

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

Léonard Boussioux¹, Jacqueline N. Lane^{2*†}, Miaomiao Zhang²,
Vladimir Jacimovic^{2,3}, & Karim R. Lakhani²

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

²Harvard Business School & The Digital Data and Design (D³) Institute at Harvard

³Continuum Labs

*Corresponding author: jnlane@hbs.edu

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

Abstract

The rapid advances in generative AI have the potential to reshape organizational innovation, raising uncertainty about the role of human solvers in this new era of augmented intelligence. We initiated a crowdsourcing challenge focused on sustainable, circular economy business opportunities, comparing the capabilities of GPT-4 and human solvers in generating novel and valuable solutions. The challenge attracted a diverse range of global solvers from various industries. 300 evaluators assessed a randomized selection of 13 out of 234 human and AI solutions, totaling 3,900 evaluator–solution pairs. Our findings reveal that, although AI solutions delivered more environmental and financial value—possibly due to a tendency to align with the central patterns seen in their training—human outputs were rated as more innovative, including extreme outcomes at the right tail of the novelty distribution. Our analysis of the rich solution text using natural language processing techniques reveals considerable overlap in semantic dissimilarity metrics between human and AI responses, but humans still exhibit greater linguistic nuances than AI. This study illuminates the promise of AI in augmenting human crowdsourcing for solving complex organizational problems and sets the groundwork for a possible integrative human-AI approach to innovative problem-solving.

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

We are grateful to Harvard Business School Research Associate Justin Ho, Program for Research in Markets & Organizations (PRIMO) fellow Stella Jia, who supported the data analysis for this project as well as Laboratory for Innovation Science at Harvard (LISH) lab manager Kate Powell who offered oversight and coordination for the research protocols. We also appreciate the Data to Actionable Knowledge Lab and the Human-Computer Interaction groups at Harvard for their feedback. We used GPT-4 for writing assistance. All errors remain our own.

The best answer to the question, “Will computers ever be as smart as humans?” is probably “Yes, but only briefly.”

—Vernor Vinge

Introduction

The success of organizational innovation is closely tied to a firm’s ability to search for the most profitable or best solution to a high-impact problem (Gavetti et al. 2005, Nickerson and Zenger 2004, Nonaka 1994). By expanding the range of independent approaches along the technical frontier (Abernathy and Rosenbloom 1969, Dahan and Mendelson 2001, Terweisch and Ulrich 2009), a firm can enhance the effectiveness of its search process, leading to superior innovation performance (Katila and Ahuja 2002, Vissa et al. 2010). Over the past decades, firms have increasingly engaged external solvers to source diverse viewpoints and deploy more approaches to problems that fall outside their core competencies (Afuah and Tucci 2023, Boudreau et al. 2011, Chesbrough 2003, Fayard 2023, von Hippel 2005, Lifshitz-Assaf 2018). Due to uncertainty about the best approach to solve such problems, attracting a large number of independent solvers can lead to alternative solutions that increase the likelihood of attaining an extreme, high-quality outcome (Boudreau et al. 2011, Dahan and Mendelson 2001).

The current, rapidly evolving landscape of technological innovation opens up the potential for artificial intelligence to reshape the boundaries of organizational innovation (Berg et al. 2023, Dell’Acqua et al. 2023, Anthony et al. 2023), similar to how external solvers have enhanced firms’ problem-solving effectiveness (Jeppesen and Lakhani 2010). Whereas AI covers a wide range of machine-driven tasks and processes, generative AI emphasizes creating new, previously non-existent content (Briot et al. 2017, Goodfellow et al. 2014, Manica et al. 2023) such as text, images, and music, based on patterns learned from existing data (Brown et al. 2020, Radford et al. 2018, 2019, Russell and Norvig 2010), in a scalable and cost-effective manner (Gaessler and Piezunka 2023, Lou and Wu 2022).

Recent studies on Large language models (LLMs)—a subset of generative AI designed to understand and produce text based on extensive training from vast amounts of written language (Bubeck et al. 2023)—suggest that LLMs can enhance individual productivity (Noy and Zhang 2023) and attain human levels of creativity (Berg et al. 2023, Girotra et al. 2023). In this study, we investigate, within the context

of crowdsourcing innovative solutions, the distinct capabilities of human solvers and LLMs in developing novel and valuable solutions. Novel solutions are original or rare ideas that depart from existing knowledge, and valuable solutions are useful ideas that generate economic and social returns (Corley and Gioia 2011, Piezunka and Dahlander 2015, Boudreau et al. 2016a, Berg 2016). High-quality solutions often balance novelty and value (Amabile 1988, Ghosh and Wu 2021, Kaplan and Vakili 2015).

Although it is plausible that LLMs have achieved a level of “intelligence” that may rival those of humans on productivity and creativity in some domains (Dell’Acqua et al. 2023, Noy and Zhang 2023), most studies focus on comparing human and AI capabilities on specific knowledge work tasks that are predominant in day-to-day organizational operations. These tasks often necessitate creative thinking but are characterized by clear problem formulation (Baer et al. 2013) and often backed by copious data to make accurate predictions, representations, and measurements (Cheng et al. 2022, Higgins et al. 2017, Lou and Wu 2022, Peterson et al. 2021, Vicinanza et al. 2022).

We know relatively little about whether LLMs can develop novel solutions for a firm’s problems that cannot be solved internally (Boudreau et al. 2011, Lifshitz-Assaf 2018, Townsend et al. 2018). Firms have turned to crowdsourcing to increase the number and diversity of perspectives on its problems (Jeppesen and Lakhani 2010). However, several challenges can impede its effective implementation. These challenges may include assembling and managing a crowd or filtering through a high volume of solutions to select the best ideas (Piezunka and Dahlander 2015, 2019). Moreover, increasing the number of participants in any contest, including crowdsourcing contests, decreases the likelihood of winning, which reduces incentives to exert effort and lowers overall innovation outcomes (Boudreau et al. 2011, Che and Gale 2003, Taylor 1995). This tension between the potential benefits of attracting many ideas from diverse solvers and the organizational difficulties of launching, managing, and incentivizing the crowd presents a compelling opportunity for AI to improve the effectiveness of a firm’s innovative search processes.

We propose that the rapid advances of LLMs (Du et al. 2023, Kıcıman et al. 2023, Min et al. 2022, OpenAI 2023, Vaswani et al. 2017, Wei et al. 2023) offer the potential to leverage the unique strengths of human solvers and LLMs for solving strategic, high-impact problems through a new, deliberate division of

labor (Agrawal et al. 2023, Anthony et al. 2023, Christensen and Knudsen 2020, Puranam 2021). Whereas human ingenuity results in a greater variety of paths that may result in more novel solutions, the exceptional computational proficiencies of LLMs may yield solutions of higher value.

The strength of LLMs resides in their ability to process and analyze vast amounts of data efficiently to produce outputs that suit human tastes (Argyle et al. 2022, Ouyang et al. 2022). These capabilities, such as adeptness in identifying patterns, trends, and correlations (Lou and Wu 2022), may enable LLMs to integrate knowledge from a wide range of disciplines effortlessly and combine existing ideas that are closely aligned with human dispositions (Agrawal et al. 2018, Cockburn et al. 2018, Lou and Wu 2022, Cheng et al. 2022). Essentially, LLMs operate on probabilistic foundations and generate responses word-by-word, typically selecting words that have a high probability of coming next based on training (Brown et al. 2020, Foster 2023, Gu et al. 2017, Radford et al. 2019, Yang et al. 2019). As a result, their outputs often gravitate toward the central or most frequent patterns seen in their training. This tendency suggests that LLMs may produce valuable solutions that appeal to a broad audience. However, their novelty may be limited in contexts that have limited prior templates, knowledge, and training data on which to base their responses (McCoy et al. 2023).

In contrast, human intelligence is celebrated for its abstract thinking and ingenuity (Amabile 1983, Cheng et al. 2022). Crowdsourced human solvers have a unique set of experiences, biases, and cultural backgrounds that ultimately create diverse and unique perspectives on the problem at hand. As a diverse human crowd's experiences are more varied and influenced by a wider range of cognitive and experiential factors, this breadth might increase the variance in the distribution of ideas generated by humans compared to LLMs. This increased variance might increase the likelihood of achieving occasional yet exceptionally novel outliers in the idea-generation process. Hence, highly novel ideas may more likely emerge from humans than AI.

To advance our understanding of how the future of crowdsourcing may be facilitated by both human intelligence and algorithmic capabilities, we partnered with Continuum Lab, an AI company to develop a crowdsourcing challenge about new business ideas on the circular economy. Our study involved

234 human- and AI-generated solutions, evaluated by 300 individuals, totaling 3,900 evaluator–solution pairs. Moreover, because evaluators were randomly assigned solutions to evaluate, our estimated relationships between the solution source and the assessed novelty, value, and quality of the solutions can be interpreted as causal estimates.

Our results are threefold. First, we highlight the distinct strengths of humans and LLMs in the crowdsourcing of novel and valuable solutions. Whereas humans generate more novel solutions, LLMs consistently produce outputs of higher perceived value. A closer look at the distribution of these solutions reveals that the broader range of human responses, compared to LLMs, is more likely to produce statistically rare, highly novel outcomes at the upper tail of the distribution. Second, we assessed the impact of both prompt engineering and model configuration on the novelty and value of responses. Building on recent work suggesting that LLM outputs may be less diverse than human responses (Dell’Acqua et al. 2023), we find that using an explicit differentiation sentence in the prompt between successive solutions can effectively elevate the novelty of the LLM’s responses without compromising their value. Third, we use natural language processing (NLP) techniques to investigate the text of the solutions to investigate semantic differences between human and AI solutions. Our average pairwise cosine distance using embedding vectors from the Bidirectional Encoder Representations from Transformer (BERT) Language Model (Devlin et al. 2019) indicates that humans exhibit greater semantic dissimilarity than AI, suggesting that a global human crowd’s advantage lies in its diversity of outputs that remains unmatched by LLMs.

Overall, our study contributes to the strategy and innovation literatures by illustrating the relative strengths of human-centered and AI-created solutions. We suggest that the future of innovative search processes will likely involve a division of labor between human and artificial intelligence: human ingenuity can be best utilized to develop statistically rare, highly novel solutions, while LLMs are best positioned for producing valuable outputs with broad appeal. Although the speed and scalability of LLM-generated solutions will allow for the generation of more solutions at relatively low marginal costs, the likelihood of achieving a high-quality outcome will depend on how well LLM solutions represent independent

approaches to the problem. This highlights the potential of collaborations between human solvers and AI, which could surpass the capabilities of either force working alone.

A Statistical View of Human and AI Crowdsourcing

We build on the statistical view of innovation, in which the quality of ideas tends to follow a normal distribution. Whereas most ideas are clustered around a mean level of quality, the right tail of the quality distribution corresponds to those that are statistically rare in both novelty and value (Dahan and Mendelson 2001, Terweisch and Ulrich 2009).

Crowdsourcing leverages a diverse pool of thinkers with differing backgrounds and experiences to increase the variance of new ideas. The goals of crowdsourcing are twofold: first, to generate a large number of parallel paths to yield many ideas; and second, to expand the range in quality of these ideas. Accordingly, crowdsourcing enhances the odds of identifying a novel and valuable idea that falls on the extreme, right tail of the distribution (Terweisch and Ulrich 2009). However, a direct implication of these joint goals is that crowdsourcing can be resource-intensive (Piezunka and Dahlander 2019) and statistically inefficient. This is particularly the case when the contest organizers care about maximizing innovation performance for a few top ideas as opposed to many average ideas (Dahan and Mendelson 2001, Girotra et al. 2010). Yet the process of analyzing a vast array of ideas necessitates a rigorous evaluation stage (Boudreau et al. 2016a, Piezunka and Dahlander 2015) to discern exceptionally high-quality outcomes from a multitude of solutions, many of which might not meet the desired criteria (Terwiesch and Ulrich 2009). In addition, the quest for such exceptional outcomes can be further complicated by diminishing contribution effort as the size of a contest grows (Boudreau et al. 2016b, Che and Gale 2003, Taylor 1993, Terwiesch and Xu 2008).

The statistical rarity of high-quality ideas from crowdsourcing might stem from the inherent challenge of simultaneously attaining both novelty and value in a proposed solution. Firms tend to prioritize one outcome over the other (Fang et al. 2010, Rindova and Petkova 2007), and attaining both novelty and value in a solution can sometimes correspond to conflicting goals (Ethiraj and Levinthal 2009, Ghosh and Wu 2022, Simon 1964). The rising capabilities of LLMs present an opportunity to enhance how firms generate novel and valuable solutions to solve complex problems at efficient speed and scale (Bubeck et al.

2023). In particular, a clear strength of LLMs over humans is their ability to generate timely and cost-effective outcomes (Muehlhauser and Salamon 2012) while adeptly handling significant workloads without compromising quality, which is challenging for humans due to their inherent physical and cognitive limitations (Huber and Power 1985, Nadkarni and Barr 2008).

Human Crowds, Large Language Models, and the Creation of Novelty and Value

We propose that LLMs can deliver solutions with higher value than their human counterparts, whereas humans' advantage lies in their ability to produce novel responses that emerge from human ingenuity and experience. To deepen our comprehension of the innovative capabilities of both humans and LLMs in the context of crowdsourcing, we explore the mechanisms underlying their ideation process. Our theoretical framework is informed by the characteristics of OpenAI's Generative Pre-trained Transformer 4 (GPT-4) as of summer 2023. GPT-4 is an advanced AI model designed to understand and generate human-like text based on large corpuses of generic text.

GPT-4 uses complex neural network structures known as transformers (Vaswani et al. 2017), which enable it to understand and predict language with high accuracy. The transformers employ a dynamic attention mechanism that enables the model to "pay attention" or consider the context of a sentence by taking note of nearby words (Ash and Hansen 2023, Bahdanau et al. 2015). During its training, GPT-4 learns through a hierarchical structure composed of numerous interconnected "neurons" or processing units, each contributing to the model's ability to infer the significance of every word within its textual environment. This involves complex algorithms derived from machine learning principles, particularly self-supervised learning, where the model refines its predictions by learning from examples with known outcomes.

Initially, GPT-4 undergoes a vast pre-training regime on a diverse and extensive dataset collected from a wide variety of sources on the internet (e.g., books, websites, scientific articles, forums, and news outlets that cover an extensive array of topics, genres, and subjects in numerous languages) to acquire a broad linguistic foundation. It is trained based on sophisticated optimization algorithms for model parameters that can be adjusted to minimize prediction errors iteratively. After its initial pre-training on a

general dataset, the model is fine-tuned on more specialized datasets to become proficient on specific tasks or to improve performance in certain domains.¹

Indeed, the LLMs' initial tendencies to mirror the most prevalent patterns in their training data can sometimes result in low-quality answers or misalignments with human tastes, such as biased or undesirable outputs. To mitigate these issues, LLMs may undergo a machine-learning refinement process known as Reinforcement Learning with Human Feedback (RLHF), wherein human evaluators shape the LLM's outputs to align more closely with ethical standards, societal expectations, and high-quality content. During this phase, human evaluators assess generated text samples from the trained model, providing ratings that inform updates to the model's parameters. This ensures that the model progressively favors outputs aligned with human values and preferences. This process of generating samples, obtaining human feedback, and updating the model is critical, as it further fine-tunes the model's parameters to favor text samples that receive positive human evaluations, progressively steering the model towards text that resonates positively with a general audience.

GPT-4 is a representative example of advanced language models, which tend to operate based on similar foundational principles (see Appendix B for a detailed overview of inference processes of LLMs). Its primary strength lies not only in its ability to generate fluent and coherent text efficiently with a relatively small marginal cost (Brands et al. 2023, Horton 2023) but also to understand and manipulate task instructions in various ways, such as summarizing, translating, or answering an extremely broad set of questions (Bubeck et al. 2023). LLMs have already demonstrated remarkable capabilities in a variety of domains and tasks, including comprehension (Bubeck et al. 2023), code (Chen et al. 2021), causal reasoning (Kiciman et al. 2023), market research (Brands et al. 2023), and experimentation (Horton 2023). Moreover, GPT-4 can produce responses that encompass a spectrum of linguistic styles and perspectives (McCoy et al. 2023). Using strategic prompt engineering, a user can direct the model to leverage specific aspects of its

¹ The training process of GPT-4 has never been made publicly available, and the complicated procedures of training an LLM is beyond the scope of this study. We refer to the GPT-4 technical report for extended reading: <https://cdn.openai.com/papers/gpt-4.pdf>.

training, revealing the nuanced impact of input phrasing on the resulting outputs, even when the prompts contain similar objectives or semantic meaning. These capabilities position LLMs as versatile tools equipped to simulate a diverse range of human-like perspectives (Argyle et al. 2022), presenting a strong case for their use in crowdsourcing scenarios where varied viewpoints are essential and desired (Jeppesen and Lakhani 2010).

