human evaluators observe when rating the novelty of the solutions. Although both AI levels 1 and 3 yield solution dissimilarity measures that show a positive correlation with novelty, it is noteworthy that the AI level 2 solution indicates no correlation.<sup>3</sup> In contrast, the dissimilarity scores all exhibit a small negative correlation with the value scores, with the human solutions being the largest in magnitude ( $\rho = -0.11$ ). These contrasting correlations between the dissimilarity scores, solution novelty, and value confirm that the evaluators are judging novelty and value using different metrics. We find suggestive evidence that the human evaluators perceive solution novelty based on their degree of departure from other solutions.

Figure 3. Distribution of Solution Dissimilarity by Level (N = 234 solutions)

---

<sup>3</sup> The absence of a correlation between solution dissimilarity and novelty for level 2 might be attributed to its unique design elements. This level of AI was mainly tailored to mimic certain crowd personas or field experts. Therefore, it was programmed to generate solutions from specific job perspectives, locations, industries, and maturity stages, as outlined in the prompts. In practice, this led to a narrower range of responses compared to other levels, as these predefined personas inherently constrained the model's outputs. This factor reduced the semantic and syntactic diversity of AI level 2 solutions.

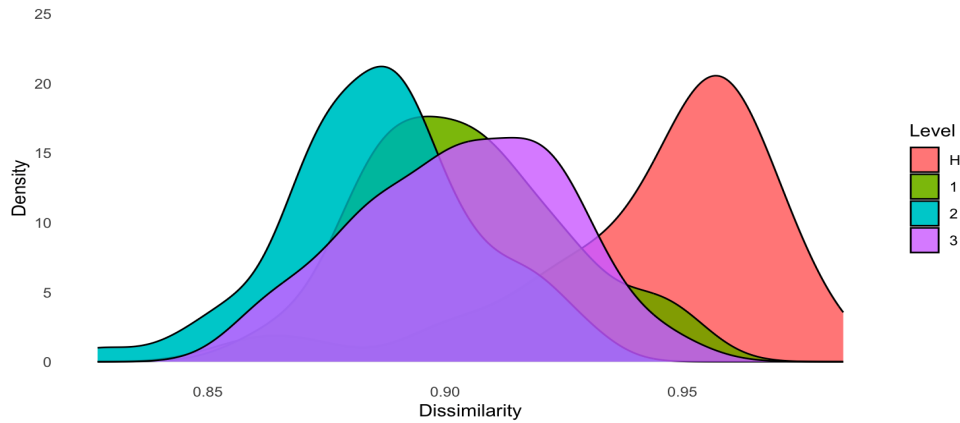
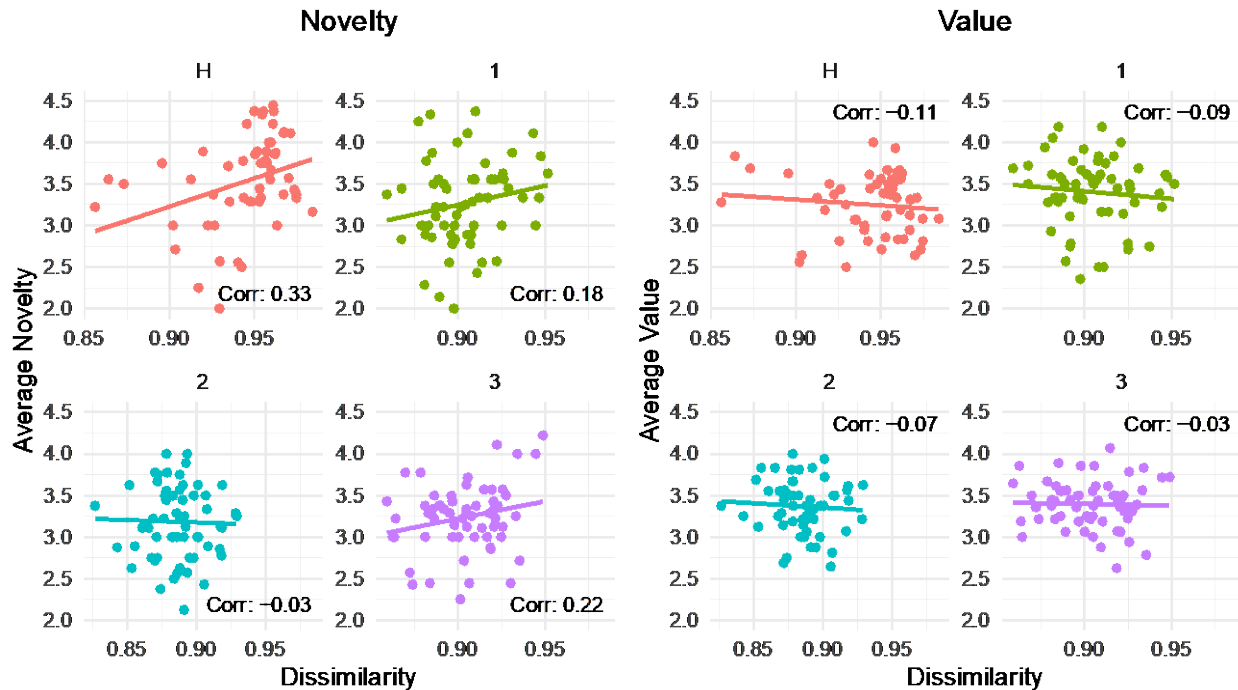


Figure 4. Scatterplot of Average Novelty and Value Scores with Solution Dissimilarity by Level (N = 234 solutions)

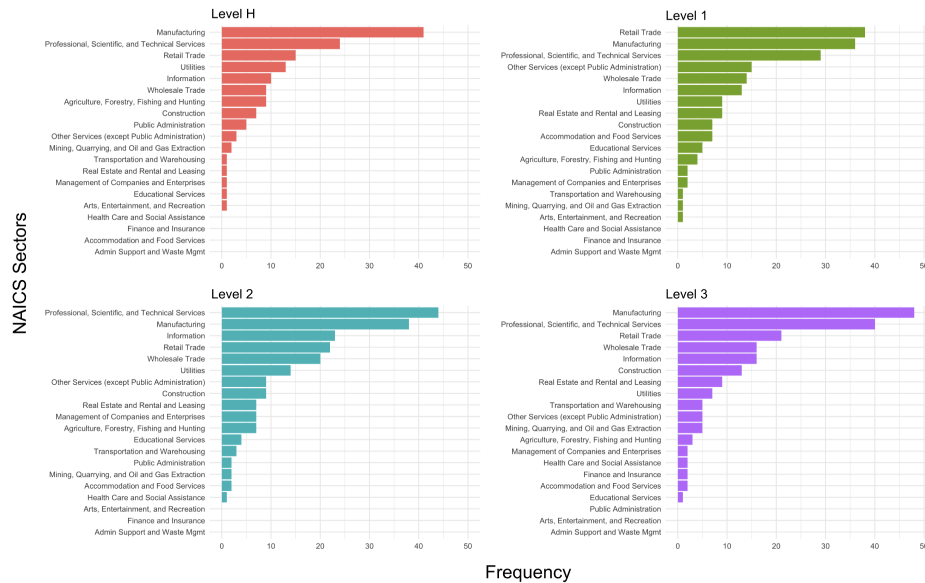


### 4.3. ChatGPT Results of Industry Classifications

Next, we use OpenAI’s GPT-4 to classify solutions into distinct industry groups. We chose the standardized 2-digit NAICS (North American Industry Classification System) sectors from 2022. For these classification tasks, we did not restrict the number of classifications per solution, meaning that a given solution could fall into one or more industries of application. Figure 5 displays the diversity of industry applications, sorted by the aggregated frequency of industry classes. Here, we see “Manufacturing” and “Professional, Scientific, and Technical Services” are consistently represented as the top two industries. This indicates that these are common bases

across ideas generated by humans and AI followed by other industries such as “Retail Trade” or “Information.”