Returning to the relationship between the tendency for LLMs to generate more valuable responses than their human counterparts, we posit that the RLHF training stage refines the model's outputs so that they become increasingly representative of a consensus view, which enhances their acceptance across diverse audiences. This adaptability positions LLMs to offer solutions that address common concerns and meet societal needs, making them particularly valuable and well-received across a wide range of users (Kim and DellaPosta 2021, Sharkey and Kovacs 2014).

By contrast, humans face cognitive and physical limitations (Levinthal and March 1993), made ever more challenging by the growing "burden of knowledge" (Jones 2009). This increasing specialization has made it progressively more difficult for humans to gather the requisite range of perspectives (Nagle and Teodoridis 2019, Lane et al. 2023) to find the appropriate combinations of knowledge (Fleming 2001, Fleming and Arts 2018) that will produce valuable new ideas. Unlike LLMs, humans' responses may change based on external and internal factors, such as mood, fatigue, stress, or recent experiences, which can significantly affect their contributions, leading to variabilities in their responses (Christian and Ellis 2011, Argote and Miron-Spektor 2011, Lerner et al. 2015). Humans also have personal biases, preferences, motivations, and opinions (Amabile 1985, Benjamin et al. 2013, Caplin et al. 2015, Harrison and Klein 2007, Kahneman and Lovallo 1993). The diversity of these subjective elements can lead to different interpretations of the same problem and, consequently, varied solutions that may not align with the consensus or central perspective (Jeppesen and Lakhani 2010).

Such variations in human responses are likely to be amplified in human crowds (Goldenberg and Gross, 2020), as they are composed of individuals from various cultural, geographical, educational, and experiential backgrounds. The differences between the human crowds and LLM intelligence suggest that

each entity has unique strengths that will contribute distinctively to the novelty and value of their generated solutions. LLMs, by their design, will produce consistent outputs (Lou and Wu 2022) that align with established frameworks, practices, and preferences that resonate with a broad audience (Kim and DellaPosta 2021, Ouyang et al. 2022). These solutions will likely be recognized and readily accepted for their applicability and practicality and thus perceived as valuable. In contrast, a human crowd yields alternative perspectives that may result in differentiated but less consistent outputs.

However, LLMs also have their drawbacks, which may offer human solvers an advantage in developing novel outputs over LLMs. Current LLMs are trained on vast yet *finite* datasets (McCoy et al. 2023), which can limit the uniqueness of solutions that they are capable of generating. LLMs are trained to predict the next word in a sequence, and producing “out-of-the-box,” statistically rare ideas may be challenging for these systems (Bojinov 2023, Li et al. 2023, McCoy et al. 2023). A recent study shows LLMs have yet to achieve human-level creativity across a broad range of evaluation criteria (Chakrabarty et al. 2023). By contrast, a consortium of different individuals with deep contextual understanding and expertise in specific topics may allow for diverse problem-solving approaches that might not be immediately obvious to other humans or algorithms. Humans may draw connections between disparate concepts under uncertainty (Fleming 2001, Jeppesen and Lakhani 2010) that cannot be easily replaced by structured pattern recognition. This greater variation in ideas may statistically increase the likelihood of outliers that push the boundaries of innovation (Jeppesen and Lakhani 2010, Terweisch and Ulrich 2009). Hence, compared to LLM-generated ideas, human ingenuity may support generating highly novel solutions.

In short, in a crowdsourcing contest, we propose that human solvers are more inclined to generate novel ideas compared to LLM responses, while LLMs are likely better-suited to produce solutions of higher value.

Hypothesis. In comparing human crowd-based and LLM idea generation, LLM-generated solutions are perceived as more valuable, whereas human crowd-generated solutions are perceived as more novel.

Research Design and Methods

Setting

Crowdsourcing Context. We partnered with Continuum Lab, an AI company, and Freelancer.com, an online marketplace, to launch a crowdsourcing challenge seeking new business ideas focused on sustainable, circular economy business opportunities. We selected the circular economy due to its multidisciplinary nature—spanning environmental science, economics, design, and engineering—making it a complex problem in a highly uncertain context to evaluate the innovativeness of human and AI solutions rigorously. This global issue demands both broad cultural insights and knowledge of specific topics for meaningful contributions. The challenge ran from January 30, 2023 to May 15, 2023. Participants were encouraged to submit real-life use cases of how companies can implement the circular economy concepts in their businesses. The circular economy is an economic system that emphasizes eliminating waste and continual resource reuse, contrasting with the traditional “take, make, dispose” linear model.² 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 Value*, *Financial Value*, and *Feasibility and Scalability of Implementation*.

All participants submitted their solutions using a Google Form. We also collected their 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, we 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.

LLM Idea Generation. We use GPT-4 to generate various solutions in response to the same crowdsourcing challenge of developing sustainable, circular economy business ideas. Prompt engineering, which involves creating effective input prompts for the AI (Brown et al. 2020, Ray 2023), greatly affects the AI’s output

² See <https://www.epa.gov/circulareconomy/what-circular-economy>.

quality and relevance. Therefore, we used three alternative prompt engineering approaches to produce the AI-generated solutions. In Appendix A, we provide details on the specific prompts used to generate the AI solutions.

Moreover, recent work suggests that LLMs may produce homogenized outputs, potentially reducing the diversity of ideas (Dell’Acqua et al. 2023). To address this concern, we implemented two alternative configurations of GPT-4 aimed at diversifying outputs: (1) multiple-instance solutions and (2) single-instance solutions with differentiation instruction (see Appendix A).

For the first configuration, each distinct instance of GPT-4 generates its solution independently from the same input prompt. While the model and prompt remain identical across instances, the sampling methods intrinsic to LLMs mean that each instance can produce varied responses because the model samples from the probability distribution of possible next words or considers several high-probability next words rather than simply selecting the next word with the highest probability.

The second configuration uses a single instance of GPT-4 to generate multiple solutions successively, one at a time. By incorporating a differentiation instruction between successive responses, a single instance of GPT-4 will attempt to differentiate its successive responses from previous ones, enabling a potentially deeper exploration of the solution space. Intuitively, 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. In contrast, the second configuration resembles an individual solver who iteratively proposes distinct ideas. For both configurations, we set the temperature parameter, which controls the randomness of predictions, at the GPT-4 API’s default value.

API Costs and Time Spent. We generated 730 AI solutions, 315 each with the multiple and single instances of GPT-4. 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 cost \$0.037 on average. Table A1 provides sample human- and AI-generated solutions.

Evaluator Recruitment and Procedures

Our study (approved under Harvard University IRB23-0770) uses human evaluators to judge the novelty and value of human and GPT-4 solutions. First, we recruited potential evaluators on Prolific.org in July 2023 and September 2023. For both recruitment sessions, 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 (60%) or more on the skills test were invited to participate in the evaluation survey (see Appendix E.1 and E.2 for survey instruments). Overall, we recruited 300 evaluators, 145 from the first call and 155 from the second. In addition to the screening criteria, we collected demographic data on the evaluators' gender, highest level of education, field of study, and employment status. Table 1 shows the descriptive statistics of the evaluators.

We randomly selected a total of 234 solutions for human evaluation. Of these, 180 were AI-generated, and 54 were human-generated.³ For the AI solutions, we randomly selected a mix of AI-generated responses, instructed with three alternative prompts, and evenly allocated between multiple and single instance configurations (see Appendix A). We used a randomized block design to randomize the AI- and human-generated solutions into distinct blocks. Each block included 10 AI and three human solutions, totaling 13 solutions 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 (defined as the mode) quality rating among all evaluators assigned to the same solution. Each evaluator received a mean bonus of \$6.43 (s.d. = \$2.30, min = \$1, max = \$12). The total compensation per evaluator ranged from \$13 to \$24.

³ Our AI-generated solutions exceeded human solutions since our randomized selection of AI responses was done at the prompt level. Please refer to Table A1 in Appendix A for a detailed illustration of prompt engineering and single (S) and multiple (M) instance configurations.

Consistent with our research question regarding the comparative abilities of humans and LLMs to generate novel and valuable solutions, each evaluator rated the solutions for their novelty (*How different is it from existing solutions?*), environmental value (*How much does it benefit the planet?*), and financial value (*What financial value can it create for businesses?*).⁴

--- Insert Table 1 here ---

Variables

Dependent Variables. We use two main dependent variables, corresponding to the evaluator's *Novelty rating* and *Value rating* of each solution. We computed the *Value rating* by taking the average of the evaluator's environmental and financial value ratings. To examine extreme outcomes, we created binary variables for *Top novelty rating* and *Top value rating*. The *Top novelty rating* was equal to 1 if a solution received the highest novelty rating and 0 otherwise, whereas *Top value rating* was equal to 1 if an evaluator gave a solution the highest environmental and financial value rating.

Independent Variables. Our main independent variable, *AI solution*, is a dummy variable corresponding to whether the solution is human-generated (baseline) or AI-generated. We also report an alternative independent variable, *AI instance*, a categorical variable that further differentiates between the AI solutions as either *Multiple instance* or *Single instance*. This alternative independent variable enables us to develop deeper insights into how alternative configurations of GPT-4 will influence the novelty and value of the produced responses.

Other Variables. Our statistical analyses rely on the random assignment of evaluators to solutions. That being said, we add several covariates corresponding to the screening criteria (i.e., work experience, interest, and skills test score), the evaluators' demographic characteristics (i.e., gender, bachelor's degree or higher, STEM major, employment status), the solution word count, and the recruitment session cohort (i.e., July or September 2023).

⁴ Evaluators also rated each solution on its feasibility and scalability of implementation (*How likely is it to succeed and how scalable is it?*) and overall quality (*Based on the four criteria above, what is the overall quality of the solution?*). We used the quality rating to determine the evaluators' bonus payments.

Estimation Approach

We analyzed our data at the evaluator–solution block level. We use nested mixed-effects models or hierarchical linear models (Gelman and Hill, 2006, Kenny et al. 2006, Wiltermuth et al. 2023), performed using the `lmerTest` package in R (Kuznetsova et al. 2017), to account for the interdependence of data around the evaluators and solution blocks. These models appropriately account for the nesting of evaluators within solution blocks by estimating random effects (i.e., random intercepts and slopes) for both the solution blocks and the evaluators. By modeling variability at both the evaluator and solution block levels, mixed-effects models can provide more accurate estimates and standard errors than ordinary least squares (OLS) in the presence of nested data (Gelman and Hill 2006).

Results: Evaluator Ratings of Human and AI Solutions

Figure 1 compares the density distributions of the mean novelty (Panel A) and value (Panel B) evaluator ratings assigned to the human- and AI-generated solutions. Although the density distributions of the human and AI solutions have some overlap, the human solutions exhibit a higher right tail for novelty (Human: Hill estimator = 37.8 vs. AI: Hill estimator = 35.5) and a heavier left tail for value (Human: Hill estimator = 40.6 vs. AI: Hill estimator = 34.8).⁵ Whereas the mean novelty rating for human responses is higher than the AI responses (Human: mean novelty = 3.505, s.d. novelty = 0.513 vs. AI: mean novelty = 3.385, s.d. novelty = 0.373; two-tailed t -test = 1.593, $p = 0.116$), the mean value rating for the AI responses is higher than those of the human-generated solutions (Human: mean value = 3.346, s.d. value = 0.340 vs. AI: mean value = 3.523, s.d. value = 0.228; two-tailed t -test = -3.589, $p < 0.001$).

--- Insert Figure 1 here ---

Next, Figure 2 shows the density plot distributions of the mean evaluator rating by solution for novelty (Panel A) and value (Panel B) by *AI Instance*. We observe in Figure 2 (Panel A) that the single-instance AI solutions have a higher mean novelty rating than the multiple-instance AI solutions (Single

⁵ The tail index (also known as the Pareto index) is a concept that originally arose from the Pareto distribution. However, the idea of a tail index has been generalized beyond just the Pareto distribution to provide insights into the tail behavior of other distributions using methods from extreme value theory. One common method is the Hill estimator.

instance: mean novelty = 3.470, s.d. novelty = 0.340 vs. Multiple instance: mean novelty = 3.301, s.d. novelty = 0.386; two-tailed t -test = 3.121, $p = 0.002$). Moreover, there is no difference in the mean novelty rating between the single-instance AI solutions and the human solutions (Single instance: mean novelty = 3.470, s.d. novelty = 0.340 vs. Human: mean novelty = 3.504, s.d. novelty = 0.513; two-tailed t -test = -0.446, $p = 0.657$), even though the human solutions appear to have a heavier right tail.

In contrast, turning to the distribution of *Value ratings* in Figure 2 (Panel B), there is no difference between the mean value rating (Single instance: mean value = 3.534; s.d. value = 0.235 vs. Multiple instance: mean value = 3.513, s.d. value = 0.222; t -test = 0.600, $p = 0.550$) of the single- and multiple-instance AI solutions.

--- Insert Figure 2 here ---

Mixed Effects Models. Tables 2–5 report the mixed model results of *Novelty rating* (Table 2) and *Top novelty rating* (Table 3), *Value rating* (Table 4) and *Top value rating* (Table 5). In all tables, the main independent variable in Models 1–3 is *AI solution*, and the main independent variable in Models 4–6 is *AI instance*. Both Models 1 and 4 report the main effect of the solution source. Models 2 and 5 add the evaluator screening criteria, and Models 3 and 6 control for additional evaluator attributes, cohort, and solution word count. For *Top novelty rating* and *Top value rating*, we additionally report mixed effects logistic regression model results in Appendix C.

Estimated Relationships Between Solution Sources and Solution Novelty. In Table 2, Model 1 indicates that compared to human solutions, AI solutions receive a lower novelty rating on average compared to human crowdsourced solutions (Model 1: -0.124, $p < 0.001$). Models 2 and 3 indicate that the estimated coefficient remains stable and robust after adding the evaluator screening criteria (Model 2: -0.124, $p < 0.001$) as well as the evaluator attributes, cohort, and solution word count controls (Model 3: -0.140, $p < 0.001$).

In Model 4, we model the solution source as a categorical variable to differentiate between human solutions and AI multiple- and single-instance solutions. Compared to human solutions, we observe that the AI solutions generated with multiple instances of GPT-4 are rated as significantly less novel (Model 4:

-0.209, $p < 0.001$), but there is no difference between the human and single-instance AI solutions (Model 4: -0.039, *ns*). Using the `emmeans` package in R, we perform pairwise comparisons to show that the coefficients for the *Multiple Instance* and *Single Instance* AI solutions in Model 4 are significantly different from each other ($p < 0.001$). Next, Model 5 indicates that the estimated relationships remain stable and significant after adding the evaluator screening criteria (Multiple instance: -0.209, $p < 0.001$; Single instance: -0.039, *ns*) and so does Model 6, which includes the evaluator demographic attributes, cohort, and solution word count controls (Multiple Instance: -0.217, $p < 0.001$; Single Instance: -0.056, *ns*).

Next, we turn to Table 3 to investigate the relationships between the *most-novel* solutions, which achieved the *Top novelty rating*, and the solution source. In Model 1, we observe that, compared to human responses, AI solutions are 7.4 percentage points (pp) less likely to receive the top novelty rating (Model 1: -0.074, $p < 0.001$). Models 2 and 3 show that the estimated relationships remain robust after adding the evaluator, cohort, and solution controls. Model 4 splits the AI solutions into *Multiple-* and *Single-instance* solutions and indicates that both configurations are less likely than the human solutions to receive the top novelty rating (Multiple Instance: -0.088, $p < 0.001$; Single Instance: -0.059, $p < 0.001$). Once again, Models 5 and 6 indicate that the reported coefficients remain consistent and robust with the evaluator screening criteria and other controls.

Although Tables 2 and 3 highlight the perceived greater novelty of human solutions compared to AI-generated ones, instructing GPT-4 to differentiate its responses within a single instance configuration demonstrates its potential to produce notably more novel outputs on average than prompting with multiple or parallel instances.

--- Insert Tables 2 and 3 here ---

Estimated Relationships Between Solution Sources and Solution Value. Turning to Table 4, Model 1 indicates that AI solutions are rated as more valuable than human solutions (Model 1: 0.171, $p < 0.001$). We observe that this estimated relationship remains statistically significant in Models 2 and 3, which add the evaluator screening criteria (Model 2: 0.171, $p < 0.001$) and evaluator and solution controls (Model 3: 0.152, $p < 0.001$), respectively. Next, in Model 4, we use the categorical variable, *AI instance*, to

differentiate between the human, multiple- and single-instance AI solutions. We observe that compared to human-generated responses, both the multiple- and single-instance solutions are rated as more valuable (Multiple instance: 0.160, $p < 0.001$; Single instance: 0.182, $p < 0.001$). A post hoc pairwise comparison of coefficients using the `emmeans` package in R indicates that the multiple- and single-instance coefficients are not significantly different from each other ($p = 0.622$). Moreover, we note that the estimated relationships are unchanged in Models 5 and 6, which add evaluator screening criteria (Multiple Instance: 0.160, $p < 0.001$; Single Instance: 0.182, $p < 0.001$) and evaluator demographic controls, cohort, and solution word count (Multiple Instance: 0.148, $p < 0.001$; Single Instance: 0.156, $p < 0.001$).

Next, we investigate the relationships between the most valuable solutions, achieving the *Top value rating*, and the solution source. Table 5 Model 1 shows no significant difference between humans and AI regarding their likelihood of generating a highly valuable solution (Model 1: 0.019, *ns*). Models 2 and 3 indicate that the AI value advantage remains weak. Turning to Model 4, we observe that, compared to the human responses, there is no difference between the AI multiple- and single-instance configurations and the likelihood of generating a highly valuable solution (Multiple Instance: 0.020, *ns*; Single Instance: 0.017, *ns*). There is no change in the estimated relationships in Models 5 and 6.

In summary, in Tables 4 and 5, we find that the AI responses achieved higher value ratings on average than the human solutions. However, there is no difference in top value between the solutions produced by humans and AI, and the multiple- and single-instance AI configurations do not have a meaningful effect on the solution's value. One possible explanation is that the differentiation prompt in the single-instance configuration is likely to force different or unique answers that push the model towards greater novelty without changing the value of their outputs. An important insight of the single-instance configuration is that we can achieve more novel responses (see Tables 2 and 3) without compromising the perceived value of the responses.

Taken altogether, the results in Tables 2–5 support our hypothesis that LLMs generate more valuable outputs than humans and that humans produce more novel outputs that, importantly, exceed LLM novelty in the right tail of the novelty distribution.

--- Insert Tables 4 and 5 here ---

Semantic Dissimilarity between Solution Sources

In this section, we investigate the rich text of the solutions to determine the extent to which human and AI solutions exhibited a broad range of content. Intuitively, a larger dissimilarity between the content of the submitted solutions indicates a greater diversity in ideas. To this end, we compare two types of dissimilarity: (1) within-source dissimilarity, or the degree of dissimilarity between submitted solutions from the same source (human-human and AI-AI); (2) and between-source dissimilarity, or the degree of dissimilarity between submitted solutions from different sources (human-AI and AI-human).

To address this question, we followed Carlson (2022) and Park et al. (2023) differentiation measures to proxy the diversity of human and AI responses. Specifically, we utilized BERT using the transformers library in Python. The BERT architecture uses Masked Language Modeling (MLM) as the primary method of training the language model, generating embeddings that are bidirectionally aware, which allows BERT embeddings to excel in semantic representations for natural language understanding tasks (Devlin et al. 2019). This process translates each solution description into a 768-dimension vector representation, effectively capturing the semantic meanings of the submitted responses. Next, we compute the pairwise cosine distance between the vector representations of each solution pair in the dataset. For every human- or AI-generated solution, we determine the average pairwise cosine distance with respect to every other human or AI response to derive a dissimilarity measure (Carlson 2022, Park et al. 2023). This average distance is computed using within-source (i.e., within-human or AI) or between-source dissimilarity to obtain four sets of dissimilarity scores. The resulting metric provides insight into the dissimilarity of a solution to its human or AI neighbors in the vector space. A low dissimilarity score suggests a solution is akin to many others in the dataset, whereas a high score denotes distinctiveness.

Mathematically, the dissimilarity between two solutions i and j is measured as $d_{ij} = 1 - \frac{\underline{v}_i \cdot \underline{v}_j}{\|\underline{v}_i\| \|\underline{v}_j\|}$,

where \underline{v}_i and \underline{v}_j are the pre-processed BERT vectorized representations, and $d_{ij} \in [0,1]$. We then capture a solution i 's dissimilarity by averaging the cosine distances between all possible combinations of solutions

within- or across-sources.⁶ Equations (1) and (2) illustrate the nuances between the two mean dissimilarity measures:

$$\textit{Within-Source Mean Dissimilarity Score}_{i \in S} = \frac{1}{|S|-1} \cdot \sum_{j \in S} d_{ij}, \quad (1)$$

$$\textit{Across-Source Mean Dissimilarity Score}_{i \in S} = \frac{1}{|S'|} \cdot \sum_{j \in S'} d_{ij}, \quad (2)$$

where $S \in \{S_H, S_{AI}\}$ is the set of solutions that correspond to either the human or AI group of solutions, and S' denotes the complementary group to S . For instance, if $S = S_{Human}$, then $S' = S_{AI}$.

Figure 3 plots the four alternative mean dissimilarity scores for AI solutions (Panel A) and human solutions (Panel B), showcasing the distribution of semantic dissimilarity between AI and human solutions. Specifically, Figure 3 (Panel A) depicts the distribution of dissimilarities for the within-source AI-AI and between-source AI-human dissimilarities; Figure 3 (Panel B) shows the distribution of within-source human-human and between-source human-AI solutions. First, Figure 3 (Panel A) indicates a high consistency in AI-generated solutions, as the distribution is characterized by a sharp peak around a mean dissimilarity of 0.058 (s.d. = 0.008). On the other hand, the AI-human distribution concentrates around a mean dissimilarity of 0.119 (s.d. = 0.009), indicating high semantic consistency within the AI solutions and greater semantic dissimilarity between the AI-human solutions. Turning to Figure 3 (Panel B), we observe a wider distribution overall, with barely noticeable peak density in the human-AI comparison group. Even though the human-human group still exhibits some resemblance, the peak value is further to the right with more dissimilarity and fatter tails.

In summary, Figure 3 indicates greater diversity among the human solutions than the AI-generated responses, with considerable overlap between human-AI solutions. This suggests that there are opportunities to be explored involving human-AI collaboration. To enrich our understanding of the semantic differences between human- and AI-generated solutions across different prompt engineering

⁶ For the within-source comparison, we averaged across dissimilarities within the same source, i.e., for human solutions, we only constructed the cosine distance with other human solutions, and for AI, we only calculated the cosine distances compared to all other AI solutions. For the between-source comparison, we followed a similar procedure, except that we only averaged across all between-source pairs.

techniques and model configurations, we further classify the solutions based on NAICS-2 industry sectors (see Appendix D.1), rephrase human-generated solutions using GPT-4 to account for systematic linguistic discrepancies not captured by BERT embeddings (see Appendix D.2), and plot the projected the 2-dimensional principal components to visualize the semantic variation as well as novelty and value ratings among solution sources (see Appendix D.3).

--- Insert Figure 3 here ---

Discussion

We began this paper with the following question: In the era of generative artificial intelligence, how will human solvers contribute to organizational crowdsourcing efforts? To investigate this question, we partnered with Continuum Lab, an AI firm, to launch a crowdsourcing challenge to identify sustainable, circular economy business opportunities. We subsequently invited human evaluators to assess the novelty and value of the submitted solutions without revealing their sources as human- or AI-generated.

Our study yields three main insights. First, whereas LLMs consistently produce solutions of higher value at efficient speed and scale, humans generate more novel solutions. A closer look at the distribution of these solutions reveals that the broader range of human responses, compared to LLMs, increases the chances of identifying statistically rare but highly novel outcomes at the upper end of the distribution. Second, we assess the impact of model configuration on the novelty and value of responses. We find that a simple instruction reminding GPT-4 to produce unique responses can effectively elevate the novelty of the LLM's responses without compromising their value. Lastly, through NLP techniques, namely BERT vector representations, we find that humans show greater semantic dissimilarity than AI. The findings suggest several important implications for the future of crowdsourcing innovations, which we believe will likely center on human-AI collaboration.

The Future of Human and AI Idea Generation

Our study evidences GPT-4's remarkable proficiency in generating solutions for innovative problems at an unparalleled speed and scale. This algorithmic problem-solving capacity offers tremendous potential for improving the efficiency of human crowdsourcing (Boudreau et al. 2011, Che and Gale 2003, Piezunka and

Dahlander 2019, Taylor 1995). In the near term, this enhanced capability may be particularly relevant for incremental product or service enhancements, where AI's speed and productivity are unmatched by humans in generating a plethora of valuable solutions. That said, our findings also reveal that human ingenuity is still irreplaceable by AI, particularly in the pursuit of highly novel ideas. An important implication of these complementary strengths (Choudhury et al. 2020, Daugherty and Wilson 2018, Raisch and Krakowski 2021) reveals the potential for delineating ideation roles, whereby humans focus primarily on developing novel outputs while AI is tasked with producing valuable responses. This division of labor may aid in reducing the cognitive and physical burden on humans, who may need more slack time (Agrawal et al. 2018) to conjure up imaginative thoughts and develop highly novel outputs (Baer et al. 2020).

Furthermore, our study underscores the promising potential of productive human-AI collaborations, where the collective outputs may outperform either entity working in isolation (Puranam 2021). Our analysis of semantic dissimilarity scores, as presented in Figure 3, reveals interesting patterns. Human outputs display greater variability than AI-generated counterparts, but substantial overlap exists in the between-source human and AI semantic dissimilarity scores. The overlap indicates that under certain conditions, AI-generated outputs can exhibit variability comparable to human-generated content. This suggests that AI outputs can sometimes statistically mirror the unpredictability we associate with human thought and ingenuity. It is also worthwhile to note that the degree of overlap exists even though human responses may be semantically more diverse from each other than GPT-4's outputs due to the global representation of countries among the human solvers in our crowdsourcing contest. In contrast, GPT-4, trained on a curated dataset enhanced by RLHF, is designed to produce consistent and high-value outputs, which can lead to a certain degree of "standardization" and "polished" answers. These findings do not rule out the possibility that with increased computational resources, high-quality data, and better algorithms, AI has the potential to further expand the frontier of innovation by generating ideas that are both consistent and varied, thus mimicking, or even surpassing human creativity. The observed semantic overlap between humans and AI outputs reminds us of the risks of humans becoming overreliant on AI. As the application and uptake of LLMs expand, humans ought to integrate their own experiences, knowledge, personalities,

and preferences into their AI interactions. Overlooking this integration could lead to homogenized ideas, undermining the value of diverse perspectives.

To fully realize the benefits of LLMs for creative problem-solving, effective human-AI collaborations are likely to hinge upon the ability to identify an optimal structure for humans and AI to work together so that the benefits of specialization outweigh the costs of interdependence (Agarwal et al. 2023, Teodoridis 2019). As we increasingly move towards a landscape of human-AI collaboration (Jia et al. 2023, Kellogg et al. 2020, Puranam 2021), it will become crucial to design the optimal structure for this human-AI partnership, including coordination decisions around the division of tasks into a sequential, parallel, or feedback-driven iterative process (Puranam 2021).

The Evolving Capabilities of LLMs and the Role of Human Ingenuity

Our study focused on the existing capabilities of GPT-4. However, the rapid advancement of scholarly research on AI suggests that the specific nature of human-AI collaboration is likely to evolve over time. Notably, our findings comparing the single- and multiple-instance configurations revealed that carefully crafted prompt instructions under a single-instance configuration may push an LLM into the tails of its distribution to yield statistically rarer ideas. Given that we used a simple instruction to encourage differentiation, it suggests that more sophisticated prompts that seek to widen the tails of LLM responses may hold promise for generating more novel outputs. Another possibility is to increase the model's temperature parameter, which can make the text more varied and potentially more novel.

In our research, we focused on the capabilities of a single LLM. An intriguing avenue for further elevating LLM creativity is to build on the collective insight of multiple collaborating AI agents (Du et al. 2023). Beyond OpenAI's GPT-4 model, there is an array of LLMs such as PaLM (Chowdhery et al. 2022), Llama 2 (Touvron et al. 2023), and Claude (Anthropic 2023) that are swiftly advancing and beginning to rival the capabilities of GPT-4. With recent infusions of capital in these models from large firms (Amazon 2023, Microsoft 2023), their capabilities will likely continue to improve and potentially exceed those of OpenAI's GPT series with time. Importantly, because these alternative LLMs are trained on different

datasets, their collaborative output could offer more novel recombinations than a single response from GPT-4—and, hence, could develop outputs that may be more novel than individual humans.

The continued advances of LLMs bring to light the need to better understand which facets of humanity—such as humans’ intrinsic capability for intuition, domain-specific expertise, and nuanced understanding of context—will remain distinct human advantages in innovative contexts (Bernstein et al. 2023). Even though the new cohort of LLMs shows more general intelligence than previous AI models, and their capabilities will only continue to improve (Bubeck et al. 2023, Regenwetter et al. 2023, Wei et al. 2022), there may be some parts of the problem and solution space—the extreme tails of the distribution—that are only conceivable through lived and varied human experiences (Singh and Fleming 2010), serendipity (Lane et al. 2021), and social interactions with other humans (Perry-Smith and Shalley 2014, Singh 2005). In this respect, it may be ever more critical for humans to foster opportunities for exposure and connection with other humans to spark breakthroughs. As teams tend to outperform the lone genius in evoking creative outputs (Singh and Fleming 2010, Wuchty et al. 2007), one promising direction to further explore is the role that AI plays in shaping the distribution of team outcomes, both in mitigating the likelihood of poor outcomes and increasing the probability of attaining extremely creative outputs.

Human- or AI-Generated? Bias, Fairness, and Intellectual Property (IP) Concerns

Finally, as these technologies are poised to transform industry, education, and social relations, we must address concerns about fairness, reliability, and IP, especially when using AI for creative tasks. Notably, LLMs produce text that reflects their training data and has the potential to exacerbate biases about race, sex, language, and culture (Li et al. 2023). Moreover, the body of knowledge used to train these models tends to arise from well-funded institutions in high-income, English-speaking countries (Ji et al. 2023). This leaves a significant underrepresentation of perspectives from other regions, which could bias understandings of these processes to a handful of geographic and demographic hubs. Lastly, the legal landscape of LLMs on the ownership of generated content is still evolving: if an LLM produces content that mirrors or closely resembles human sub-population attitudes or expressions, can it be considered original? Or does it become a derivative of the training data, potentially infringing on the IP rights of the

original content creators? With burgeoning legal debates over the use of copyrighted materials in AI training and the repercussions for the creators' IP, as already observed in recent US and UK lawsuits and an executive order to create AI safeguards to advance and govern the development of AI,⁷ it is critical to design and implement LLMs that not only consider a broad spectrum of inputs but also adhere to policy and ethics. Hence, as we integrate LLMs into innovation processes, we must commit to deliberately crafting and educating these systems to navigate the full extent of the problem and solution space, ensuring that the final conclusions are drawn with conscientious compliance, governing guidelines, and high moral standards.

Conclusion

In conclusion, as we continue to advance in the age of AI, it is clear that the path toward generating innovative solutions to complex problems lies not in humans or machines alone but in the collaboration between humans' novel responses and machines' valuable outputs. As the latest LLMs reap unprecedented gains in intelligence over prior models, their march forward underscores a pivotal moment for humanity to harness the strengths of human ingenuity: the unparalleled ability to imagine, craft, and evoke extreme, right-tail outcomes. Our study sheds light on a new, promising direction that has the potential to reshape the dynamics of organizational strategy and innovation.

⁷ See <https://llmlitigation.com/> and <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>.

References

- Abernathy WJ, Rosenbloom RS (1969) Parallel strategies in development projects. *Mgmt. Sci.* 15(10):B486–B505.
- Afuah A, Tucci CL (2023) Reflections on the 2022 AMR decade award: Crowdsourcing as a solution to distant search. *Acad. Mgmt. Rev.* 48(4):597–610.
- Agarwal N, Moehring A, Rajpurkar P, Salz T (2023) Combining human expertise with artificial intelligence: Experimental evidence from radiology. Working paper no. 31422, National Bureau of Economic Research.
- Agrawal A, Gans J, Goldfarb A (2018) *Prediction Machines: The Simple Economics of Artificial Intelligence* (Harvard Business Press, Boston).
- Amabile TM (1996) Creativity and innovation in organizations. Harvard Business School Background Note 396-239.
- Amazon. Amazon and Anthropic announce strategic collaboration to advance generative AI. Retrieved October 23, 2023, <https://www.aboutamazon.com/news/company-news/amazon-aws-anthropic-ai>.
- Anthony C, Bechky BA, Fayard AL (2023) “Collaborating” with AI: Taking a system view to explore the future of work. *Org. Sci.* 34(5):1672–1694.
- Anthropic. Model card and evaluations for Claude models. Retrieved July 8, 2023, <https://www-files.anthropic.com/production/images/Model-Card-Claude-2.pdf>.
- Argote, L, Miron-Spektor E (2011) Organizational learning: From experience to knowledge. *Org. Sci.* 22(5):1123–1137.
- Argyle L, Busby E, Fulda N, Gubler J, Rytting C, Wingate D (2023) Out of one, many: Using language models to simulate human samples. *Pol. Anal.* 31(3):337–351.
- Arnaud A, Schminke M (2012) The ethical climate and context of organizations: A comprehensive model. *Org. Sci.* 23(6):1767–1780.
- Arts S, Fleming L (2018) Paradise of novelty—or loss of human capital? Exploring new fields and inventive output. *Org. Sci.* 29(6):1074–1092.
- Baer M, Dirks KT, Nickerson JA (2013) Microfoundations of strategic problem formulation. *Strat. Mgmt. J.* 34(2):197–214.
- Barr PS, Stimpert JL, Huff AS (1992) Cognitive change, strategic action, and organizational renewal. *Strat. Mgmt. J.* 13(S1):15–36.
- Benjamin DJ, Brown SA, Shapiro JM (2013) Who is “behavioral”? Cognitive ability and anomalous preferences. *J. Eur. Econ. Assn.* 11(6):1231–1255.
- Berg J, Raj M, Seamans R (2023) Capturing value from artificial intelligence. *Acad. Mgmt. Disc.* Forthcoming.
- Bernstein IA, Zhang YV, Govil D, Majid I, Chang RT, Sun Y, Shue A, Chou JC, Schehlein E, Christopher KL, Groth SL, Ludwig C, Wang SY (2023) Comparison of ophthalmologist and large language model chatbot responses to online patient eye care questions. *JAMA Network Open* 6(8):e2330320.
- Bojinov I (2023) *Keep Your AI Projects on Track* (Harvard Business Review Press, Boston).
- Boudreau KJ, Guinan EC, Lakhani KR, Riedl, C (2016) Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Mgmt. Sci.* 62(10):2765–2783.
- Boudreau KJ, Lacetera N, Lakhani KR (2011) Incentives and problem uncertainty in innovation contests: An empirical analysis. *Mgmt. Sci.* 57(5):843–863.
- Boudreau KJ, Lakhani KR, Menietti M (2016). Performance responses to competition across skill levels in rank-order tournaments: Field evidence and implications for tournament design. *RAND J. Econ.* 47(1):140–165.

- Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E, Lee P, Lee YT, Li Y, Lundberg S, Nori H, Palangi H, Ribeiro MT, Zhang Y (2023) Sparks of artificial general intelligence: Early experiments with GPT-4 arXiv:2303.12712. <http://arxiv.org/abs/2303.12712>.
- Briot J-P, Hadjeres G, Pacht F-D (2017) Deep learning techniques for music generation. arXiv:1709.01620. <https://arxiv.org/abs/1709.01620>.
- Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler DM, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Chess B, Clark J, Berner C, McCandlish S, Radford A, Sutskever I, Amodei D (2020) Language models are few-shot learners. arXiv:2005.14165. <https://arxiv.org/abs/2005.14165>.
- Brynjolfsson E, Li D, Raymond L (2023) Generative AI at work. NBER Working Paper No. 31161, National Bureau of Economic Research, Cambridge. <https://doi.org/10.3386/w31161>.
- Caplin A, Dean M (2015) Revealed preference, rational inattention, and costly information acquisition. *Amer. Econ. Rev.* 105(7):2183–2203.
- Carlson N (2022) Differentiation in microenterprises. *Strat. Mgmt. J.* 44(5):1141–1167.
- Chakrabarty T, Laban P, Agarwal D, Muresan S, Wu C-S (2023) Art or artifice? Large language models and the false promise of creativity. arXiv:2309.14556. <https://arxiv.org/abs/2309.14556>.
- Che Y-K, Gale I (2003) Optimal design of research contests. *Amer. Econ. Rev.* 93(3):646–671.
- Chen M, Tworek J, Jun H, Yuan Q, Pinto HPDO, Kaplan J, Zaremba W et al. (2021) Evaluating large language models trained on code. arXiv:2107.03374. <https://arxiv.org/abs/2107.03374>.
- Cheng Z, Lee D, Tambe P (2022, March) InnoVAE: Generative AI for Understanding Patents and Innovation. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3868599.
- Chesbrough HW (2003) *Open Innovation: The New Imperative for Creating and Profiting from Technology* (Harvard Business Review Press, Boston).
- Choudhury P, Starr E, Agarwal R (2020) Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strat. Mgmt. J.* 41(8):1381–1411.
- Chowdhery A, Narang S, Devlin J, Bosma M, Mishra G, Roberts A, Fiedel N et al. (2022) PaLM: Scaling language modeling with pathways. arXiv:2204.02311. <https://arxiv.org/abs/2204.02311>.
- Christensen M, Knudsen T (2020) Division of roles and endogenous specialization. *Industrial and Corporate Change* 29(1):105–124.
- Christian MS, Ellis APJ (2011) Examining the effects of sleep deprivation on workplace deviance: A self-regulatory perspective. *Acad. Mgmt. J.* 54(5):913–934.
- Christiano P, Shlegeris B, Amodei D (2018) Supervising strong learners by amplifying weak experts. arXiv:1810.08575. <https://arxiv.org/abs/1810.08575>.
- Cockburn IM, Henderson R, Stern S (2019) The impact of artificial intelligence on Innovation. Agrawal A, Gans J, Goldfarb A, eds. *The Economics of Artificial Intelligence: An Agenda* (University of Chicago Press, Chicago) 115–148.
- Corley KG, Gioia DA (2011) Building theory about theory building: What constitutes a theoretical contribution? *Acad. Mgmt. Rev.* 36(1):12–32.
- Dahan E, Mendelson H (2001) An extreme-value model of concept testing. *Mgmt. Sci.* 47(1):102–116.
- Daugherty P, Wilson HJ (2018) *Human + Machine: Reimagining Work in the Age of AI* (Harvard Business Review Press, Boston).
- Dell’Acqua F, McFowland E, Mollick ER, Lifshitz-Assaf H, Kellogg K, Rajendran S, Kraymer L, Candelon F, Lakhani KR (2023) Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24-013. <https://doi.org/10.2139/ssrn.4573321>.
- Du Y, Li S, Torralba A, Tenenbaum JB, Mordatch I (2023) Improving factuality and reasoning in language models through multiagent debate. arXiv:2305.14325. <https://arxiv.org/abs/2305.14325>.
- Gaessler F, Piezunka H (2023) Training with AI: Evidence from chess computers. *Strat. Mgmt. J.* 44(11):2724–2750.

- Gavetti, G, Levinthal DA, Rivkin JW (2005) Strategy making in novel and complex worlds: The power of analogy. *Strat. Mgmt. J.* 26(8):691–712.
- Ghosh S, Wu A (2021) Iterative coordination and innovation: Prioritizing value over novelty. *Org. Sci.* Forthcoming.
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. Ghahramani Z, Welling M, Cortes C, Lawrence N, Weinberger KQ, eds. *Advances in Neural Information Processing Systems 27: 28th Annual Conference on Neural Information Processing Systems 2014* (Curran, Red Hook, NY).
- Hargadon A, Sutton RI (1997) Technology brokering and innovation in a product development firm. *Admin. Sci. Qtrly.* 42(4):716–749.
- Huber GP, Power DJ (1985) Retrospective reports of strategic-level managers: Guidelines for increasing their accuracy. *Strat. Mgmt. J.* 6(2):171–180.
- Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, Ishii E, Bang YJ, Madotto A, Fung P (2023) Survey of hallucination in natural language generation. *ACM Computing Surveys* 55(12):1–38.
- Fayard AL (2023) Making time for social innovation: How to interweave clock time and event time in open social innovation to nurture idea generation and social impact. *Org. Sci.* Forthcoming.
- Fleming L (2001) Recombinant uncertainty in technological search. *Mgmt. Sci.* 47(1):117–132.
- Fudenberg D, Liang A (2019). Predicting and understanding initial play. *Amer. Econ. Rev.* 109(12):4112–4141.
- Fudenberg D, Liang A (2020) Machine learning for evaluating and improving theories. *ACM SIGecom Exchanges* 18(1):4–11.
- Gelman A, Hill J (2006) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- Girotra K, Meincke L, Terwiesch C, Ulrich KT (2023) Ideas are dimes a dozen: Large language models for Idea generation in innovation. SSRN. <https://doi.org/10.2139/ssrn.4526071>.
- Goldenberg A, Gross JJ (2020) Digital emotion contagion. *Trends Cog. Sci.* 24(4):316–328.
- Gu J, Cho K, Li VOK (2017) Trainable greedy decoding for neural machine translation. arXiv:1702.02429. <https://arxiv.org/abs/1702.02429>.
- Harrison DA, Klein KJ (2007) 'What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Acad. Mgmt. Rev.* 32(4):1199–1228.
- Higgins I, Dyre B, Pal A, Burgess C, Xavier L, Botvinick M, Mohamed S, Lerchner A (2017). beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. ICLR conference paper.
- Jeppesen LB, Lakhani KR (2010) Marginality and problem-solving effectiveness in broadcast search. *Org. Sci.* 21(5):1016–1033.
- Jia N, Luo X, Fang Z, Liao C (2023) When and how artificial intelligence augments employee creativity. *Acad. Mgmt. J.* Forthcoming. <https://doi.org/10.5465/amj.2022.0426>.
- Jones BF (2009) The burden of knowledge and the “death of the Renaissance man”: Is innovation getting harder? *Rev. Econ. Stud.* 76(1):283–317.
- Jones TM, yan LV (1997) The link between ethical judgment and action in organizations: A moral approbation approach. *Org. Sci.* 8(6):663–680.
- Kahneman D, Lovallo D (1993) Timid choices and bold forecasts: A cognitive perspective on risk taking. *Mgmt. Sci.* 39(1):17–31.
- Kaplan S, Vakili K (2015) The double-edged sword of recombination in breakthrough innovation. *Strat. Mgmt. J.* 36(10):1435–1457. <https://doi.org/10.1002/smj.2294>
- Katila R, Ahuja, G (2002) Something old, something new: A longitudinal study of search behavior and new product introduction. *Acad. Mgmt. J.* 45(6):1183–1194.
- Kellogg KC, Valentine MA, Christin A (2020) Algorithms at work: The new contested terrain of control. *Acad. Mgmt. Ann.* 14(1):366–410.
- Kenny DA, Kashy DA, Cook WL (2006) *Dyadic Data Analysis*. New York: Guilford Publications.

- Kıcıman E, Ness R, Sharma A, Tan C (2023) Causal reasoning and large language models: Opening a new frontier for causality. arXiv:2305.00050. <https://arxiv.org/abs/2305.00050>.
- Kovács B, Sharkey A J (2014) The paradox of publicity. *Admin. Sci. Qtrly.* 59(1):1–33.
- Kuznetsova, A, Brockhoff PB, Christensen RHB (2017) lmerTest package: Tests in linear mixed effects models. *J. of Stat. Softw.* 82(13):1–26. <https://doi.org/10.18637/jss.v082.i13>.
- Lane JN, Ganguli I, Gaule P, Guinan E, Lakhani KR (2021) Engineering serendipity: When does knowledge sharing lead to knowledge production? *Strat. Mgmt. J.* 42(6):1215–1244.
- Levinthal DA, March JG (1993) The myopia of learning. *Strat. Mgmt. J.* 14(S2):95–112.
- Lerner JS, Li Y, Valdesolo P, Kassam KS (2015) Emotion and decision making. *Ann. Rev. Psych.* 66(1):799–823.
- Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW (2023) Ethics of large language models in medicine and medical research. *Lancet Digital Health* 5(6):e333–e335.
- Lifshitz-Assaf H (2018) Dismantling knowledge boundaries at NASA: The critical role of professional identity in open innovation. *Admin. Sci. Qtrly.* 63(4):746–782.
- Lou B, Wu L (2021) AI on drugs: Can artificial intelligence accelerate drug development? Evidence from a large-scale examination of bio-pharma firms. *MIS Qtrly.* 45(3):1451–1482.
- Manica M, Born J, Cadow J, Christofidellis D, Smith JR et al. (2023) Accelerating material design with the generative toolkit for scientific discovery. *NPJ Computational Materials* 9:69. <https://doi.org/10.1038/s41524-023-01028-1>.
- Marcus G (2018) Deep learning: A critical appraisal. arXiv:1801.00631. <https://arxiv.org/abs/1801.00631>.
- McCoy RT, Smolensky P, Linzen T, Gao J, Celikyilmaz A (2023) How much do language models copy from their training data? Evaluating linguistic novelty in text generation using RAVEN. *Trans. Assoc. Comp. Ling.* 11:652–670.
- Muehlhauser L, Salamon A (2012) Intelligence explosion: Evidence and import. Eden A, Moor J, Søraaker J, Steinhart E, eds. *Singularity Hypotheses. The Frontiers Collection* (Springer, Berlin).
- Microsoft. “Microsoft and OpenAI extend partnership.” Retrieved October 23, 2023, <https://blogs.microsoft.com/blog/2023/01/23/microsoftandopenaiextendpartnership/>.
- Min S, Lyu X, Holtzman A, Artetxe M, Lewis M, Hajishirzi H, Zettlemoyer L (2022) Rethinking the role of demonstrations: What makes in-context learning work? *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (Association for Computational Linguistics, Abu Dhabi) 11048–11064.
- Nadkarni S, Barr PS (2008) Environmental context, managerial cognition, and strategic action: An integrated view. *Strat. Mgmt. J.* 29(13):1395–1427.
- Nagle F, Teodoridis F (2019) Jack of all trades and Master of Knowledge: The role of diversification in New Distant Knowledge Integration. *Strat. Mgmt. J.* 41(1):55–85.
- Nickerson JA, Zenger TR (2004) A knowledge-based theory of the firm—The problem-solving perspective. *Org. Sci.* 15(6):617–632.
- Nonaka I (1994) A dynamic theory of organizational knowledge creation. *Org. Sci.* 5(1):14–37.
- Noy S, Zhang W (2023) Experimental evidence on the productivity effects of generative artificial intelligence. *Science* 381(6654):187–192.
- Ouyang L, Wu J, Jiang X, Almeida D, Wainwright CL, Mishkin P, Zhang C, Agarwal S, Slama K, Ray A, Schulman J, Hilton J, Kelton F, Miller L, Simens M, Askell A, Welinder P, Christiano P, Leike J, Lowe R (2022) Training language models to follow instructions with human feedback. arXiv:2203.02155. <https://arxiv.org/abs/2203.02155>.
- OpenAI. (2023). GPT-4 Technical Report. arXiv:2303.08774. <https://arxiv.org/abs/2303.08774>.
- Park S, Piezunka H, Dahlander L (2023). Coevolutionary lock-in in external search. *Acad. Mgmt. J.* Forthcoming. <https://doi.org/10.5465/amj.2022.0710>.
- Park YJ, Kaplan D, Ren Z, Hsu CW, Li C, Xu H, Li S, Li J (2023) Can ChatGPT be used to generate scientific hypotheses? arXiv:2304.12208. <https://arxiv.org/abs/2304.12208>.
- Perry-Smith JE, Shalley CE (2014) A social composition view of team creativity: The role of member nationality-heterogeneous ties outside of the team. *Org. Sci.* 25(5):1434–1452.

- Peterson JC, Bourgin DD, Agrawal M, Reichman D, Griffiths TL (2021) Using large-scale experiments and machine learning to discover theories of human decision-making. *Science* 372(6547):1209–1214.
- Piezunka H, Dahlander L (2015) Distant search, narrow attention: How crowding alters organizations' filtering of suggestions in crowdsourcing. *Acad. Mgmt. J.* 58(3):856–880.
- Piezunka H, Dahlander L (2019) Idea rejected, tie formed: Organizations' feedback on crowdsourced ideas. *Acad. Mgmt. J.* 62(2):503–530.
- Puranam P (2021) Level 0 vs Level 1 Self-organization: “Invisible” vs “Hidden” hand explanations of order in organizations. SSRN. <https://doi.org/10.2139/ssrn.3959868>.
- Radford A, Narasimhan K (2018) Improving Language Understanding by Generative Pre-Training.
- Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I (2019) Language models are unsupervised multitask learners. https://d4mucfpxsywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.
- Raisch S, Krakowski S (2021) Artificial intelligence and management: The automation–augmentation paradox. *Acad. Mgmt. Rev.* 46(1):192–210.
- Ray PR (2023) ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems* 3:121–154.
- Regenwetter L, Srivastava A, Gutfreund D, Ahmed F (2023) Beyond statistical similarity: Rethinking metrics for deep generative models in engineering Design. *Computer-Aided Design* 165, 103609.
- Russell SJ, Norvig P (2010) *Artificial Intelligence: A Modern Approach* (Pearson, Upper Saddle River, NJ).
- Singh J (2005) Collaborative networks as determinants of knowledge diffusion patterns. *Mgmt. Sci.* 51(5):756–770.
- Singh J, Fleming L (2010) Lone inventors as sources of breakthroughs: Myth or reality?. *Mgmt. Sci.* 56(1):41–56.
- Taylor CR (1995) Digging for golden carrots: An analysis of research tournaments. *Amer. Econ. Rev.* 85(4):872–890.
- Teodoridis F (2018) Understanding team knowledge production: The interrelated roles of technology and expertise. *Mgmt. Sci.* 64(8):3625–3648.
- Terwiesch C, Ulrich K (2009) *Innovation tournaments: Creating and selecting exceptional opportunities* (Harvard Business Press, Boston, MA).
- Touvron H, Lavril T, Izacard G, Martinet X, Lachaux M-A, Lacroix T, Rozière B, Goyal N, Hambro E, Azhar F, Rodriguez A, Joulin A, Grave E, Lample G (2023). LLaMA: Open and efficient foundation language models. arXiv:2302.13971. <https://arxiv.org/abs/2302.13971>.
- Townsend DM, Hunt RA, McMullen JS, Sarasvathy SD (2018) Uncertainty, knowledge problems, and entrepreneurial action. *Acad. Mgmt. Ann.* 12(2):659–687.
- Tyler BB, Steensma HK (1995) Evaluating technological collaborative opportunities: A cognitive modeling perspective. *Strat. Mgmt. J.* 16(S1):43–70.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2023) Attention is all you need. arXiv:1706.03762. <https://arxiv.org/abs/1706.03762>.
- Vicinanza P, Goldberg A, Srivastava SB (2022) A deep-learning model of prescient ideas demonstrates that they emerge from the periphery. *PNAS Nexus*. 2(1):pgac275. [https://doi: 10.1093/pnasnexus/pgac275](https://doi.org/10.1093/pnasnexus/pgac275).
- Vissa B, Greve HR, Chen WR (2010) Business group affiliation and firm search behavior in India: Responsiveness and focus of attention. *Org. Sci.* 21(3):696–712.
- von Hippel E, von Krogh G (2003) Open source software and the “private-collective” innovation model: Issues for organization science. *Org. Sci.* 14(2):209–223.
- Von Hippel E (2005) Democratizing innovation: The evolving phenomenon of user innovation. *Journal für Betriebswirtschaft* 55:63–78.
- Wasserman L (2004) *All of Statistics: A Concise Course in Statistical Inference* (Springer, New York).