Figure 5. Comparisons of Industry Classifications By Solution



To further assess the diversity of each level, we used the Shannon diversity index—a metric often used to measure the diversity of species in a specific space. The Shannon index  $H$ , can be normalized to the Shannon equitability index, which takes values between 0 to 1, denoted as  $E_H$ . The Shannon index within a level that contains  $k$  total sectors composing of  $p_i$  proportion of sector  $i$  is measured as:

$$E_H = \frac{-\sum_{i=1}^k p_i \cdot \log(p_i)}{\log(k)}. \quad (3)$$

The Shannon index across the human and AI levels were similar (human level = 0.793, AI level 1 = 0.820, AI level 2 = 0.824, and AI level 3 = 0.799), which suggests that both humans and AI generated similar degrees of diversity of industry applications.

## 5. Discussion and Conclusion

This study compares the problem-solving success of generative AI and human solvers to better understand AI’s capabilities of augmenting or replacing human crowdsourcing. To address this question, we partnered with ContinuumLab.AI, a venture studio focused on AI startup innovation, incubation, and investment, to develop a crowdsourcing challenge to directly compare the business ideas generated by human and AI solvers on the topic of “Creating a Sustainable Future: Unlock the Potential of a Circular Economy.”

Our findings highlight three noteworthy patterns. First, we find that generative AI has the ability to not only mirror the diverse idea-generation capabilities of human crowdsourcing but does so with minimal cost and unprecedented speed. While the human crowdsourcing challenge took five months to generate 125 viable solutions, we were able to generate the AI solutions in less



than two hours. Second, we find no noticeable difference in the average or top quality between human and AI responses, but AI crowdsourcing achieves higher evaluation scores on value than human-generated solutions. However, human problem solvers are rated as achieving higher novelty, suggesting that the ability to create new concepts, ideas, and theories may still be a uniquely human endeavor. Third, using NLP techniques, we find that human solvers demonstrate greater diversity in solution responses than AI responses: the more unique a human solution's text is from other solvers' solutions, the more likely it is to receive a higher novelty score. Interestingly, this pattern does not replicate entirely for the AI solutions, suggesting that there may be subtle differences in how novelty is interpreted for human and AI text responses. We also note that AI and humans exhibit similar ability in generating diverse industry applications to an open and unstructured circular economy challenge.

Our study primarily contributes to the literature on open innovation and crowdsourcing. When firms encounter complex issues that exceed their internal problem-solving capabilities, a common strategy is to engage in "broadcast search" by disclosing the details of the problem at hand to the public and inviting the participation of anyone who deems themselves qualified to solve the problem (Jeppesen & Lakhani, 2010). Although the use of external human solvers is a proven approach for coming up with high-quality solutions (e.g., Jeppesen & Lakhani, 2010; Liftshitz-Assaf, 2018), the findings of our study indicate that the "crowdless future" era in favor of generative AI may be found under certain conditions. Our research has illuminated the rich interplay between human creativity and AI's generative capabilities, demonstrating the effects of prompt engineering in influencing the AI's output significantly, with the AI mimicking the perspectives of a diverse crowd to create context-specific outputs. More specifically, when the objective of the problem-solving process is to generate value for the firm, our finding suggests that AI may even outperform human solvers.

Our study also contributes to the literature on radical and incremental innovation by shedding light on the potential tradeoffs between novelty and value (Amabile, 1983; Ghosh & Wu, 2021; Kaplan & Vakili, 2015; Singh & Fleming, 2010) for human and AI crowdsourcing. While humans are more likely to generate innovative responses, AI is more likely to create valuable outputs with higher environmental and financial value. As both novelty and value are critical to the creativity of crowdsourced ideas (Amabile, 2018), although generative AI may overcome some of the design, expenditure, and execution challenges that human crowdsourcing entails, we conclude it cannot completely replace human intelligence, especially when organizations prioritize novelty over all other criteria. Future research should investigate how to optimize prompt engineering to balance these crucial aspects, aiming for solutions that are not only context-specific and valuable but also novel.

Moreover, we contribute to the emerging literature on the implications of generative AI on organizational processes (e.g., Byrnjolfsson et al. 2023; Noy & Zhang, 2023; Gilardi et al., 2023; Horton, 2023). Similar to conventional crowdsourcing, AI integration also demands internal adjustments within organizations. For instance, seeking solutions from external resources, including off-the-shelf LLM models, entails internal organizing parties transitioning from being problem-solvers to managers and overseers of human and AI contributors. Importantly, similar privacy issues and ethical concerns persist and might even be amplified with the adoption of LLMs. Therefore, while using AI to derive solutions could potentially enhance efficiency, exposing proprietary and sensitive information to third-party APIs introduces additional risks and

obligations. An organization's evaluation and selection of ideas could influence the external generation of ideas, possibly leading to a "coevolutionary lock-in" that counteracts the initial diversity benefits that crowdsourcing promises (Park et al., 2022). Even with sophisticated prompting and fine-tuning techniques or the development of proprietary in-house models, organizations may find themselves caught in a loop of self-reinforcing biases and narrowing perspectives. Hence, as organizations venture into the realm of AI crowdsourcing, they must remain cognizant of these pitfalls and strive to continuously pay attention to what information can be disclosed, maintain and encourage diversity in idea generation, and manage an unbiased and systematic internal evaluation system.

Although our study offers several insights, there are, nevertheless, some empirical limitations that point to opportunities for future work. First, we focused on a specific type of problem, the circular economy, which corresponds to a broad yet relatively recent challenge for which few proven solutions exist. Hence, our findings might not be generalizable to other types of problem-solving that involve deep technicalities, abstract reasoning, artistic tastes, or simply a more specific style of problem formulation. Future work can extend our findings into different contexts, settings, and problem types. Second, despite the promises of AI crowdsourcing, the ever-evolving development of state-of-the-art models and the indeterministic nature of LLMs prompt a word of caution. The complexity behind training data and algorithms can add another layer of unpredictability, causing varying results even with identical inputs. When engaging in multi-step workflows, these variabilities can compound, potentially hindering the application of LLMs in tasks requiring high precision and reliability. Just as in human crowdsourcing, the process of AI ideation must include human supervision to avoid such issues. Lastly, in this study, we solely rely on OpenAI's GPT-4 model, but other readily available foundational and domain-specific models also exist. Hence, examining the capabilities of other LLMs to replace or augment human-generated solutions remains another promising avenue for future work.

In sum, as we advance into the age of AI, it is clear that the path toward generating innovative solutions to complex problems lies not in machines alone but in the interplay between machine-generated outputs and human expertise. The future of crowdsourcing may not rest solely on AI or human intelligence but on a robust partnership between the two, fostering an era of augmented crowdsourcing that combines the best of both worlds. Our study sheds light on a new, promising approach that has the potential to reshape the dynamics of organizational strategy and innovation.

## References

- Afuah, A., & Tucci, C. L. (2012). Crowdsourcing as a solution to distant search. *Academy of Management Review*, 37(3), 355–375.
- Allen, G., He, G., & Gadiraju, U. (2023). Power-up! What can generative models do for human computation workflows? journal title? volume etc.? or is this book with a publisher?
- Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, 45(2), 357–377.
- Amabile, T. M. (1996). Creativity and innovation in organizations. journal title? volume etc.? or is this book with a publisher?
- Amabile, T. M. (2018). *Creativity in Context: Update to the Social Psychology of Creativity*. Routledge.