- Wei J, Tay Y, Bommasani R, Raffel C, Zoph B, Borgeaud S, Yogatama D, Bosma M, Zhou D, Metzler D, Chi EH, Hashimoto T, Vinyals O, Liang P, Dean J, Fedus W (2022) Emergent abilities of large language models. *Transactions on Machine Learning Research*. <https://openreview.net/forum?id=yzkSU5zdwd>.
- Wiltermuth S, Gubler T, Pierce L (2023) Anchoring on Historical Round Number Reference Points: Evidence from Durable Goods Resale Prices. *Organization Science* 34(5):1839-1863.
- Wuchty S, Jones BF, Uzzi B (2007) The increasing dominance of teams in production of knowledge. *Science*, 316(5827):1036–1039.
- Yang L-C, Chou S-Y, Yang Y-H (2017) MidiNet: A convolutional generative adversarial network for symbolic-domain music generation. arXiv:1703.10847. <https://arxiv.org/abs/1703.10847>.
- Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov R, Le QV (2020) XLNet: Generalized autoregressive pretraining for language understanding. arXiv:1906.08237. <https://arxiv.org/abs/1906.08237>.

Table 1. Evaluator Summary Statistics (N = 300 evaluators)

<i>Screening Criteria</i>	
Level of interest	3.91 (0.82)
Knowledge test score	2.45 (1.16)
Work experience	4.68 (5.18)
<i>Demographics</i>	
Female	37.0%
Bachelor's or higher	61.7%
Employed or self-employed	86.0%
STEM degree	53.3%

Note. standard deviations reported in parentheses.

Table 2. Nested Mixed Effects Models of Evaluator Ratings of Novelty Rating on Solution Source (Human or AI)

<i>Dependent Variable: Novelty Rating</i>						
	(1)	(2)	(3)	(4)	(5)	(6)
AI Solution	-0.124*** (0.035)	-0.124*** (0.035)	-0.140*** (0.035)			
AI Instance (Multiple)				-0.209*** (0.038)	-0.209*** (0.038)	-0.217*** (0.039)
AI Instance (Single)				-0.039 (0.038)	-0.039 (0.038)	-0.056 (0.039)
Intercept	3.508*** (0.046)	3.381*** (0.180)	3.242*** (0.215)	3.508*** (0.046)	3.381*** (0.180)	3.262*** (0.215)
N	3900	3900	3900	3900	3900	3900
# blocks	18	18	18	18	18	18
# evaluators	300	300	300	300	300	300
Screening criteria	N	Y	Y	N	Y	Y
Controls	N	N	Y	N	N	Y
Log-Likelihood	-5430.30 df = 5	-5434.96 df = 8	-5439.89 df = 15	-5419.78 df = 6	-5424.45 df = 9	-5430.94 df = 16

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

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

Table 3. Nested Mixed Effects Models of Evaluator Ratings of Top Novelty Rating on Solution Source (Human or AI)

<i>Dependent Variable: Top Novelty Rating (0/1)</i>						
	(1)	(2)	(3)	(4)	(5)	(6)
AI Solution	-0.074*** (0.012)	-0.074*** (0.012)	-0.079*** (0.012)			
AI Instance (Multiple)				-0.088*** (0.014)	-0.088*** (0.014)	-0.091*** (0.014)
AI Instance (Single)				-0.059*** (0.014)	-0.059*** (0.014)	-0.065*** (0.014)
Intercept	0.208*** (0.015)	0.189*** (0.057)	0.176** (0.068)	0.208*** (0.015)	0.189*** (0.057)	0.179** (0.068)
N	3900	3900	3900	3900	3900	3900
# blocks	18	18	18	18	18	18
# evaluators	300	300	300	300	300	300
Screening criteria	N	Y	Y	N	Y	Y
Controls	N	N	Y	N	N	Y
Log-Likelihood	-1329.62 df = 5	-1338.23 df = 8	-1356.98 df = 15	-1330.04 df = 6	-1338.65 df = 9	-1358.06 df = 16

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

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

Table 4. Nested Mixed Effects Models of Evaluator Ratings of Value Rating on Solution Source (Human or AI)

	<i>Dependent Variable: Value Rating</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
AI Solution	0.171*** (0.025)	0.171*** (0.025)	0.152*** (0.026)			
AI Instance (Multiple)				0.160*** (0.028)	0.160*** (0.028)	0.148*** (0.028)
AI Instance (Single)				0.182*** (0.028)	0.182*** (0.028)	0.156*** (0.029)
Intercept	3.351*** (0.038)	3.080*** (0.172)	3.017*** (0.200)	3.351*** (0.038)	3.080*** (0.172)	3.018*** (0.200)
N	3900	3900	3900	3900	3900	3900
# blocks	18	18	18	18	18	18
# evaluators	300	300	300	300	300	300
Screening criteria	N	Y	Y	N	Y	Y
Controls	N	N	Y	N	N	Y
Log-Likelihood	-4294.88 df = 5	-4293.91 df = 8	-4288.06 df = 15	-4297.25 df = 6	-4296.28 df = 9	-4290.79 df = 16

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

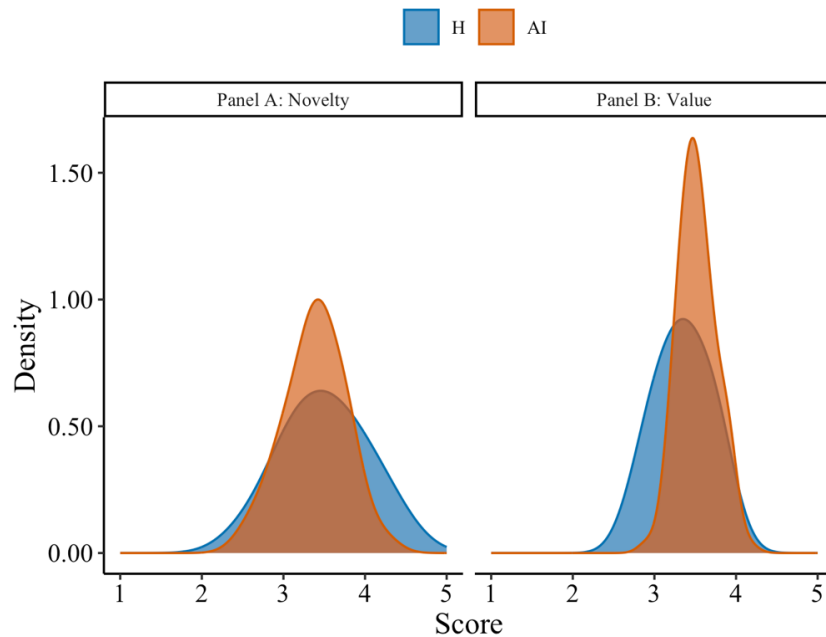
Notes. This table presents mixed-model (hierarchical linear modeling) results from evaluator ratings of solution novelty, with 300 evaluators nested in eighteen solution blocks. Models 2-3 and 5-6 include the screening criteria: Work Experience, Level of Interest, and Knowledge Test Score. Models 3 and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

Table 5. Nested Mixed Effects Models of Evaluator Ratings of Top Value Rating on Solution Source (Human or AI)

	<i>Dependent Variable: Top Value Rating (0/1)</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
AI Solution	0.019 (0.014)	0.019 (0.014)	0.012 (0.014)			
AI Instance (Multiple)				0.020 (0.016)	0.020 (0.016)	0.016 (0.016)
AI Instance (Single)				0.017 (0.016)	0.017 (0.016)	0.007 (0.016)
Intercept	0.268*** (0.019)	0.071 (0.085)	0.065 (0.102)	0.268*** (0.019)	0.071 (0.085)	0.064 (0.102)
N	3900	3900	3900	3900	3900	3900
# blocks	18	18	18	18	18	18
# evaluators	300	300	300	300	300	300
Screening criteria	N	Y	Y	N	Y	Y
Controls	N	N	Y	N	N	Y
Log-likelihood	-1975.78 df = 5	-1977.58 df = 8	-1990.34 df = 15	-1979.13 df = 6	-1980.93 df = 9	-1993.51 df = 16

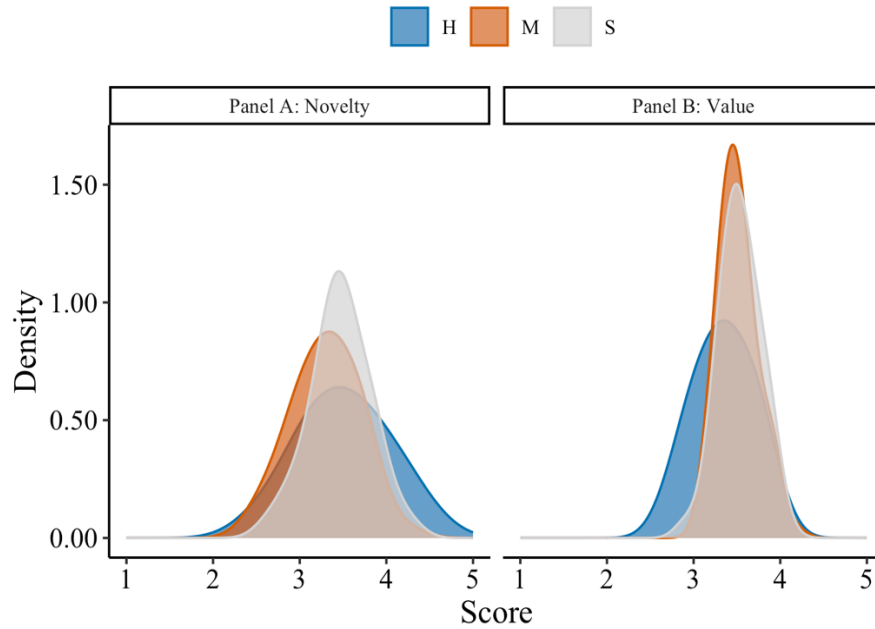
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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

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

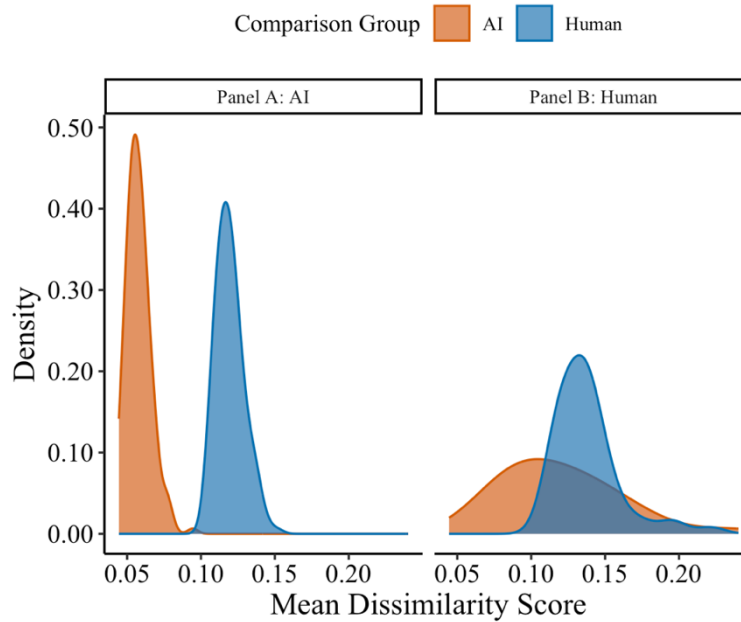
Note. The mean solution ratings for novelty and value are computed as the average of all evaluator scores assigned to a solution.

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



Note. The mean solution ratings for novelty and value are computed as the average of all evaluator scores assigned to a solution.

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



Appendix

Appendix A: Detailed Explanations Prompt Building Mechanisms and Model

As shown in Table A1, we use three prompt engineering levels to generate the AI solutions, comprising both multiple and single instance configurations for each level. This resulted in six distinct configurations by level and instance, corresponding to Levels 1S, 1M, 2S, 2M, 3S, and 3M.

- Level 1 uses the initial problem description for baseline comparison of human and AI solutions.
- Level 2 adds individual characteristics of the 125 human solvers, simulating the context of the original human crowd.
- Level 3 introduces expert persona prompting, leveraging GPT-4’s vast data to generate diverse, industry-specific solutions by mimicking expert personas from 23 industries.

Table A1. 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.
Multiple Instance	Each distinct instance of GPT-4 generates its solution independently from the same input prompt.	Intuitively, a multiple instance configuration mimics the concept of independent crowd solvers, as each instance of GPT-4 operates independently.
Single Instance	A single instance of GPT-4 generates multiple solutions successively, one at a time, with a differentiation instruction between successive responses.	Intuitively, a single instance configuration resembles an individual solver who iteratively proposes distinct ideas.

Table A2. Examples of Human and AI Problems and Solutions for Top Value by Configuration

Example Problem and Solution	Configuration
<p><u>Problem:</u> A significant portion of food waste is generated by restaurants, supermarkets, and households, which not only squanders resources but also adds to the growing problem of landfills. In particular, the disposal of food waste in landfills results in the generation of methane, a greenhouse gas twenty times more potent than carbon dioxide. Hence, finding an efficient and economic method to manage food waste reduces the need for new landfills and helps mitigate climate change.</p> <p><u>Solution:</u> To address this issue, I propose the creation of local bio-energy centers. The idea is to partner with restaurants, markets, and households to collect their organic waste. This waste will then be converted into fertilizer for community gardens and parks, as well as biogas, which can be used to generate electricity.</p> <p>First, there would be waste collection bins at strategic locations so businesses and households can conveniently dispose of their organic waste. The waste would be regularly collected and brought to the bio-energy centers for processing.</p> <p>The organic waste, once in the bio-energy center, will be treated in a process called anaerobic digestion. Anaerobic digestion is basically a series of biological processes in which microorganisms break down biodegradable material in the absence of oxygen. This process results in the production of biogas, which consists mainly of methane.</p> <p>The methane generated can be used as a renewable source of energy, to generate heat and electricity, while nutrient-rich digestate (a byproduct of the process) can be used as a low-cost, efficient fertilizer to promote local agriculture and horticulture.</p> <p>This solution provides multiple benefits: reduction in waste that would have gone to landfills; production of clean, renewable energy; providing organic, chemical-free fertilizers for agricultural purposes; and contributing to the local economy. The feasibility and scalability of this model have been proven in several European countries. The financial impact is expressed in both cost savings (as waste management costs are reduced) and income sourcing (sale of energy, fertilizer). The environmental benefits are also clear, as this process drastically reduces the release of harmful greenhouse gases that contribute to climate change.</p>	1M
<p><u>Problem:</u> The rapid depletion of natural resources, exacerbated by traditional linear economic practices, is a pressing issue. Consider the automobile industry, industrial equipment, metal products, and the construction sector. These industries, though economically significant, are notorious for contributing significantly to environmental degradation. The issues span from resource mining, energy-intensive manufacturing processes, and end-product disposal, often in landfills. There's a pressing need for innovative strategies to extend the lifecycle of products, reduce waste, and conserve natural resources.</p>	2M