- Baldwin, C., & von Hippel, E. (2011). Modeling a paradigm shift: From producer innovation to user and open collaborative innovation. *Organization Science*, 22(6), 1399–1417.
- Bennett, J., & Lanning, S. (2007). The Netflix Prize. *Proceedings of KDD Cup and Workshop*, 3–6. (this confirms the pagination: <https://www.cs.uic.edu/~liub/KDD-cup-2007/proceedings/The-Netflix-Prize-Bennett.pdf>.)
- Berg, J. M. (2016). Balancing on the creative highwire: Forecasting the success of novel ideas in organizations. *Administrative Science Quarterly*, 61(3), 433–468.
- Berg, J. M., Raj, M., & Seamans, R. (in press). Capturing value from artificial intelligence. *Academy of Management Discoveries*.
- Brynjolfsson, E., Li, D., & Raymond, L. (2023). Generative AI at work. National Bureau of Economic Research Working Paper W31161. <https://doi.org/10.3386/w31161>
- Boudreau, K. J., & Lakhani, K. R. (2015). “Open” disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology. *Research Policy*, 44(1), 4–19.
- Brown, Y. et al. <https://arxiv.org/pdf/2005.14165.pdf>.
- Catalini, C., & Tucker, C. (2016). Seeding the S-Curve? The role of early adopters in diffusion. National Bureau of Economic Research Working Paper Series, no. 22596. <https://doi.org/10.3386/w22596>.
- Cheng, Z., Lee, D., & Tambe, P. (2022). InnoVAE: Generative AI for understanding patents and innovation. Available at <http://dx.doi.org/10.2139/ssrn.3868599>.
- Chesbrough, H. W. (2003). *Open innovation: The New Imperative for Creating and Profiting from Technology*. Harvard Business Press.
- Dahan, E., & Mendelson, H. (2001). An extreme-value model of concept testing. *Management Science*, 47(1), 102–116.
- Dahlander, L., Jeppesen, L. B., & Piezunka, H. (2019). How organizations manage crowds: Define, broadcast, attract, and select. *Managing Inter-Organizational Collaborations: Process Views (Research in the Sociology of Organizations, Vol. 64)*, Emerald Publishing Limited, 239–270.
- Dahlander, L., & Piezunka, H. (2020). Why crowdsourcing fails. *Journal of Organization Design*, 9, 1–9.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *ArXiv Preprint:1810.04805*.
- Dewar, R. D., & Dutton, J. E. (1986). The adoption of radical and incremental innovations: An empirical analysis. *Management Science*, 32(11), 1422–1433.
- Fernandes, R., & Simon, H. (1999). A study of how individuals solve complex and ill-structured problems. *Policy Sciences*, 32(3), 225–245.
- Ghosh, S., & Wu, A. (2021). Iterative coordination and innovation: Prioritizing value over novelty. *Organization Science*. volume etc.?
- Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30).
- Girotra, K., Meincke, L., Terwiesch, C., & Ulrich, K. T. (2023). Ideas are dimes a dozen: Large language models for idea generation in innovation. Available at <https://ssrn.com/abstract=4526071>.
- Girotra, K., Terwiesch, C., & Ulrich, K. T. (2010). Idea generation and the quality of the best idea. *Management Science*, 56(4), 591–605. <https://doi.org/10.1287/mnsc.1090.1144>.
- Gruber, M., MacMillan, I. C., & Thompson, J. D. (2008). Look before you leap: Market opportunity identification in emerging technology firms. *Management Science*, 54(9), 1652–1665.
- Haase, J., & Hanel, P. H. P. (2023). Artificial muses: Generative artificial intelligence chatbots have risen to human-level creativity. <https://arxiv.org/abs/2303.12003>.

- Horton, J. J. (2023). Large language models as simulated economic agents: What can we learn from Homo Silicus? National Bureau of Economic Research Working Paper Series, no. 31122. <https://doi.org/10.3386/w31122>.
- Jeppesen, L. B., & Lakhani, K. R. (2010). Marginality and problem-solving effectiveness in broadcast search. *Organization Science*, 21(5), 1016–1033.
- Kaplan, S., & Vakili, K. (2015). The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal*, 36(10), 1435–1457.
- Koput, K. W. (1997). A chaotic model of innovative search: Some answers, many questions. *Organization Science*, 8(5), 528–542.
- Lane, J., Ganguli, I., Gaule, P., Guinan, E., & Lakhani, K. R. (2021). Engineering serendipity: When does knowledge sharing lead to knowledge production? *Strategic Management Journal*, 42(6), 1215–1244.
- Laursen, K., & Salter, A. (2006). Open for innovation: The role of openness in explaining innovation performance among U.K. manufacturing firms. *Strategic Management Journal*, 27(2), 131–150.
- Lifshitz-Assaf, H. (2018). Dismantling knowledge boundaries at NASA: The critical role of professional identity in open innovation. *Administrative Science Quarterly*, 63(4), 746–782.
- Lovett, M. C., & Anderson, J. R. (1996). History of success and current context in problem solving: Combined influences on operator selection. *Cognitive Psychology*, 31(2), 168–217.
- Marengo, L., Dosi, G., Legrenzi, P., & Pasquali, C. (2000). The structure of problem-solving knowledge and the structure of organizations. *Industrial and Corporate Change*, 9(4), 757–788.
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187–192.
- Page, S. (2008). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*, rev. ed. Princeton University Press.
- Park, S., Piezunka, H., & Dahlander, L. (2023). Coevolutionary lock-in in external search. *Academy of Management Journal*.
- Piezunka, H., & Dahlander, L. (2015). Distant search, narrow attention: How crowding alters organizations' filtering of suggestions in crowdsourcing. *Academy of Management Journal*, 58(3), 856–880.
- Rindova, V. P., & Petkova, A. P. (2007). When is a new thing a good thing? Technological change, product form design, and perceptions of value for product innovations. *Organization Science*, 18(2), 217–232.
- Shane, S. (2000). Prior knowledge and the discovery of entrepreneurial opportunities. *Organization Science*, 11(4), 367–472.
- Singh, J., & Fleming, L. (2010). Lone inventors as sources of breakthroughs: Myth or reality? *Management Science*, 56(1), 41–56.
- Sørensen, J. B., & Stuart, T. E. (2000). Aging, obsolescence, and organizational innovation. *Administrative Science Quarterly*, 45(1), 81–112.
- Terwiesch, C., & Loch, C. H. (2004). Collaborative prototyping and the pricing of custom-designed products. *Management Science*, 50(2), 145–158.
- Terwiesch, C., & Ulrich, K. T. (2009). *Innovation Tournaments: Creating and Selecting exceptional Opportunities*. Harvard Business Press.
- Von Hippel, E. (1986). Lead users: A source of novel product concepts. *Management Science*, 32(7), 791–805.
- Von Hippel, E. (2005). Open source software projects as user innovation networks. In J. Feller, B. Fitzgerald, S. A. Hissam, K. R. Lakhani (eds.), *Perspectives on Free and Open Source Software*, 267–278. MIT Press.
- Von Hippel, E. (2006). *Democratizing Innovation*. MIT Press.

- Von Hippel, E. & von Krogh, G. (2003). Open source software and the “private-collective” innovation model: Issues for organization science. *Organization Science*, 14(2), 209–223.
- Von Hippel, E., & von Krogh, G. (2016). Crossroads—Identifying viable “need–solution pairs”: Problem solving without problem formulation. *Organization Science*, 27(1), 207–221.
- Von Krogh, G., & von Hippel, E. (2006). The promise of research on open source software. *Management Science*, 52(7), 975–983.

## Appendix A: Detailed Explanation of Prompt-Building Mechanisms and Prompts Used

### A.1 Detailed Explanation of Prompt-Building Mechanism

We elaborate on the specific mechanism used to construct the prompts for the AI-generated solutions in the study. Our implementation uses the Python programming language on GoogleColab and leverages OpenAI’s GPT-4 model. We interact with the GPT-4 model using OpenAI’s ChatCompletion API, which is designed for conversational tasks and allows for multi-turn exchanges with the model by including a series of structured messages as inputs.

Each message included in the API request is categorized by role and content. The “role” attribute is assigned as either “system” or “user.” The “system” role provides high-level instructions or context for the conversation, while the “user” role prompts the model to generate specific outputs based on the given task.

The specific Python function utilized for the API request was `openai.ChatCompletion.create()`, which accepts two parameters: `model` and `messages`. The “model” parameter specifies the AI model being used, which in this case was set to “gpt-4.” The “messages” parameter is a list of structured messages to be delivered to the model. Each message in the list is a dictionary containing two keys: “role” and “content.”

In our study, the content of the “system” message was set to the “context,” a string that provides the general context of the problem to be solved exactly as it was shared with humans, potentially augmented with additional information such as solver characteristics or persona details based on the prompt level. The “user” message’s content was set to a specific “content” string, which contains the template to be used for the answer.

The function call in our code is thus:

```
response = openai.ChatCompletion.create(
    model="gpt-4",
    messages=[
        {"role": "system", "content": default_context},
        {"role": "user", "content": content}
    ])
```

Upon execution, the function returns a “response,” which contains the AI-generated solution. This solution is then used for further evaluation and comparison in the context of our study. This iterative process of prompt creation and AI response generation was carried out 750 times to match the 125 human-generated solutions for each level and configuration.

## **A.2 Modifying Prompt-Qualifying Adjectives to Match Distribution of Human-Generated Answers**

Given the diversity and variation inherent in human-generated responses, we aimed to replicate this natural dispersion by slightly adjusting each ChatGPT prompt. Specifically, we incorporated a range of qualifying adjectives indicative of the expected answer length in the prompts.

We defined a set of adjectives: “highly detailed and elaborate,” “succinct,” “brief,” “concise,” “short,” “comprehensive,” “long,” “5-paragraph,” “3-paragraph,” “medium-length,” “very precise and elaborate,” “20-sentence.” The Python code iteratively selected adjectives from these sets in a predefined random sequence and injected them into the prompts.

By employing a variety of qualifying adjectives for the expected length of the problem and solution, we generated a range of AI responses that mirrored the distribution of human-generated solutions more closely. This enhancement further refined our experiment, offering a richer comparison of the capabilities between human-generated and AI-generated solutions.

## **A.3 Prompts used**

We provide the prompts we used for each level below:

### **[Level 1M-2M-3M]**

#### **Context:**

We are excited to announce an opportunity for freelancers to collaborate with researchers at the Digital, Data, and Design Institute 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.

Here is an example: Car Sharing in order to reduce the carbon footprint associated with driving.

Submit your real-life use cases on how companies can implement the circular economy in their businesses. New ideas are also welcome, even if they are 'moonshots'. Your suggestions will help Harvard researchers understand the impact of the circular economy on business. Let's get creative and revolutionize the world through the circular economy!

Your goal is to win the top monetary prizes. Judges will use the following evaluation criteria:

- \* Novelty (How different is it from existing solutions?)
- \* Environmental Impact (How much does it benefit the planet?)
- \* Financial Impact (What financial value can it create for businesses?)
- \* Feasibility and Scalability of Implementation (How likely is it to succeed and how scalable is it?)

**[Level 2M adds the following]**

To answer the question, you will take the perspective of the following persona:

You are a [Job Title] located in [Continent]. You propose a solution that applies to [Industry of Solution]. The maturity of your solution is [Maturity].

Examples of such persona (not included in the prompt):

- You are a Executives, Managers, and Entrepreneurs, located in Africa. You propose a solution that applies to Food, Beverages, Packaging and Waste Management. The maturity of your solution is Proof of Concept.
- You are a Technical and Creative Professionals, located in Europe. You propose a solution that applies to Transportation and Logistics. The maturity of your solution is Ideation.

**[Level 3M adds the following]**

To answer the question, you will take the perspective of the following persona:

[Expert Name] who has expertise in [Expert Field].

**Prompt:**

Answer the following two questions to propose a circular economy idea that could win the challenge according to the evaluation criteria.

Problem: Tell us about the problem your solution is meant to solve.

Solution: Describe the solution in your own words.

Use the following template to answer:

**[Level 1M]**

Problem: <Write a high quality, *ADJECTIVE1* answer.>

Solution: <Write a high quality, *ADJECTIVE2* solution.>

**[Level 2M-3M]**

Problem: <Write a high quality, *ADJECTIVE1* answer, **corresponding to the personality, inspiration, and knowledge of your persona.**>

Solution: <Write a high quality, *ADJECTIVE2* solution, **corresponding to the personality, inspiration, and knowledge of your persona.**>

**Context for Level 1S-2S-3S:**

We are excited to announce an opportunity for freelancers to collaborate with researchers at the Digital, Data, and Design Institute 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.

Here is an example: Car Sharing in order to reduce the carbon footprint associated with driving.

Submit your real-life use cases on how companies can implement the circular economy in their businesses. New ideas are also welcome, even if they are ‘moonshots’. Your suggestions will help Harvard researchers understand the impact of the circular economy on business. Let’s get creative and revolutionize the world through the circular economy!

Your goal is to win the top monetary prizes by satisfying to the maximum the following evaluation criteria.

- \* Novelty (How different is it from existing solutions?)
- \* Environmental Impact (How much does it benefit the planet?)
- \* Financial Impact (What financial value can it create for businesses?)
- \* Feasibility and Scalability of Implementation (How likely is it to succeed and how scalable is it?)

Answer the following two questions to propose a circular economy idea that could win the challenge according to the evaluation criteria.

Problem: Tell us about the problem your solution is meant to solve.

Solution: Describe the solution in your own words.

Use the following template to answer:

**[Level 1S]**



Problem: <Write a high quality answer.>

Solution: <Write a high quality solution.>

We will ask to answer these questions several times, and make sure each new answer tackles a different problem than the previous ones and proposes a different solution.

### **[Level 2S-3S]**

Problem: <Write a high quality answer, corresponding to the personality, inspiration, and knowledge of your persona.>

Solution: <Write a high quality solution, corresponding to the personality, inspiration, and knowledge of your persona.>

We will give you a series of different personas, and make sure each new answer from each persona tackles a different problem than the previous ones and proposes a different solution. Make sure each answer to correspond to the perspective, characteristics, and knowledge of your persona.

### **Prompt:**

Give a new high quality, *ADJECTIVE1* Problem and high quality, *ADJECTIVE2* Solution. Make sure to tackle a different problem than the previous ones and propose a different solution. Make also sure your answers satisfy the evaluation criteria (novelty, environmental impact, financial impact, feasibility and scalability).

### **[Level 2S adds the following]**

The persona you embody for this answer:

You are a [Job Title] located in [Continent]. You propose a solution that applies to [Industry of Solution]. The maturity of your solution is [Maturity].

### **[Level 3S adds the following]**

The persona you embody for this answer:

[Expert Name] who has expertise in [Expert Field].

## **A.4 Expert personas from the randomly sampled Level 3 answers evaluated**

Level 3M:

(Mette Hay, Home Furnishings), (Hubertus Muehlhaeuser, Electrical Equipment), (Ramon Laguarta, Containers & Packaging), (Kenichiro Yoshida, Consumer Electronics), (David Steiner, Waste Management), (Michael Green, Building Products), (William L. McComb, Household Appliances), (Howard Schultz, Food), (David Abney, Transportation & Logistics), (Lisa P. Jackson, Software & IT Services), (Dave Lennard, Building Products), (Simon Segars, Electrical Equipment), (Stephen Kieran, Buildings), (Akio Toyoda, Automobiles & Tires), (Jensen Huang, Technology/ Hardware Products), (Tom Linebarger, Construction Machinery), (Emmanuel Faber, Food), (Catherine Howarth, Financials), (James Timberlake, Buildings), (Ren Zhengfei,

Consumer Electronics), (Alex Gorsky, Health Care Products), (Stella McCartney, Apparel & Textiles), (Paul Polman, Forest Products), (Elon Musk, Automobiles & Tires), (Bill Browning, Construction Materials), (Andrew Martin, Home Furnishings), (Tim Cook, Technology/ Hardware Products), (Alex Keith, Cosmetics), (David Greenberg, Cosmetics), (Jean-Paul Agon, Cosmetics)

#### Level 3S:

(Satya Nadella, Software & IT Services), (Kenichiro Yoshida, Consumer Electronics), (Andrew Liveris, Construction Materials), (Akio Toyoda, Automobiles & Tires), (Mark Eames, Metals: Beyond Big 4), (Michael Dell, Technology/ Hardware Products), (Larry Fink, Financials), (Lisa Su, Technology/ Hardware Products), (Catherine Howarth, Financials), (Tim Cook, Technology/ Hardware Products), (Pat Gelsinger, Technology/ Hardware Products), (Richard Adkerson, Metals: Beyond Big 4), (Ramon Laguarta, Containers & Packaging), (Paul Polman, Forest Products), (Tom Linebarger, Construction Machinery), (Thomas Rau, Home Furnishings), (Uday Yadav, Electrical Equipment), (Leif Johansson, Health Care Products), (Berry Wiersum, Forest Products), (Jan Jenisch, Construction Materials), (Marc Benioff, Software & IT Services), (Stephen Kieran, Buildings), (John Hayes, Containers & Packaging), (Rick Fedrizzi, Buildings), (Mark Bitzer, Household Appliances), (Ivan Glasenberg, Metals: Beyond Big 4), (John Elkington, Construction Materials), (Lance Fritz, Transportation & Logistics), (Denise Morrison, Food), (Bill Browning, Construction Materials)

### **A.5 Initial Formulation of the Challenge for Humans**

(after removing the administrative details to participate in the challenge)

We are excited to announce an opportunity for freelancers to collaborate with researchers at the Digital, Data, and Design Institute (D<sup>3</sup>) (<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.