<p><u>Solution:</u> My solution proposes the 'Resource Lifecycle Extension and Re-utilization Platform' (ReLEaRP). This would function as a multi-sector digital marketplace combining the aspects of sharing, leasing, reusing, repairing, refurbishing and recycling of products and materials, thus encapsulating the concept of a Circular Economy.</p> <p>In this ecosystem, businesses can list their waste materials or old equipment, which might be useful to other industries. For instance, the 'waste' metal from an automobile company could be valuable raw material for a construction firm that uses metal extensively. Businesses can also offer their idle assets for lease, fostering resource sharing.</p> <p>The platform would also facilitate repair and refurbish services. Businesses with impaired equipment can link with repair service providers to get them fixed rather than disposed of. Similarly, used products can be refurbished and re-introduced to the market. The platform can also engage recycling companies to convert waste materials to useful products.</p> <p>The effectiveness of such a system can create a ripple effect across businesses. It encourages waste minimization, promotes resource conservation, reduces production cost by using reclaimed materials and expands the potential for new revenue streams from leasing and selling refurbished products.</p> <p>Moreover, implementation appears feasible. The concept leverages existing digital marketplace models, can start in small local circles and be gradually scaled to bigger markets. Initially, the model can start with pilot sectors like automobile and construction, and once successful, the model can be replicated to other industries, promoting a circular economy at a larger scale.</p> <p>From an environmental standpoint, the extended product lifecycles, decreased waste, and reduced resource mining will lessen the negative environmental impact. Financially, businesses stand to save on raw material costs, generate revenues from idle assets and waste, and reduce landfill costs.</p> <p>Overall, the concept of 'Resource Lifecycle Extension and Re-utilization Platform' significantly contributes to ushering in a more sustainable and economically viable circular economy.</p>	
<p><u>Problem:</u> Despite emerging advancements in green technology and renewable energy, transportation and logistics remain heavily reliant on traditional, non-renewable energy sources. This is a pressing problem, particularly in a world where climate change is increasingly urgent. More specifically, the process of shipping and freight transportation continues to emit a significant amount of greenhouse gases. Today's non-circular economy largely rests upon the use of single-use containers and extended freight delivery routes, often crisscrossing locations unnecessarily due to suboptimal planning. Consequently, heavy costs for businesses, inefficient processes, and environmental damage proliferate. In essence, the problem relates to the emission-intensive, expensive and inefficient ways of transporting goods.</p> <p><u>Solution:</u> I propose a combination of intelligent route planning, reusable</p>	3M

<p>packaging and better use of data as a circular economy solution.</p> <p>Firstly, advanced AI systems can optimize delivery routes in real-time, increasing efficiency while reducing fuel consumption and emissions. Key to achieving this is incorporating machine learning algorithms and AI to effectively use data related to current traffic, weather, and delivery locations. Such systems can lead to fewer miles driven and a reduction in unnecessary idling, therefore reducing costs and emissions.</p> <p>Secondly, reusable packaging solutions can be adopted, directly minimizing waste produced by the industry. These durable containers can be recycled and used multiple times with their life cycle significantly extended. Also, returnable packing materials frequently occupy less space than their disposable counterparts when empty, leading to expanded transport efficiency.</p> <p>Lastly, using data to forecast demand and match it with supply in the industry can reduce the number of empty kilometers traveled by vehicles. By applying predictive analytics, we can determine demand trends which help in optimizing cargo load and reducing unnecessary trips.</p> <p>This model is not only adaptable and scalable to varying business sizes and needs but also capable of economic pay-off due to reduced fuel consumption and packing costs. Plus, it bears the potential to have a significant positive environmental impact. Utterly congruous with the circular economy principles, this proposition ensures longer product utility, optimized resource usage, efficient supply, and lesser wastage.</p>	
<p><u>Problem:</u> Electronic waste, or 'e-waste', is a growing concern. With new devices continuously entering the market, old electronics are too often discarded, often improperly. These devices can contain toxic materials hazardous to both environmental and human health. The existing recycling rate for e-waste is low, and valuable, finite resources present in these devices, such as rare metals, are not sufficiently recovered.</p> <p><u>Solution:</u> The proposed 'Electronics Lifecycle Optimization Initiative (ELOI)' would be based on these principles:</p> <ol style="list-style-type: none"> 1. Extended Producer Responsibility (EPR): Enforce stringent EPR regulations that require electronic manufacturers to manage the lifecycle of their products, including end-of-life disposal and recycling. 2. Design for Disassembly and Recycling: Encourage electronics manufacturers to design devices with their entire life cycle in mind. This means making it easy to disassemble devices for recycling and to use materials that can be efficiently recovered. 3. E-Waste Recycling Kiosks: Establish e-waste recycling kiosks in public spaces, stores, and community centers to facilitate convenient e-waste disposal. 	1S

<p>4. Urban Mining: Promote urban mining, the process of reclaiming compounds and elements from products, buildings, and waste, which includes e-waste recycling as a chief component, helping recover valuable and precious metals.</p> <p>The initiative's environmental benefits would come from reducing toxic e-waste, reducing the need for new raw material extraction, and reclaiming valuable resources. It creates financial value by generating a market for second-hand components and reclaimed materials. EPR regulations provide an incentive for manufacturers to become involved in the second-hand market, opening up new opportunities for revenue.</p> <p>With strong regulations and increasing awareness about e-waste, this initiative's feasibility is high. Its scalability extends to every locale with electronic consumers, effectively making it a global strategy. It satisfies all evaluation criteria, making it a strong contender for the circular economy challenge.</p>	
<p><u>Problem:</u> The construction industry in Asia produces a vast quantity of surplus materials - such as bricks, cement, wood, and metal - that are discarded after projects, leading to significant waste. The automobiles and industrial equipment sector generate a significant amount of scrap metal and used parts. With the existing linear 'create-use-discard' model, these valuable resources are often wasted, causing environmental harm and financial loss. The challenge here is creating a sustainable, circular solution that optimally utilizes these resources and minimizes waste.</p> <p><u>Solution:</u> I propose the 'Relove & Rebuild' initiative, a platform that connects construction companies, automakers, and industrial houses with smaller businesses or individuals who need these surplus materials or parts.</p> <p>'Builders Bay' in the platform would list excess materials from construction sites. Smaller builders, artists, DIY enthusiasts, can purchase them at discounted prices, preventing these materials from going to waste.</p> <p>'Revamp Garage' would list the scrap metal and used parts from automobiles and industrial equipment. Local repair shops, artisans, hobbyists can buy these parts to refurbish or to create new products.</p> <p>'Green Points' would be given for each transaction, encouraging participation and reinforcing sustainable behavior.</p> <p>By reducing waste, the initiative significantly lowers the environmental impact. Financially, savings are made by both parties - the seller recovers some costs from surplus, while the buyers gain access to cheaper resources.</p> <p>The solution requires a user-friendly digital platform and engagement from the industries. Once it gains traction, the model can be scaled across different regions, increasing its environmental and financial impacts.</p>	2S

<p>'Relove & Rebuild' presents a novel way to approach surplus materials, harnessing the power of the circular economy to transform waste into wealth, and fostering an ecosystem of sustainable practices in construction and manufacturing industries.</p>	
<p><u>Problem:</u> In the Construction Materials industry, concrete production poses a significant environmental challenge. Concrete's primary ingredient, cement, is produced from limestone, which requires high-temperature kilns and emits vast amounts of CO₂. Combined with the large volumes of concrete used worldwide, the cement industry accounts for about 7% of global CO₂ emissions. Managing these emissions is imperative for combating climate change and fostering a circular economy within the construction sector.</p> <p><u>Solution:</u> my proposition encapsulates an innovative approach termed "Green Concrete Revolution," composed of "Low-Carbon Concrete Production," "Concrete Recycling," and "Carbon Capture and Usage."</p> <p>The first strategy, "Low-Carbon Concrete Production," revolves around producing concrete with less dependence on Portland Cement, the main cause of high emissions in concrete production. This goal could be accomplished by replacing a part of cement with industrial by-products like fly ash or slag. Researchers are also exploring new types of cement that require lower kiln temperatures, thereby reducing energy use and emissions. New concrete compositions, like self-healing concrete or ultra-high-performance concrete, despite their higher initial cost, require less frequent replacement and repair, extending their lifecycle and reducing emissions in the long run.</p> <p>Secondly, "Concrete Recycling" becomes imperative. When demolishing concrete structures, recycling and reusing the concrete waste in new projects should be standardized. It can reduce the demand for virgin aggregate and cement, translating into lesser resource extraction and emissions.</p> <p>Lastly, "Carbon Capture and Usage" refers to the integration of carbon capture technology in cement plants. The captured CO₂ can be injected into fresh concrete, where it becomes locked once the concrete hardens, effectively making this material a carbon sink.</p> <p>The aggregate effect of these solutions produces a substantial environmental benefit, significantly lowering the industry's carbon emissions. The financial impact is two-fold; although low-carbon concrete and carbon capture technology may be more expensive initially, they offer the potential for operational cost savings due to lesser energy use, and the long-term lifecycle cost savings provide an economic edge. Moreover, with climate policies tightening worldwide, embracing these green initiatives can help avoid future regulatory restrictions or penalties. In terms of feasibility, the technology for these solutions is available and is being progressively improved. The scalability of this approach is global due to the universal use of concrete in construction. Adopting the "Green Concrete Revolution" can fundamentally shift the Construction Materials industry towards a sustainable and circular economic model.</p>	<p>3S</p>
<p><u>Problem:</u> Roughly 1/3 of all food produced for human consumption goes to waste. 87% is caused by</p>	<p>H</p>

<p>perishable items. Waste caused by perishable products in supermarkets and online grocery stores has a negative effect on the environment in addition to the financial loss.</p> <p><u>Solution:</u> To help reduce the waste caused by perishable products in supermarkets and online grocery stores, a food waste solution has been developed by Wasteless. It is a proprietary dynamic pricing algorithm for products based on their expiration date. This allows retailers to deploy AI to mark off prices on perishable products. The system is managing the inventory of fresh products in relation to expiring dates and the consumers' attitudes towards them. Wasteless's pricing engine employs a branch of machine learning called –'Reinforcement Learning–^a. This allows the pricing engine to quickly learn how consumers respond to dynamic pricing so it can then find the optimal discounting policy. The retailers will achieve a remarkable increase in revenue by optimizing pricing and incentivizing their shoppers to buy products that would otherwise go to waste. Suppliers will leverage prices to consumers to better match supply with demand, reduce overstock situations, and ensure constant product availability at the point of sale.</p>	
--	--

Table A3. Examples of Human and AI Problems and Solutions for Top Novelty by Configuration

Example Problem and Solution	Configuration
<p><u>Problem:</u> A significant portion of food waste is generated by restaurants, supermarkets, and households, which not only squanders resources but also adds to the growing problem of landfills. In particular, the disposal of food waste in landfills results in the generation of methane, a greenhouse gas twenty times more potent than carbon dioxide. Hence, finding an efficient and economic method to manage food waste reduces the need for new landfills and helps mitigate climate change.</p> <p><u>Solution:</u> To address this issue, I propose the creation of local bio-energy centers. The idea is to partner with restaurants, markets, and households to collect their organic waste. This waste will then be converted into fertilizer for community gardens and parks, as well as biogas, which can be used to generate electricity.</p> <p>First, there would be waste collection bins at strategic locations so businesses and households can conveniently dispose of their organic waste. The waste would be regularly collected and brought to the bio-energy centers for processing.</p> <p>The organic waste, once in the bio-energy center, will be treated in a process called anaerobic digestion. Anaerobic digestion is basically a series of biological processes in which microorganisms break down biodegradable material in the absence of oxygen. This process results in the production of biogas, which consists mainly of methane.</p> <p>The methane generated can be used as a renewable source of energy, to generate heat and electricity, while nutrient-rich digestate (a byproduct of the process) can be used as a low-cost, efficient fertilizer to promote local agriculture and horticulture.</p> <p>This solution provides multiple benefits: reduction in waste that would have gone to landfills; production of clean, renewable energy; providing organic, chemical-free fertilizers for agricultural purposes; and contributing to the local economy. The feasibility and scalability of this model have been proven in several European</p>	1M

<p>countries. The financial impact is expressed in both cost savings (as waste management costs are reduced) and income sourcing (sale of energy, fertilizer). The environmental benefits are also clear, as this process drastically reduces the release of harmful greenhouse gases that contribute to climate change.</p>	
<p><u>Problem:</u> Africa is fraught with challenges pertaining to food and beverage waste, inefficient packaging methods, and lack of waste management techniques. Every year, million tons of plastic and other packaging materials end up in landfills, deteriorating the environment and posing a threat to human health. Simultaneously, a significant percentage of food and beverages are wasted due to inadequate storage and distribution systems. This overall inefficiency leads to economic losses and emissions of greenhouse gases contributing to climate change.</p> <p><u>Solution:</u> We propose a solution titled "Zero Waste and Nutrient Circularity in Food, Beverages & Packaging through Bio-Conversion Technology." The idea centers around the use of black soldier flies (<i>Hermetia illucens</i>), which can consume different types of organic waste, including food and beverage waste. This process will convert waste into larvae, which in turn can be used as a protein source for animal feed. On the other hand, the residual waste can be used as a nutrient-rich biofertilizer to replenish agricultural lands.</p> <p>In terms of packaging, we envision transitioning to zero-waste packaging solutions. One such approach is the introduction of edible, biodegradable packaging material made from natural substances such as seaweed.</p> <p>In conjunction, a widespread, "Return, Reward, and Recycle." initiative would encourage consumers to return their used packaging for responsible recycling or composting, incentivized through discounts or other benefits.</p> <p>The solution's novelty lies in integrating biological conversion and eco-friendly packaging into a unified waste management system. It has a substantial environmental impact, reducing plastic waste and emissions, while the financial impact is realized through cost savings in waste management, creation of saleable products, and sustainable brand image enhancement. It is feasible with public participation and scalable through public-private partnerships and legislative support.</p>	2M
<p><u>Problem:</u> The beverages industry, in which I specialize, makes a substantial contribution to global waste – particularly plastic waste – since packaging is predominantly single-use. Despite recycling efforts, a substantial percentage of these bottles, cans, and containers end up in landfills or the natural environment. The problem is multifaceted and contributes to several global challenges: resource depletion due to virgin plastic production, waste management difficulties, and pollution, not to mention the carbon emissions associated with production, transportation, and recycling. Additionally, this linear production model puts financial pressure on businesses as fluctuations in raw material prices can heavily impact costs. The industry is in need of a transformative, circular solution.</p> <p><u>Solution:</u> The solution lies in a "Beverage-as-a-Service" (BaaS) model that leverages IoT technology for the reuse of beverage containers and drives customer engagement in the circular economy. Instead of the traditional, linear 'produce-consume-dispose' model, containers (bottles, cans) would be designed for reusability, embedded with a chip that tracks their lifecycle. Upon purchase, customers pay a small deposit – they</p>	3M

<p>are essentially leasing the container.</p> <p>When the container is empty, customers return it to a collection point (grocery stores, vending machines, collection stations). The IoT chip then communicates the container's return, and the deposit is refunded to the customer digitally. Cleaned and refilled, the containers then loop back into the market – no need for recycling or waste treatment.</p> <p>Simultaneously, the digital trail created by the IoT chips will enable data collection on consumption habits. For businesses, this means data-driven decisions and personalized marketing. Even more, cost savings as fluctuations in raw material prices will have less impact, reducing financial risk. Plus, the deposit system will encourage a high return rate, ensuring a steady supply of reusable containers.</p> <p>This is a scalable solution that can start with select locations and products before expanding. It creates a closed-loop system, thus minimizing environmental impact by eliminating waste and reducing production need. While certain infrastructure investments are necessary, such as cleaning and refill systems and IoT technology, the long-term benefits in cost savings and sustainability make this a feasible solution for the beverage industry.</p>	
<p><u>Problem:</u> Single-use meal containers and packaging in food delivery and take-away services generate massive amounts of waste. Most of these containers are made from non-recyclable or difficult-to-recycle materials due to food contamination issues. This ultimately exacerbates problems related to waste disposal and resource depletion. While some delivery services and restaurants have started to provide more sustainable packaging options, the scale of the issue requires more comprehensive and effective solutions.</p> <p><u>Solution:</u> A 'Reusable Food Container Service (RFCS)' could offer a sustainable and innovative approach to this.</p> <ol style="list-style-type: none"> 1. Durable Containers: Collaborate with restaurants and food delivery services to introduce durable, reusable, and standardized meal containers. These containers, made of safe and long-lasting materials, would come in different standard sizes that can accommodate various types of food. 2. Deposit Model: Customers would pay a small refundable deposit for the reusable container at the point of ordering their meal. Alternatively, subscribers to the RFCS could obtain a set of containers to use for multiple food orders. 3. Collection and Cleaning: After use, dishes could be returned in designated collection points or directly to partnered restaurants/delivery services during the next order. The containers would then be professionally cleaned in a central location, ready for reuse in line with hygiene standards. 4. Awareness: Promote the importance of sustainability in food delivery habits and personalize it - show users how many single-use containers they have saved from waste by choosing RFCS. <p>The RFCS offers a unique approach that extends the idea of reusable shopping bags and water bottles to the widespread practice of food take-away and delivery. It</p>	1S

<p>contributes to reducing packaging waste, which is a significant environmental problem in urban areas worldwide. It also provides a model for delivering economic benefits through saving on disposable packaging costs and creating revenue from deposits and subscription fees. The main challenges for feasibility include establishing efficient cleaning and distribution logistics, ensuring hygiene standards, and getting buy-in from businesses and customers. However, the popularity of take-away services and the increasing awareness around packaging waste suggests that the scalability of this concept could be wide-reaching. The RFCS has the potential to revolutionize the way we package and consume take-out food, turning a linear process into a cyclical one that safeguards our planet and resources.</p>	
<p><u>Problem:</u> In a number of African cities and towns, a significant amount of food and beverages is wasted as a result of mismanagement, lack of proper storage facilities, and logistical challenges in the supply chain. Simultaneously, the vast usage of non-reusable and non-recyclable packaging materials contributes to growing littering issues and landfills. Many households and small businesses lack the knowledge or resources to manage these issues effectively, thereby leading to considerable environmental and socio-economic challenges, such as pollution, waste of valuable resources, added stress on landfills and waste management infrastructure, and lost economic opportunities.</p> <p><u>Solution:</u> To address these challenges, I propose the development of an innovative 'Smart Food Waste & Packaging Management System', focusing on the principles of the circular economy. This system would essentially comprise an integrated network of IoT-enabled 'Smart Bins' for households and businesses, a centrally coordinated collection and redistribution system, a waste-to-energy micro-scale facility, and an educational mobile app platform.</p> <p>The Smart Bins connected to Wi-Fi, would segregate food waste and packaging materials and alert the central system when full, enabling planned and timely pickups. The collected food waste would be redirected through two paths based on their conditions - safely edible items could be donated to local food banks, while spoiled food items would be processed at a micro-scale waste-to-energy facility to generate bioenergy. The collected packaging waste would be sorted and recycled.</p> <p>Simultaneously, an interactive mobile application would educate users about proper waste segregation, the concept of 'food rescue', recycling options for packaging materials, and the benefits of reducing food and packaging waste. Consumers could also be incentivized through a reward system within the app to promote proactive participation.</p> <p>This solution would massively reduce food waste and packaging littering, curtail the load on landfills, generate bioenergy, and foster sustainable behavior among the consumer base. Financially, it has potential to create revenue from the sale of bioenergy and recycled packaging materials, while the reduction in waste collection and landfilling costs for municipalities also makes it an attractive prospect. This system might require initial investment, partnerships with tech companies for app development and IoT setup, and collaboration with municipalities for implementation. However, once proven successful, it could be scaled up across different cities throughout Africa. The proposed system stands out in its novelty due to its integrated approach towards managing food waste and packaging, its use of smart technology for effective waste management, and its focus on consumer education - all working synergistically to revolutionize waste management in Africa.</p>	2S

<p><u>Problem:</u> The 'beyond Big 4' metals sector faces challenges related to the tailings issue. Mining for metals like cobalt, nickel, or rare earths often result in large quantities of leftover rock and processed material, known as tailings. These tailings are stored in dams but pose severe environmental risks, including groundwater contamination and dam failures, leading to catastrophic spills. It's critical to manage these tailings both for environmental protection and for repurposing these waste materials into valuable resources.</p> <p><u>Solution:</u> I suggest the implementation of a comprehensive "Tailings Transformation Strategy" based on "Innovative Tailings Reprocessing" and "Stable Storage Measures."</p> <p>"Innovative Tailings Reprocessing" involves developing technologies to efficiently extract residual valuable metals from tailings. Advances in extraction techniques, including bioleaching or hydrometallurgical methods, can make it feasible to recover these metals that would otherwise go to waste. Additionally, tailings can also be repurposed for other uses such as construction material, thereby minimizing waste.</p> <p>"Stable Storage Measures" focus on investing in the infrastructure and techniques to securely store tailings. This might involve more secure dam designs, real-time monitoring systems, or even elimination of tailings dams by shifting towards dry stacking methods.</p> <p>The "Tailings Transformation Strategy" would dramatically mitigate the environmental risks associated with tailings while also creating new sources of metallic resources, reducing the need for further raw material extraction. On the financial side, the recovered metals could add to revenue streams, potentially offsetting the costs of improved storage measures. As for feasibility, it relies on advancing technologies, some of which are already in use to a certain extent. Considering the global mining industry's scale, the potential for scalability is substantial. In these ways, the "Tailings Transformation Strategy" can enable the metals sector's significant strides towards a circular economy.</p>	3S
<p><u>Problem:</u> Since the construction industry is the largest user of coal (used to fire the bricks) in the country and a significant cause of air pollution, it also leads to the loss of fertile topsoil. Therefore, I came up with an ingenious solution to this pertinent problem by creating bricks, made from foundry dust and waste plastic!</p> <p><u>Solution:</u> To overcome this ecological problem, we came up with the bricks, made from foundry dust and waste plastic. There is no use of water in the making of these bricks. Also, more water is saved during the construction process, as the walls built with these bricks do not need to be cured with water. Also, the best part of this idea is that while 70 percent of the total sand can be reused, the other 30 percent, in the form of foundry dust, is too fine to be used again.</p> <p>Also, with enhanced technology we evolve this idea in making interlocking bricks, which essentially work like Lego blocks.</p>	H

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, 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 (D³) (<https://d3.harvard.edu/>) at Harvard to source the most innovative and cutting-edge circular economy solutions for the business world.

Circular Economy is a simple idea.

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

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

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 D³ 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 D³ 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: Technical Details of Inference Mechanism of LLMs

The inference stage in LLMs is the phase where the model generates text based on the input provided. This stage follows the pre-training and fine-tuning phases and relies on the Transformer architecture that underpins the model.⁸

The inference process encompasses the following steps:

1. **Input Tokenization:** The input text is tokenized into subwords or tokens using a tokenizer trained on the same corpus as the language model.
2. **Token Embedding:** Tokens are converted into numerical vectors, known as embeddings, which capture semantic and syntactic information.
3. **Positional Encoding:** To provide information about the sequential order of the tokens, positional encodings are added to the embeddings since the Transformer architecture does not inherently understand the sequential nature of the text data.
4. **Transformer Processing:**
 - a. The embeddings pass through multiple layers of the Transformer, each consisting of self-attention mechanisms (see B.1 for details) and feed-forward neural networks.
 - b. Self-attention (see B.1 for details) allows the model to weigh the importance of different parts of the input sequence when generating each token in the output sequence.
 - c. The feed-forward networks apply further transformations to the attention-weighted embeddings.
5. **Output Token Generation:** After tokenizing and encoding a prompt, this leaves a block of data representing our input as the machine understands it, including meanings, positions, and relationships between words. The model uses the final layer's output to estimate the probability distribution over the next token. GPT-4 employs a sampling strategy that calculates the probability distribution of the next token based on the context provided by the input sequence and the internal representations learned during training.

B.1 Self-Attention Mechanism

The benefits of self-attention for language processing increase as the model scales. Simply put, it allows LLMs to take context from beyond sentence boundaries, giving the model a greater understanding of how and when a word is used. The idea is formalized with a *self-attention function*, which takes as input a sequence of initial token embeddings and outputs a sequence of new token embeddings that allow the initial embeddings to interact. Let $(p^0_{d,1}, p^0_{d,2}, \dots, p^0_{d,N})$ be the initial embeddings that make up a document. The new embedding at each position n is given by,

$$p^1_{d,n} = \sum_{n'=1}^{N_d} w_{(d,n),n'} p^0_{d,n'} \text{ where } \sum_{n'=1}^{N_d} w_{(d,n),n'} = 1.$$

That is, each embedding in the transformed sequence is itself a weighted average of the embeddings in the initial sequence. The non-negative attention weights $w_{(d,n),n'}$, which are estimated during model training,

⁸ We direct our technical readers to find more details on how transformer generates outputs, using an encoder-decoder framework, multi-headed self-attention, as well as positional encoding here: <http://jalammr.github.io/illustrated-transformer/>.

determine which pairs of (potentially distant) tokens interact to form each context-sensitive word embedding in the final document representation.⁹ We refer our technical readers to Phung & Hutter (2022) for formal algorithms for Transformers.¹⁰

B.2 Decoding Strategies and Sampling Methods

Several sampling methods can be used to select the next word from the probability distribution. It is unclear which method GPT-4 uses precisely:

- Greedy Sampling: Chooses the token with the highest probability, leading to deterministic outputs.
- Random Sampling: Selects a token randomly based on the probability distribution, allowing for varied outputs.
- Top-k Sampling: Restricts the sampling pool to the top-k most likely tokens, balancing variety and coherence.
- Top-p (Nucleus) Sampling: Chooses from a subset of tokens that cumulatively make up to a certain probability p , focusing on high-probability tokens while maintaining diversity.

Once the next token is sampled, the model continues the process autoregressively, generating one token at a time and feeding the updated sequence back into the model until a termination condition is met, such as the end-of-sequence token or a specified maximum length. It is worth noting that previous studies have shown that custom decoding methods could significantly improve a language model’s output for a specific task.¹¹ Despite the potential shown in tailored decoding techniques, there seems to be a declining trend in their popularity. This may be attributed to two primary factors: 1) the increasing propensity for close-sourced models, which are less amenable to user-defined decoding adaptations, and 2) the improvements in baseline performance of pre-trained LLMs, which diminish the perceived need for such customizations.

Nevertheless, decoding adjustments could serve as an avenue for enhancing the generation of less common, or “long-tail,” solutions. Two ways to generate more nuanced and varied outputs from LLMs is using simple modifications, such as adjustment of temperature settings or the manipulation of top-k/top-p parameters. These exploratory steps could pave the way for potentially unlocking new capabilities within these systems.

The “temperature” hyperparameter influences the randomness or “creativity” of the model’s outputs. When generating text, the model calculates a probability distribution over possible following words. The temperature modifies the sampling from this distribution through a softmax calculation. At a higher temperature value, the distribution is flatter, rendering the output more random as it considers a broader range of word possibilities. Conversely, a lower temperature results in sharper distribution, leading to more deterministic outputs, wherein the model is more inclined to pick the most probable next word. For this study, we set the temperature at GPT-4 API’s default value.

Top-k sampling is a method where the model restricts its choice of the next word to the k most likely options. The value of k determines the breadth of the model’s consideration set; a smaller k leads to a higher probability that the model will select a more common or expected word, resulting in text that is typically more coherent but less varied. A larger k value gives the model leeway to pick less probable words, thereby increasing novelty and variation in the output but potentially at the cost of coherence and predictability.

Top-p, or nucleus sampling, takes a different approach by choosing from a dynamic set of options. Rather than selecting from the top k possibilities, it selects from the smallest set of words whose cumulative

⁹ <https://www.annualreviews.org/doi/pdf/10.1146/annurev-economics-082222-074352>.

¹⁰ <https://arxiv.org/pdf/2207.09238.pdf>.

¹¹ <https://arxiv.org/pdf/2110.07640.pdf>.

probability exceeds the threshold p . This means the model considers a range of words just enough to sum up to the probability p , allowing for dynamic and context-dependent variation in the number of words considered. With a lower p value, the model's outputs are more focused and less random, similar to having a lower temperature or smaller k . As p increases, the model can sample from a wider array of words, thus injecting more randomness and diversity into the generated text, akin to increasing the temperature or k value.

In practice, adjusting top- k and top- p can help balance between the generation of common, highly probable text and more diverse or surprising outputs. By tuning these parameters, one can calibrate the LLM to produce outputs that range from safe and predictable to novel and wide-ranging, thus enhancing the generation of less common, "long-tail" solutions.

B.3 Output Generation

The output generation is the culmination of the inference stage, where the sequence of predicted tokens is converted back into human-readable text. The model's ability to generate fluent and contextually appropriate text is a direct result of the complex interaction between its learned parameters and the inference-time sampling strategies.

Appendix C: Logistic Regression Analysis of AI and Human Solutions

Table C1. Nested Mixed Effects Logistic Models of Evaluator Ratings of Top Novelty Rating on Solution Sources (Human or AI)

<i>Dependent Variable: Top Novelty Rating (Logistic)</i>						
	(1)	(2)	(3)	(4)	(5)	(6)
AI Solution	-0.624*** (0.107)	-0.624*** (0.107)	-0.644*** (0.108)			
AI Instance (Multiple)				-0.778*** (0.125)	-0.779*** (0.125)	-0.785*** (0.125)
AI Instance (Single)				-0.484*** (0.120)	-0.484*** (0.120)	-0.510*** (0.121)
Work experience		-0.035+ (0.021)	-0.035 (0.021)		-0.035+ (0.021)	-0.035 (0.021)
Level of interest		0.174 (0.116)	0.180 (0.115)		0.175 (0.116)	0.180 (0.115)
Knowledge test score		-0.144+ (0.082)	-0.142+ (0.084)		-0.144+ (0.082)	-0.142+ (0.084)
Intercept	-1.603*** (0.133)	-1.772*** (0.518)	-1.897** (0.613)	-1.605*** (0.133)	-1.775*** (0.519)	-1.873** (0.614)
N	3900	3900	3900	3900	3900	3900
# blocks	18	18	18	18	18	18
# evaluators	300	300	300	300	300	300
Controls	N	N	Y	N	N	Y
Log-Likelihood	-1483.24 df = 4	-1479.09 df = 7	-1474.08 df = 14	-1479.86 df = 5	-1475.71 df = 8	-1471.22 df = 15
R2 Marg.	0.015	0.031	0.045	0.018	0.035	0.048
R2 Cond.	0.370	0.372	0.372	0.374	0.375	0.374

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

(1) (2) (3) (4) (5) (6)

Notes. This table presents mixed effects logistic model results from evaluator ratings of solution top novelty dummy, with 300 evaluators nested in eighteen solution blocks.¹² Models 3 and 6 include the following covariates: Gender, Highest Level of Education, Major, Employment Status, Cohort Session, and Solution Word Count. Standard errors are in parentheses. The number of points per axis for evaluating the adaptive Gauss-Hermite approximation to the log-likelihood is set to zero for convergence.

Table C2. Nested Mixed Effects Logistic Models of Evaluator Ratings of Top Value Rating on Solution Source (Human or AI)

<i>Dependent Variable: Top Value Rating (Logistic)</i>						
	(1)	(2)	(3)	(4)	(5)	(6)
AI Solution	0.128 (0.101)	0.129 (0.101)	0.091 (0.102)			
AI Instance (Multiple)				0.140 (0.111)	0.140 (0.111)	0.119 (0.112)
AI Instance (Single)				0.117 (0.111)	0.117 (0.111)	0.061 (0.113)
Work experience		-0.023 (0.022)	-0.024 (0.022)		-0.023 (0.022)	-0.024 (0.022)
Level of interest		0.480*** (0.133)	0.484*** (0.132)		0.480*** (0.133)	0.484*** (0.132)
Knowledge test score		-0.182+ (0.095)	-0.179+ (0.096)		-0.182+ (0.095)	-0.179+ (0.096)
Intercept	-1.368*** (0.136)	-2.700*** (0.594)	-2.710*** (0.702)	-1.368*** (0.136)	-2.700*** (0.594)	-2.718*** (0.702)
N	3900	3900	3900	3900	3900	3900
# blocks	18	18	18	18	18	18
# evaluators	300	300	300	300	300	300
Controls	N	N	Y	N	N	Y
Log-Likelihood	-1924.69 df = 4	-1914.09 df = 7	-1904.31 df = 14	-1924.67 df = 5	-1914.06 df = 8	-1904.12 df = 15
R2 Marg.	0.001	0.042	0.066	0.001	0.042	0.067
R2 Cond.	0.476	0.481	0.483	0.485	0.482	0.484

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

Notes. This table presents mixed effects logistic model results from evaluator ratings of solution top value dummy, with 300 evaluators nested in eighteen solution blocks. Models 3 and 6 include the following covariates: Gender, Highest Level of Education, Major, Employment Status, Cohort Session, and Solution Word Count. Standard errors are in parentheses. The number of points per axis for evaluating the adaptive Gauss-Hermite approximation to the log-likelihood is set to zero for convergence.