Here is an example: Car Sharing in order to reduce the carbon footprint associated with driving.

Here is more information on circular economy:

<https://ellenmacarthurfoundation.org/topics/circular-economy-introduction/overview>

Submit your real-life use cases on how companies can implement the circular economy in their businesses. New ideas are also welcome, even if they are 'moonshots'. Your suggestions will help Harvard researchers understand the impact of the circular economy on business. Let's get creative and revolutionize the world through the circular economy!

Our team will be evaluating your entries using the following criteria:

- \* Novelty (How different is it from existing solutions?)
- \* Environmental Impact (How much does it benefit the planet?)
- \* Financial Impact (What financial value can it create for businesses?)
- \* Feasibility and Scalability of Implementation (How likely is it to succeed and how scalable is it?)

The best overall solution will receive a \$1,000 prize and be presented with an official trophy from the D<sup>3</sup> Institute at Harvard.

Top 500 best solutions will receive a cash prize of \$10 each. The same freelancer could potentially win multiple prizes if they had submitted more than one winning entries! Apart from the cash prize, the winning entries will receive a letter from a Harvard faculty member and an official certificate from Harvard D<sup>3</sup> Lab.

Some creators of the best solutions will also receive an invitation to participate in an exclusive two-day virtual Circular Economy Catalyst Event at Harvard Business School in April 2023 to learn how leading companies, startups, and investors are engaging in the circular economy.

## Appendix B: Survey Materials

### B.1 Sample Solutions by Level

Level	Problem	Solution
H	<p>These days the production of clothing and selling of clothes is unsustainable and needs transformation. However, many of the changes are already underway, and soon there will be a change in our environment, economies, and closets. As there is an increase in environmental awareness among governments, industries, brands, and consumers, the need for sustainability throughout the fashion sector is paramount. Most importantly, accounting for 10% of global emissions, the fashion industry is</p>	<p>Fast fashion is our current clothing reality that is a business model based on low manufacturing costs and frequent consumer purchases of new items. Multiple social and environmental costs are not considered in consumer price tags. Hence, this makes the model unsustainable.</p> <p>Improving demand forecasting is one of the most important and key to reducing up to 40% of fabrics. These are the ones that never make it to end customers. Some companies use cloud-based supply chain management software to keep inventory levels low. They also automatically reorder those which sell well. Most companies can save money by lowering unsold inventory, with environmental factors becoming a cost-saving driver.</p> <p>Apart from that, prominent industry players now offer repair services, adapting models for a sustainable future. To replace revenue lost from fewer repeat purchases, retailers embracing repair will create new revenue streams, and this service can become a brand differentiator. According to a recent study, one can</p>

	<p>increasingly under the spotlight to reduce its carbon footprint and become more circular. The circular economy aims to eliminate waste and reuse and recycle materials. It can regenerate nature to encourage a more prosperous future for our planet.</p>	<p>change the thread with which clothes are sewn to increase garment recyclability to 90% and cut the carbon footprint by 50%.  One of the most challenging pieces of clothing to recycle is jeans. Through laser technology and biodegradable threads, one can create stone-wash patterns. Similarly, buttons are easy to unscrew by recyclers and planet-friendly dyes. The climate-conscious pair of jeans will have the same classic look.</p> <p>Throughout the fashion industry, buy-back schemes are becoming increasingly popular. The typical model entails returning pre-worn items to the original brand for store credit. The brand can repair, resell or recycle the garment. Moreover, it is also extending the product lifecycle. While reducing waste, buy-back schemes offer value to the brand and consumer.</p> <p>The circular economy economic value is obvious, and in terms of monetary value, it is estimated that the world economy could be improved by \$192 billion by 2030. It is possible if the fashion industry addresses the problems presented by its current linear economy</p>
<p>H</p>	<p>The problem this solution is meant to solve is the difficulty in managing and tracking the lifecycle of software products and IT services in the context of the circular economy. In traditional linear models, software products and services are often used for a limited time and discarded, leading to large amounts of electronic waste. This leads to a need for a more sustainable approach to software and IT services that maximize the use of resources and minimize waste. The fast-paced and rapidly evolving nature of the software and IT industries and the limited transparency and traceability of software products and services also</p>	<p>The solution to the problem of managing the end-of-life of software and IT services in the context of the circular economy is to shift from a linear take-make-dispose model to a more sustainable, closed-loop system. This can be achieved through a variety of approaches, such as designing software products for reuse and recyclability, implementing circular procurement practices, and developing closed-loop systems for the disposal of electronic waste.</p> <p>Innovative business models and technology solutions can also play a critical role in supporting the circular economy in the software and IT services industry. This could include the development of new software platforms that are designed for reuse and recycling, the creation of certification and labeling schemes to promote sustainable practices, and the implementation of circular procurement processes that incentivize the use of circular solutions.</p> <p>By addressing the challenges of the linear model and promoting the use of circular practices, the software and IT services industry can contribute to a more sustainable and resilient future for businesses, communities, and the planet as a whole</p>

	make it challenging to implement circular economy practices in a way that is economically viable and technically feasible.	
1	<p>The construction industry contributes significantly to the global waste generated each year. This comes from both the construction of new buildings or infrastructure and the demolition or renovation of existing ones. It is estimated that this industry is responsible for 40% of the world's total waste. This waste primarily consists of concrete, metal, glass, brick, and wood, among other materials. The disposal of these materials into landfills leads to growing environmental concerns such as soil and water pollution and greenhouse gas emissions.</p>	<p>To tackle this, we propose launching a digital marketplace platform for the exchange of construction materials. The platform would connect builders, contractors, or individuals involved in construction and demolition projects who have a surplus of materials, with those in need of these materials. This will promote the reuse and recycling of construction materials, reducing their disposal in landfills, and the demand for new materials, thus minimizing waste production and the industry's overall environmental impact. Additionally, it will cut costs by creating a new revenue stream for selling unneeded materials and reducing the expenditure on new materials. Potential challenges towards implementation involve regulation and standardization of the materials and overcoming resistance to change within the industry. However, with growing environmental awareness and regulation, the adaptation to such platforms is increasing, offering scalability and feasibility for this circular economy solution.</p>
1	<p>The fashion industry today is the second most polluting industry in the world as there is a significant amount of waste produced throughout its supply chain, from production to disposal. The problem goes beyond environmental concerns as it also leads to a significant moral and economic waste.</p>	<p>We propose a solution of a closed-loop system, used in the fashion industry that includes textile recycling and clothes rental services. Firstly, textile recycling: Brands should partner with recycling organizations to turn used clothes into new clothes. The idea is to turn pre-loved clothes into the raw material for new products, creating less demand for new raw materials and thus, reducing the fashion industry's overall carbon footprint. Secondly, clothes rental services: Brands should also offer the option to rent clothes. Many people only wear a certain outfit a few times before throwing it away or leaving it unused in their wardrobe. By offering rental services, fewer clothes will need to be produced annually, significantly reducing waste. This is also financially beneficial to the business, as a single piece of clothing can create recurring revenue whereas selling it only brings a one-time profit. This solution is not only favorable for the environment by reducing waste and emissions but also helps businesses cut down costs on</p>

		production and increase profits.
2	Rapid urbanization and consumerism in Asian markets have led to an unnecessary increase in waste generation from different sectors such as Apparel, Electronics, Consumer Goods, Automobiles, Industrial Equipment, Metal Products, Buildings and Construction. Tremendous amounts of resources are spent on creating goods, only for them to end up in landfills once their perceived usefulness diminishes for the consumer. This linear consumer model of make, use, and dispose is not sustainable in the long run, and it is adversely affecting both environmental and economic aspects of societies.	<p>To address this problem, I propose a multi-tier Global-Marketplace based on the Circular Economy model (GMC). Built on a digital platform, the GMC will connect manufacturers, suppliers, consumers, and recycling industries across various sectors - Apparel, Electronics, etc. The platform aims to extend the lifecycle of products by promoting sustainable practices like sharing, leasing, reusing, and recycling. The Concept: GMC would have a dual functionality. A marketplace for leasing, sharing, and reselling of used goods. For example, a consumer can lease an electronic device or an automobile when needed instead of buying it. Here, apart from consumers, manufacturers will also take the responsibility to facilitate leasing or sharing by offering "Product-as-a-service." A platform for channelizing end-of-life goods towards recycling industries. End-of-life goods would be collected, refurbished, or disassembled to recycle and retrieve valuable materials, which can be supplied back to the manufacturers. The Implementation Strategy:</p> <ul style="list-style-type: none"> <li>-To ensure large-scale adoption, businesses will be incentivized based on their Environmental, Social, and Corporate Governance (ESG) score, which will increase with their level of involvement in the circular economy practices.</li> <li>-The marketplace will prioritize displaying products from businesses with a high ESG score to promote their products and services.</li> <li>-The digital platform will be powered by blockchain technology to ensure the traceability of the products/services and transparent transactions.</li> <li>-Appropriate use of AI technology can help in predicting the end-of-life of goods and therefore, plan the strategies for their effective recycling.</li> </ul> <p>Impact: This approach will break the linear production and consumption model. The environmental impact would be significant in terms of waste reduction, efficient resource utilization, and decreased carbon emissions. For businesses, this model opens up new revenue streams in the form of leasing, reselling, providing repair services, and retrieving valuable materials from disposed goods. This model is scalable, which can involve more sectors, and also feasible as it is based on digital technologies, which are rapidly advancing and widely accepted. This will also offer job opportunities in sectors like refurbishing and repair.</p>
2	In waste management, the	I envision a circular economy solution termed "Food

	<p>problem that strikes me the most and begs for an urgent solution is the excessive amount of food waste that businesses produce. This food waste not only contributes to a substantial portion of solid waste occupying landfill spaces, but also generates methane, a potent greenhouse gas. Additionally, it is a tragic misplacement of resources, given the large population of undernourished people globally.</p>	<p>Waste to Energy". This involves setting up anaerobic digestion systems in businesses or communities that produce significant amount of food waste such as hotels, supermarkets, and dining halls. These systems convert the organic waste into biogas that can be harnessed for energy. The residual, nutrient-rich, harmless substance can be utilized as compost for agriculture, in place of chemically synthetic fertilizers. This solution is not just environmentally beneficial but also creates financial value for businesses by reducing their waste management costs, offering an in-house sustainable energy source, and potentially generating revenue through selling the compost. The implementation of this system is quite feasible with existing technology and scalable across various sizes of businesses and communities. If championed by businesses, it could significantly change their environmental footprint and prove to be a responsible and profitable step towards a sustainable circular economy.</p>
<p>3</p>	<p>Within the Technology &amp; Hardware Products sector, one critical environmental concern is the lifecycle management of electronic components, particularly those containing rare and precious metals. These components, used in a broad array of devices like smartphones, laptops, or servers, are often not reclaimed or recycled efficiently at the end of their life. This linear handling leads to increased electronic waste (e-waste) accumulation and lays waste to the valuable resources embedded in these components. Also, the continual extraction of these metals for new components often entails high environmental impacts and resource depletion. A transition to a more circular approach,</p>	<p>I propose the "Circular Component Campaign" based on "Component Leasing," "Modular Device Design," and "Dedicated E-Waste Handling Programs." Firstly, "Component Leasing" offers an innovative business model alteration. Instead of selling components outright, customers could lease them for a designated period. Once the components reach their end-of-life or need an upgrade, they could be returned to the company, ensuring that the valuable materials they contain are efficiently recovered instead of being discarded. Secondly, for leasing to work effectively, we need to work closely with device manufacturers to incorporate "Modular Device Design." Devices built with an easy-to-disassemble design would facilitate the swift replacement or upgrading of individual components, thereby enhancing their ability to be leased and later recovered. Thirdly, to effectively manage these returned components and any other e-waste our products generate, we need "Dedicated E-Waste Handling Programs." By setting up collection centers and employing state-of-the-art recycling technologies, we can ensure that not only the precious metals but as much material as possible is reclaimed from the devices. The "Circular Component Campaign" can significantly reduce e-waste generation, curb resources depletion, and reduce the environmental footprint of our products. Financially speaking, the recovered metals present a valuable secondary resource stream, likely leading to cost savings. As for feasibility, although</p>

	<p>focusing on efficient, end-of-life recovery and reuse of these components, is hence of crucial importance.</p>	<p>challenging, it builds upon existing practices such as product leasing and e-waste handling frameworks, with room for improvement and expansion. And considering the ubiquity of electronic devices and components, the potential for scalability is vast. Hence, the "Circular Component Campaign" offers a solid blueprint for a transition towards the circular economy in the Technology &amp; Hardware Products industry.</p>
3	<p>A significant environmental burden within the Containers &amp; Packaging industry lies in the over-reliance on single-use packing; this includes non-recyclable plastic packaging broadly used in various sectors, specifically FMCG, perishables, and e-commerce. Each year, billions of such packages end up in landfills or oceans, contributing to pollution and harming wildlife. Moreover, the production of this packaging is resource and energy-intensive and contributes substantially to carbon emissions. Therefore, the urgent need to transition to more sustainable packaging solutions within the industry is quite evident.</p>	<p>I propose a multi-pronged "Packaging Revolution" strategy, consisting of 'Design for Recyclability,' 'Alternate Material Integration,' and 'Return &amp; Reuse Schemes'. 'Design for Recyclability' involves creating packaging that is easy to recycle. This principle requires fundamental changes in the material selection, design, labelling, and manufacturing process to ensure that the packaging can be effectively sorted and recycled in existing infrastructure. 'Alternate Material Integration' explores the use of materials that have a lower environmental impact, like bioplastics, paper, or even edible materials where appropriate. Investment in R&amp;D is required to develop these alternative materials to offer similar functionality as traditional plastic packaging. Lastly, 'Return &amp; Reuse Schemes' encourages consumers to return the packaging after use. Leveraging deposit schemes, reverse vending, or home pickups could encourage more consumers to return packaging, which can then be cleaned and reused, thereby extending the life of packaging material. The 'Packaging Revolution' strategy could significantly reduce garbage generation, raw material extraction, energy consumption, and the associated environmental impacts. From a financial perspective, tapping into the growing preference for sustainable packaging could boost market share, and reusing packaging could bring cost savings over time. With recycling technologies improving and consumers' increasing willingness to participate in return schemes, the strategy is viable. Considering the ubiquity of packaging in almost every sector, these practices can be universally implemented, propelling the Containers &amp; Packaging industry towards a circular economy.</p>

**B.2 Screening Survey Materials**

Q0 Welcome to this 5-minute screening survey. We will ask you a few questions about yourself and your domain knowledge in the circular economy. You will be paid \$1 for completion of the screening.



Depending on your responses, we will determine your eligibility to participate in the follow-up evaluation task.