Appendix D: Additional Text Analysis of AI and Human Solutions

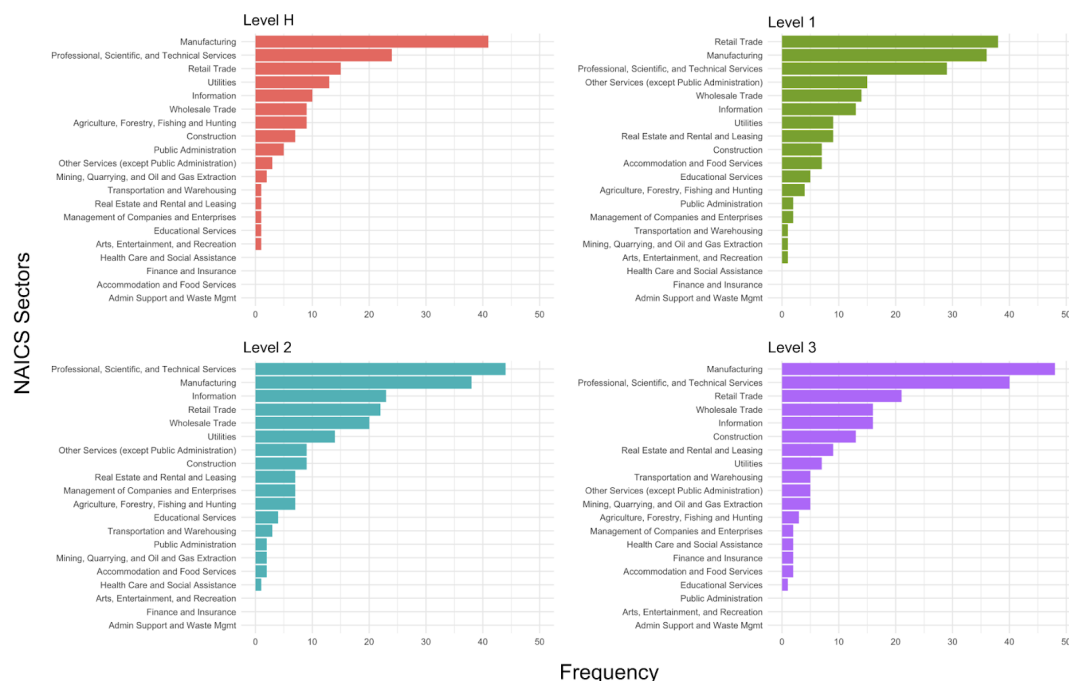
D.1 ChatGPT Results of Industry Classifications

To further explore the rich-text content of our human- and AI-generated solutions, we used OpenAI's GPT-4 to classify them 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

¹² <https://cran.r-project.org/web/packages/lme4/lme4.pdf>.

of application. Figure C1 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 D1. Comparisons of Industry Classifications By Human and AI Levels



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 equitability index within a level that contains k total sectors composing of proportion p of sector i is measured as:

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

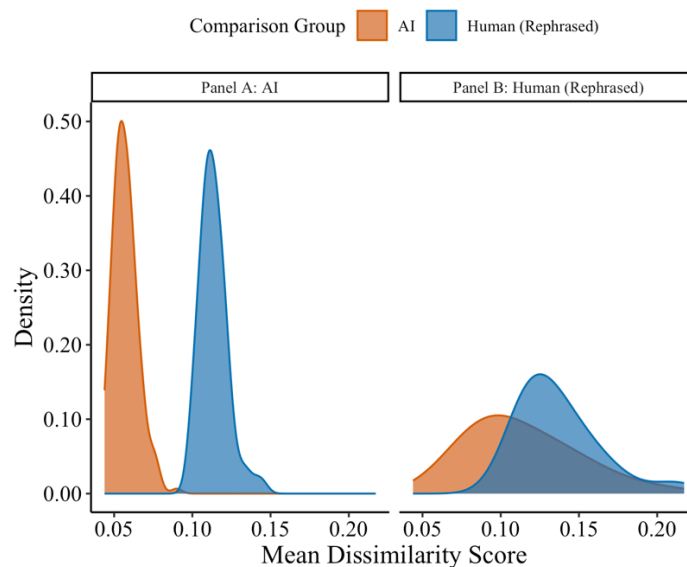
We calculated Level Human = 0.793, AI Level 1 = 0.820, AI Level 2 = 0.824, and AI Level 3 = 0.799, which suggests that AI and human levels generated similar degrees of diversity of industry applications.

D.2 Considerations on the Limitations of BERT Embeddings for AI and Human Semantic Diversity Measurement

Studies have engaged with both crowdsourced and GPT-based methodologies for generating a breadth of ideas, aiming to gauge the extent of the diversity of generated ideas. Notwithstanding, rigorous testing of various similarity metrics grounded in embedding techniques has unveiled that language embeddings may misinterpret the same idea presented in variant linguistic forms as disparate entities. To account for these linguistic discrepancies between human language, we passed the human solutions through the GPT-4 model while preserving the core idea of the solutions using the prompt, “*Rewrite this in good English while keeping the same content and ideas.*” We then recomputed the mean dissimilarity scores for the GPT-4 rephrased human solutions and the GPT-4 solutions. This first step ensures that we are using a consistent model to rearticulate ideas generated by human crowds vs. AI. Our hope is that this strategy could standardize linguistic expressions, thereby allowing the dissimilarity metric to more accurately

reflect conceptual rather than phrasal diversity. As shown in Figure D2, we found the distribution of mean dissimilarity score between the GPT-4-rephrased version of human solutions and AI solutions does not differ substantially from Figure 3, which reassures that the visual evidence of diverse semantics shown may not be attributed to language use differences between human and AI only, although we still do not rule out the possibility that differing linguistic expressions of identical solutions are still assessed as dissimilar.

Figure D2. Density plot distributions of the mean dissimilarity score by solution source. Comparison made within- and across-AI (Panel A) or GPT-4-rephrased-human (Panel B) sources.



D.3 Principal Component Analysis

To further exploit the rich text of our solutions, we transformed the text into BERT word embeddings and performed principal component analysis (PCA), a statistical technique for dimensionality reduction, to project the text of the solutions onto 2-dimensional space. PCA reduces the dimensionality of the text data, which can assist with plotting, visualizing, and identifying patterns in the data. Figures D3-D5 showcase that the AI solutions appear to cluster more centrally, while human solutions exhibit a broader spread, suggesting greater variance in the PCA space. This spatial distribution indicates the underlying diversity intrinsic to humans, compared to AI-generated solutions. Moreover, although there is a degree of overlap, each level occupies a relatively distinguishable region in the PCA space, hinting at underlying differences in the characteristics across sources of the solutions.

Figure D3. PCA projections by human and AI

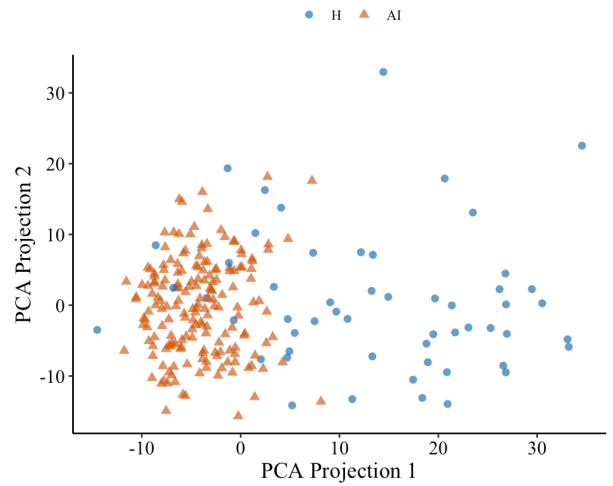


Figure D4. PCA projections by human and AI instances (M/S)

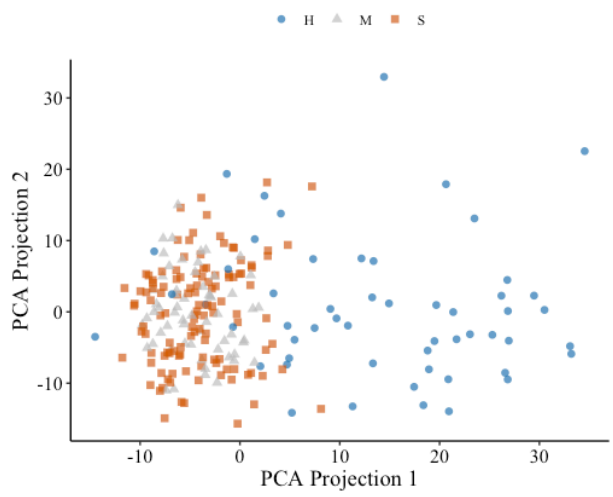
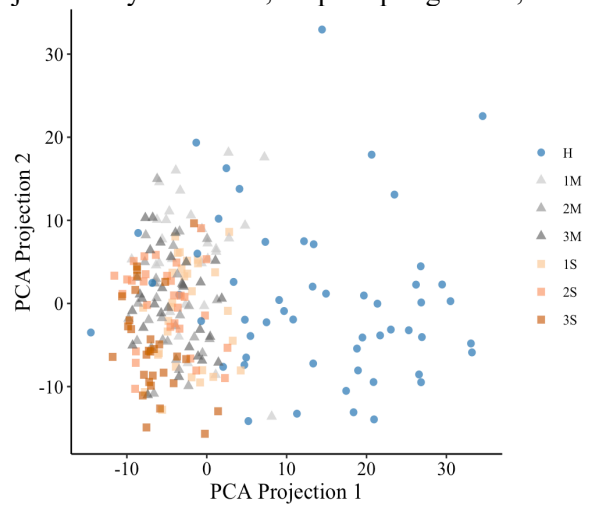


Figure D5. PCA projections by all human, AI prompting levels, and model configurations



Lastly, we color-coded the aggregated novelty and value ratings of each human- and AI-generated solutions on the PCA plots. The spread and density of points in Figure D6 suggest that highly novel ideas are scattered, arising from more diverse regions, most of which originate from humans. Figure D7, on the

other hand, color-codes aggregated value ratings and suggests that high-value solutions are more confined to a specific area within the PCA projections, indicating that value may tend to concentrate more in certain regions of the solution space than novelty.

Figure D6. PCA projections colored by novelty

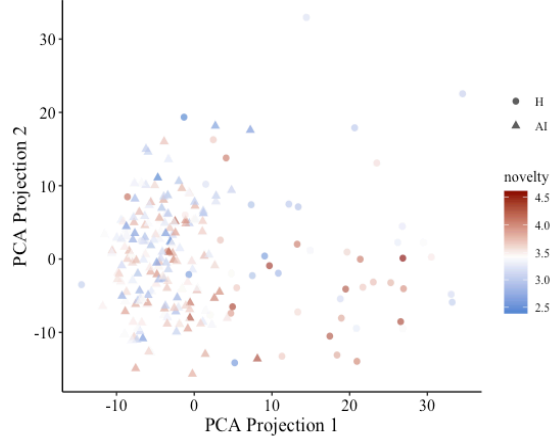
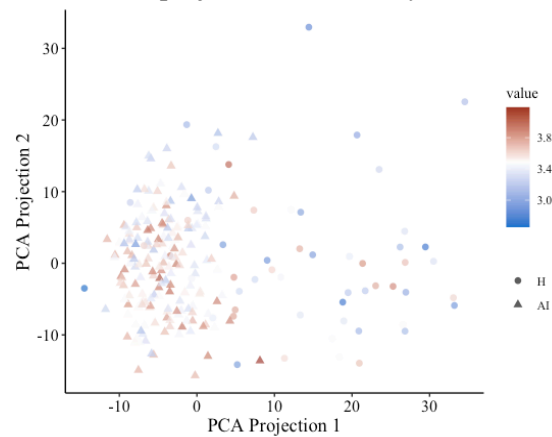


Figure D7. PCA projections colored by value



Appendix E: Survey Materials

E.1 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)

- o 45-54 (5)
- o 55-64 (6)
- o 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? (1 ... > 20 years) (6)
	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)	○	○	▼ less than 1 year (1 ... > 20 years (6))
Construction Materials (7)	○	○	▼ less than 1 year (1 ... > 20 years (6))
Consumer Electronics (8)	○	○	▼ less than 1 year (1 ... > 20 years (6))
Containers & Packaging (9)	○	○	▼ less than 1 year (1 ... > 20 years (6))
Cosmetics (10)	○	○	▼ less than 1 year (1 ... > 20 years (6))
Electrical Equipment (11)	○	○	▼ less than 1 year (1 ... > 20 years (6))
Financials (13)	○	○	▼ less than 1 year (1 ... > 20 years (6))
Food (14)	○	○	▼ less than 1 year (1 ... > 20 years (6))
Forest Products (15)	○	○	▼ less than 1 year (1 ... > 20 years (6))
Health Care Products (16)	○	○	▼ less than 1 year (1 ... > 20 years (6))
Home Furnishings (17)	○	○	▼ less than 1 year (1 ... > 20 years (6))
Household Appliances (18)	○	○	▼ 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)

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)

End of Block: Congrats

Display This Question:

*If Congratulations! You have been selected to participate in the evaluation task. We expect the foll...
= 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."

E.2 Evaluation Survey Instructions and Demographic Information

Start of Block: Consent

Q00 **Information**

The following is a short summary of this study to help you decide whether to be a part of this study. More detailed information is listed later in this form.

Why am I being invited to take part in a screening for this research study?

We invite you to take part in this study screening because you are over 18 years old and reside in the United States.

What should I know about a research study?

Someone will explain this research study to you.

Whether or not you take part is up to you.

Your participation is completely voluntary.

You can choose not to take part.

You can agree to take part and later change your mind.

Your decision will not be held against you.

Your refusal to participate will not result in any consequences or any loss of benefits that you are otherwise entitled to receive.

You can ask all the questions you want before you decide.

Why is this research being done?

The goal of this study is to understand how people evaluate the creativity of crowdsourced solutions.

Crowdsourcing leverages many diverse perspectives to improve the likelihood of getting high-value solutions. We want to understand how people evaluate these solutions.

How long will the research last and what will I need to do?

We expect that you will be in this research study for up to 30 minutes. You will be asked to evaluate 13 solutions to a crowdsourcing challenge using an evaluation framework. You will be asked to evaluate these solutions based on multiple criteria including novelty, feasibility, and impact). After this task, you will be asked to answer a short demographics questionnaire.

Is there any way being in this study could be bad for me?

We don't believe there are any risks from participating in this research.

Will being in this study help me in any way?

There are no benefits to you from your taking part in this research. We cannot promise any benefits to others from your taking part in this research. However, possible benefits to others include creating frameworks to more successfully evaluate high-value solutions to problems.

What happens if I do not want to be in this research?

Participation in research is completely voluntary. You can decide to participate, not participate, or discontinue participation at any time without penalty or loss of benefits to which you are otherwise entitled. Your alternative to participating in this research study is to not participate.

Detailed Information

The following is more detailed information about this study in addition to the information listed above.

What happens if I say yes, but I change my mind later?

You can leave the research at any time; it will not be held against you. Any data you created will be destroyed and not used for research.

Is there any way being in this study could be bad for me? (Detailed Risks)

We will do our best to protect your data during storage and when they are shared. However, there remains a possibility that someone could identify you. There is also the possibility that people who are not supposed to might access your data and samples. In either case, we cannot reduce the risk to zero.

If I take part in this research, how will my privacy be protected? What happens to the information you collect?

Efforts will be made to limit the use and disclosure of your Personal Information, including name and email if provided, to people who have a need to review this information. We cannot promise complete secrecy. Organizations that may inspect and copy your information include the IRB and other representatives of this organization.

Some of your data from the screening survey will be merged with your data from the main study upon completion of the main study tasks. If you do not agree to take part in the main study or leave the study early, we will destroy your screening data and any data created for the main study.

If identifiers are removed from your identifiable private information that are collected during this research, that information could be used for future research studies or distributed to another investigator for future research studies without your additional informed consent.

Compensation

If you agree to take part in this research study, you will receive a base pay of \$12. Additionally, you will have the opportunity to earn \$13 in bonuses for each solution where your rating is aligned with the consensus or the mode rating among all evaluators rating the same solution. In other words, for each solution where your rating aligns with the mode rating of the other evaluators, you will receive an extra \$1 of compensation, for a total of \$13 across the 13 solutions you will be asked to rate. The maximum total compensation you will receive is \$25.

Who can I talk to?

If you have questions, concerns, or complaints, or think the research has hurt you, talk to the research

team at <redacted>.

This research has been reviewed and approved by [redacted] Institutional Review Board (“IRB”). You may talk to them at [redacted] if:

- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.
- You have questions about your rights as a research subject.
- You want to get information or provide input about this research.

Do you consent to participate in this study?

- Yes (1)
- No (2)

End of Block: Consent

Start of Block: Instructions

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).

End of Block: Instructions

Start of Block: evaluation_0 - Jul 10, 2023

[13 problem and solution pairs each followed by the matrix box below]

	1 (Poor) (1)	2 (Below average) (2)	3 (Average) (3)	4 (Above average) (4)	5 (Excellent) (5)
Novelty (How different is it from existing solutions?) (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Feasibility and Scalability of Implementation (How likely is it to succeed and how scalable is it?) (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Environmental Impact (How much does it benefit the planet?) (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Financial Impact (What financial value can it create for businesses?) (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quality (Based on the four criteria above, what is the overall quality of the solution?) (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

End of Block: evaluation_0 - Jul 10, 2023

Start of Block: evaluation_1 - Jul 10, 2023

End of Block: evaluation_1 - Jul 10, 2023

....

....

Start of Block: evaluation_17 - Jul 10, 2023

End of Block: evaluation_17 - Jul 10, 2023

Start of Block: Demographics

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)

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.

End of Block: End