First, what is your Prolific ID?

---

Q1 Where are you located?

- United States (1)
- Outside United States (2)

Q2 How old are you?

- under 18 (1)
- 18-24 (2)
- 25-34 (3)
- 35-44 (4)
- 45-54 (5)
- 55-64 (6)
- 65 or older (7)

End of Block: Welcome

Start of Block: Screening1 - self-claimed interest + expertise

Q3 How interested are you in the problem of circular economy?

	1 (little to no interest) (1)	2 (2)	3 (moderate interest) (3)	4 (4)	5 (very much interest) (5)
Interest (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q4 List all industries you have previously worked (outside of educational experience)?

	Have you worked in this field?		If yes, for how many years?
	Yes (1)	No (2)	
Apparel & Textiles (1)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Automobiles & Tires (2)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Beverages (3)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Building Products (4)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Buildings (5)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))

Construction Machinery (6)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Construction Materials (7)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Consumer Electronics (8)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Containers & Packaging (9)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Cosmetics (10)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Electrical Equipment (11)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Financials (13)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Food (14)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Forest Products (15)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Health Care Products (16)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Home Furnishings (17)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Household Appliances (18)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))

Industrial Machinery (19)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Metals: beyond big 4 (20)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Software & IT Services (21)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Technology/ Hardware Products (22)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Transportation & Logistics (23)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Waste Management (24)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))
Other (25)	<input type="radio"/>	<input type="radio"/>	▼ less than 1 year (1 ... > 20 years (6))

End of Block: Screening2 - work exp

Start of Block: Screening3 - skills

Q5 Welcome to the skills test! Below you will find 5 multiple choice questions related to your claimed area of expertise. Please answer them carefully. *To avoid plagiarism, we will be grading not only based on your accuracy but also on your time of completion.*

Which of the following principles is NOT associated with a circular economy?

- Waste as a resource (1)
- System effectiveness (2)
- Long-term usage (3)

- Linear consumption (4)
- User of renewable energy (5)

Q6 The circular economy envisions waste as:

- A necessary byproduct of production (1)
- A resource that should be minimized (2)
- A cost to be managed and reduced (3)
- An unavoidable aspect of human activity (4)
- An indicator of inefficient resource use (5)

Q7 In the circular economy, what does “product as a service” mean?

- Renting out products as services (1)
- Transforming products into services (2)
- Charging for the service a product provides, rather than the product itself (3)
- Offering complimentary services with the product (4)
- Selling services instead of products (5)

Q8 What is one potential challenge of transitioning to a circular economy?

- Initial investment cost (1)
- Increased product durability (2)
- Reduced reliance on non-renewable resources (3)
- Reduction in waste production (4)
- Increased use of renewable energy (5)

Q9 Which of the following industries has commonly adopted the circular economy model?

- Fossil fuel energy production (1)
- Single-use plastic manufacturing (2)
- Furniture manufacturing (3)
- Fast-fashion clothing (4)
- Lead-acid battery production (5)

End of Block: Screening3 - skills

Start of Block: Congrats

Q10 Congratulations! You have been selected to participate in the evaluation task.

We expect the follow-up evaluation survey to take ~30 minutes. We will first ask you some demographic information about yourself, followed by evaluations of 13 solutions provided for a large research institution to understand the impact of circular economy on business.

We will pay you \$12 for your time and effort. Additionally, you will have the opportunity to receive up to \$13 in bonuses depending on your performance, for a maximum compensation of \$25.

Are you willing to participate in the follow-up evaluation task?

Yes (1)

No (2)

*Display This Question:*

*If Q10 = Yes.*

Q11-0 Thanks for your interest! The follow-up evaluation survey will be active on Prolific soon. It will pop up in your Prolific feed after we custom-invite everyone who is eligible to participate based on the screening survey. The title of the study will be something like “Evaluating Circular Economy Solutions.”

### **B.3 Evaluation Survey Instructions and Demographic Information**

QID1 Thank you for participating in our research study!

What is your Prolific ID?

---

QID2 Instructions

Today you will be asked to evaluate 13 circular economy ideas to help a large research institution understand the impact of circular economy on business. We expect this survey to take ~30 minutes. We will first ask you some demographic information about yourself, followed by the evaluation task.

Your task is to rate the solutions based on their performance across four criteria:

Novelty: How different is it from existing solutions?

Environmental Impact: How much does it benefit the planet?

Financial Impact: What financial value can it create for businesses?

Feasibility and Scalability of Implementation: How likely is it to succeed and how scalable is it?

Based on these four criteria, you will then assess the overall quality of the solution. Please rate each solution on a scale from 1 to 5, with 1 being poor and 5 being outstanding. All solutions are of the format 'Problem' (problem identified) followed by 'Solution' (proposed solution).

After completing the survey, don't forget to click the Prolific URL at the end of the survey to claim your base payment of \$12. We will manually review your solutions to assess your bonus payment amount (up to \$13).

QID106 What is the highest level of education you have completed?

- Some high school, no diploma (1)
- High school graduate, diploma or the equivalent (2)
- Trade/technical/vocational training (3)
- Associate degree (4)
- Bachelor's degree (5)
- Master's degree (6)
- Professional degree (7)
- Doctorate degree (8)

QID107 In which field did you complete your highest level of education?

- Arts and Humanities (1)
- Social Sciences (2)
- Business (3)



- Life Sciences (4)
- Physical Sciences (5)
- Engineering (6)
- Technology/Computer Science (7)
- Education (8)
- Health and Medicine (9)
- Other (10) \_\_\_\_\_

QID108 What is your current employment status?

- Employed (1)
- Self-employed (2)
- Unemployed (3)
- Student (4)
- Retired (5)

QID111 What is your gender?

- Male (1)

Female (2)

Prefer not to say (3)

Q480 The circular economy concept that involves sharing, leasing, reusing, repairing, refurbishing and recycling existing materials and products as long as possible is:

	1 (outside my field of expertise) (1)	2 (2)	3 (at the boundary of my field of expertise) (3)	4 (4)	5 (inside my field of expertise) (5)
Expertise (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q481 How well do you understand the key factors that are critical for addressing problems related to circular economy?

	1 (none) (1)	2 (2)	3 (moderate) (3)	4 (4)	5 (very well) (5)
Level of understanding (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

End of Block: Demographics

Start of Block: End

QID112 Thanks for your participation! Please click the button below to be redirected back to Prolific and register your submission.

If you have any additional comments or feedback, please feel free to leave them below.

---

**Appendix C: Ordinary Least Squares (OLS) Regressions of Evaluator Ratings on AI and Human Solutions**

Table C1. OLS Regressions of Evaluator Ratings on AI vs. Human Solutions

VARIABLES	Novelty Model 1	Feasibility Model 2	Value Model 3	Quality Model 4
Baseline = Human				
AI-generated	-0.407*** (0.155)	-0.0140 (0.159)	0.228** (0.108)	0.0624 (0.137)
Constant	3.538*** (0.0488)	3.147*** (0.0447)	3.266*** (0.0352)	3.274*** (0.0462)
Evaluator FE	Y	Y	Y	Y
Problem Adj. FE	Y	Y	Y	Y
Solution Adj. FE	Y	Y	Y	Y
Observations	1,885	1,885	1,885	1,885
R-squared	0.032	0.018	0.016	0.009
Number of evaluators	145	145	145	145

Robust standard errors in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table C2. OLS Regressions of Evaluator Ratings on AI Configuration (Level and Instance) vs. Human Solutions

VARIABLES	Novelty Model 1	Feasibility Model 2	Value Model 3	Quality Model 4
Baseline = Human				
Level 1M	-0.516** (0.200)	-0.00737 (0.172)	0.221* (0.128)	-0.00822 (0.165)
Level 2M	-0.255 (0.186)	-0.00170 (0.174)	0.157 (0.130)	0.146 (0.164)
Level 3M	-0.643*** (0.181)	-0.0828 (0.203)	0.316** (0.130)	-0.0391 (0.171)
Level 1S	-0.246 (0.193)	-0.0365 (0.176)	0.248* (0.130)	0.104 (0.164)
Level 2S	-0.186 (0.185)	0.129 (0.166)	0.175 (0.123)	0.195 (0.153)

Level 3S	-0.493*** (0.172)	-0.0494 (0.201)	0.222 (0.138)	0.0234 (0.181)
Constant	3.538*** (0.0481)	3.147*** (0.0446)	3.266*** (0.0351)	3.274*** (0.0460)
Evaluator FE	Y	Y	Y	Y
Problem Adjective FE	Y	Y	Y	Y
Solution Adjective FE	Y	Y	Y	Y
Observations	1,885	1,885	1,885	1,885
R-squared	0.013	0.044	0.020	0.018
Number of evaluators	145	145	145	145

Robust standard errors in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table C3. OLS Regressions of Top Ratings on AI vs. Human Solutions

VARIABLES	Top Novelty Model 1	Top Feasibility Model 2	Top Value Model 3	Top Quality Model 4
Baseline = Human				
AI generated	-0.267*** (0.0542)	-0.00713 (0.0444)	0.00723 (0.0272)	-0.0403 (0.0432)
Constant	0.218*** (0.0185)	0.0989*** (0.0121)	0.0345*** (0.00737)	0.115*** (0.0149)
Evaluator FE	Y	Y	Y	Y
Problem Adj. FE	Y	Y	Y	Y
Solution Adj. FE	Y	Y	Y	Y
Observations	1,885	1,885	1,885	1,885
R-squared	0.039	0.016	0.015	0.010
Number of evaluators	145	145	145	145

Robust standard errors in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table C4. OLS Regressions of Top Rating on AI Configuration (Level and Instance) vs. Human Solutions

VARIABLES	Top Quality Model 1	Top Novelty Model 2	Top Feasibility Model 3	Top Impact Model 4	Env. Top Impact Model 5	Fin.
Baseline = Human						
Level 1M	-0.0315 (0.0553)	-0.304*** (0.0645)	0.0207 (0.0514)	0.102 (0.0713)	0.000184 (0.0448)	
Level 2M	0.00603 (0.0504)	-0.223*** (0.0683)	0.0220 (0.0498)	-0.0279 (0.0774)	0.00841 (0.0482)	
Level 3M	-0.0928	-0.294***	-0.0849	0.00157	0.0213	

	(0.0598)	(0.0647)	(0.0532)	(0.0718)	(0.0453)
Level 1S	-0.0106	-0.271***	0.0540	0.0981	-0.0290
	(0.0504)	(0.0628)	(0.0459)	(0.0664)	(0.0433)
Level 2S	0.0183	-0.206***	0.0669	0.00479	-0.00791
	(0.0450)	(0.0663)	(0.0461)	(0.0760)	(0.0420)
Level 3S	-0.0969*	-0.286***	-0.0765	-0.00260	-0.0245
	(0.0513)	(0.0697)	(0.0495)	(0.0698)	(0.0412)
Constant	0.115***	0.218***	0.0989***	0.218***	0.0713***
	(0.0146)	(0.0184)	(0.0116)	(0.0149)	(0.00949)
Evaluator FE	Y	Y	Y	Y	Y
Problem Adj. FE	Y	Y	Y	Y	Y
Solution Adj. FE	Y	Y	Y	Y	Y
Observations	1,885	1,885	1,885	1,885	1,885
R-squared	0.016	0.042	0.028	0.017	0.015
Number of evaluators	145	145	145	145	145

Robust standard errors in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## Appendix D: Implementation Details for Industry Classification of Answers

### D.1 Prompt-Building Mechanism and Modifying Prompts

The same prompt-building mechanisms described in A.1 were used to classify solutions generated by AI and humans under NAICS sectors. Before the solutions were classified, we prompted ChatGPT to individually summarize the 234 solutions. Then, using the summarized text, we prompted ChatGPT to classify it under NAICS sectors that were the most relevant to the text. In the prompt, we provided ChatGPT with a description of each sector taken from the NAICS manual. This allows ChatGPT to narrow down the information relevant to each sector and facilitate a successful classification.

### D.2 Prompts used for summarization

#### Context:

You are a summarizer model. Summarize the following text in one or two sentences.

#### Prompt:

[solution text]

### D.3 Prompts used for classification

#### Context:

You are a text classifier model. You will use a list of 20 industries from the North American Industry Classification System (NAICS) to classify a given text into one of them. Here are the 20 possible industries:

1. Agriculture, Forestry, Fishing and Hunting

Activities of this sector are growing crops, raising animals, harvesting timber, and harvesting fish and other animals from farms, ranches, or the animals' natural habitats.

2. Mining, Quarrying, and Oil and Gas Extraction

Activities of this sector are extracting naturally occurring mineral solids, such as coal and ore; liquid minerals, such as crude petroleum; and gases, such as natural gas; and beneficiating (e.g., crushing, screening, washing, and flotation) and other preparation at the mine site, or as part of mining activity.

3. Utilities

Activities of this sector are generating, transmitting, and/or distributing electricity, gas, steam, and water and removing sewage through a permanent infrastructure of lines, mains, and pipe.

4. Construction

Activities of this sector are erecting buildings and other structures (including additions); heavy construction other than buildings; and alterations, reconstruction, installation, and maintenance and repairs.

5. Manufacturing

Activities of this sector are the mechanical, physical, or chemical transformation of materials, substances, or components into new products.

6. Wholesale Trade

Activities of this sector are selling or arranging for the purchase or sale of goods for resale; capital or durable nonconsumer goods; and raw and intermediate materials and supplies used in production, and providing services incidental to the sale of the merchandise.

7. Retail Trade

Activities of this sector are retailing merchandise generally in small quantities to the general public and providing services incidental to the sale of the merchandise.

8. Transportation and Warehousing

Activities of this sector are providing transportation of passengers and cargo, warehousing and storing goods, scenic and sightseeing transportation, and supporting these activities.

9. Information

Activities of this sector are distributing information and cultural products, providing the means to transmit or distribute these products as data or communications, and processing data.

#### 10. Finance and Insurance

Activities of this sector involve the creation, liquidation, or change in ownership of financial assets (financial transactions) and/or facilitating financial transactions.

#### 11. Real Estate and Rental and Leasing

Activities of this sector are renting, leasing, or otherwise allowing the use of tangible or intangible assets (except copyrighted works), and providing related services.

#### 12. Professional, Scientific, and Technical Services

Activities of this sector are performing professional, scientific, and technical services for the operations of other organizations

#### 13. Management of Companies and Enterprises

Activities of this sector are the holding of securities of companies and enterprises, for the purpose of owning controlling interest or influencing their management decisions, or administering, overseeing, and managing other establishments of the same company or enterprise and normally undertaking the strategic or organizational planning and decision-making role of the company or enterprise.

#### 14. Administrative and Support and Waste Management and Remediation Services

Activities of this sector are performing routine support activities for the day-to-day operations of other organizations.

#### 15. Educational Services

Activities of this sector are providing instruction and training in a wide variety of subjects.

#### 16. Health Care and Social Assistance

Activities of this sector are providing health care

#### 17. Arts, Entertainment, and Recreation

Activities of this sector are operating or providing services to meet varied cultural, entertainment, and recreational interests of their patrons.

#### 18. Accommodation and Food Services

Activities of this sector are providing customers with lodging and/or preparing meals, snacks, and beverages for immediate consumption.

#### 19. Other Services (except Public Administration)

Activities of this sector are providing services not elsewhere specified, including repairs, religious activities, grantmaking, advocacy, laundry, personal care, death care, and other personal services.

#### 20. Public Administration

Activities of this sector are administration, management, and oversight of public programs by Federal, State, and local governments.

**Prompt:**

Classify the following circular economy idea into the 20 industry categories. If there are multiple relevant industries then list multiple.

[solution text]

Use the following template to answer:

<Write the name(s) of the industry that encapsulates accurately and as closely as possible the content of the idea. If there are multiple, separate each by a colon.>