Competition in Creative Tasks, Gender Gaps and the Mitigating Role of AI: Experimental Evidence

Federico Atzori, Luca Corazzini, Valeria Maggian, Filippo Pavesi, Massimo Scotti

ISSN 1827-3580

Working Papers 2026 **No. 25**

Edizioni Ca'Foscari

Università Ca'Foscari Venezia | Department of **Economics**

2026 — No. 25

Competition in Creative Tasks, Gender Gaps and the Mitigating Role of AI: Experimental Evidence

Federico Atzori
CIMEO, Sapienza University,
Rome, Italy
federico.atzori@unimib.it

Luca Corazzini
University of Milan – Bicocca, Italy
luca.corazzini@unimib.it

Valeria Maggian
Ca' Foscari University of Venice, Italy
valeria.maggian@unive.it

Filippo Pavesi
LIUC University, Castellanza, Italy
fpavesi@liuc.it

Massimo Scotti
LIUC University, Castellanza, Italy
mscotti@liuc.it

Federico Atzori	→ Sapienza University
Luca Corazzini	→ University of Milan – Bicocca
Valeria Maggian	→ Ca' Foscari University of Venice
Filippo Pavesi	→ LIUC University
Massimo Scotti	→ LIUC University

Abstract We examine gender differences in competitiveness when performance depends on creativity, which is central to many high-skilled jobs, and investigate how generative AI may affect them. Our study is based on a laboratory experiment in which participants complete a creative task and then choose whether to enter a tournament compensation scheme, with access to ChatGPT varying across treatments. While men and women achieve similar levels of creative performance, they differ in competition entry, with women displaying a lower willingness to compete in environments without AI assistance. This gender gap in competition becomes smaller when AI is available. To investigate the mechanisms behind this pattern, we construct a measure of underconfidence based on the difference between participants' beliefs about their relative performance and their actual rank. We show that underconfidence predicts competition entry and that, besides improving creative performance, AI is associated with lower underconfidence, particularly among women.

Keywords Gender Gap in Competition; Creativity; Generative AI; Laboratory Experiment

JEL Codes C91, J16, J24, O33

Acknowledgements We thank Thomas Bassetti, Maria Bigoni, Antonio Cabrales, Piergiuseppe Morone, Stefano Verzillo, and the participants in the workshop "Experimental Approaches to Gender Inequality: Stereotypes, Norms, and Women's Achievements" in Bologna, the "Behavioral Economics for Sustainability" summer school in Rome, and seminars at the JRC's CC-ME Competence Centre on Microeconomic Evaluation in Ispra and at the University of Hamburg for their useful suggestions and comments. Financial support from the project "Gender Stereotypes and Women's Achievements (GENS)," PRIN - 2022 N. 20229RB8HY is gratefully acknowledged. We are in debt to Andrea Albarea, Chiara Grigoletti, and all the staff of the VERALabEx for helping run the experiment. The experimental protocol received ethics approval from the Ethical Committee of Ca' Foscari University of Venice (n. 2/2025). During the preparation of this work, the authors used the *GPT-4.2* API which was integrated into the *oTree* interface as a writing support tool for participants in the laboratory experiment. Additional minor editorial assistance, limited to text refinement, was provided using *ChatGPT 5.2*. After using these tools, the authors carefully reviewed and edited the content as needed and take full responsibility for the content of the published article.

Data availability statement All data, *STATA* do files, and replication codes are available at https://data.mendeley.com/datasets/77bg529jbx/1.

Address for correspondence:
Luca Corazzini
Department of Economics,
Management, and Statistics
University of Milan – Bicocca
Piazza dell'Ateneo Nuovo, 1
20126 Milan (MI), Italy
e-mail: luca.corazzini@unimib.it

Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta
San Giobbe, 30121, Venice, Italy

The *Working Papers* series
is available only online at
https://edizionicafoscari.it/it/edizioni4/riviste/dipartimento-di-economia-working-papers/.
For editorial correspondence,
please contact: wp.dse@unive.it

This Working Paper is published under the auspices of the Department of Economics of Ca' Foscari University of Venice. The opinions expressed herein are those of the authors and do not necessarily reflect those of the Department. The *Working Papers* series is intended to disseminate preliminary or incomplete research and to encourage discussion and feedback. When citing this paper, its provisional nature should be taken into account. This Working Paper may subsequently be revised and submitted for publication.

Università Ca'Foscari Venezia

1 Introduction

The gender gap in labor market outcomes, spanning wages, promotion rates, and representation in upper-level positions, is one of the most discussed issues in the economic debate (Blau and Kahn, 2017). Among the mechanisms proposed to explain these disparities, gender differences in willingness to compete have received particular attention. A large literature, initiated by the seminal experiment of Niederle and Vesterlund (2007), documents that women are less likely than men to select into competition, even when gender differences in performance are absent. This gap in competition entry has indeed been shown to predict educational and labor-market choices (Buser et al., 2014; Reuben et al., 2017; Buser et al., 2026) and is considered an important driver of gender differences in career progression. Most of this evidence, however, concerns competition in structured tasks, such as arithmetic or verbal ability tests, with little attention to settings in which performance depends on creativity.

Yet creativity is increasingly regarded as one of the most valuable skills in modern labor markets and a key determinant of career success, particularly in managerial, entrepreneurial, and leadership positions, where generating novel and valuable ideas is central (Anderson et al., 2014; Katz et al., 2022). Whether these gender differences extend to creative environments remains, however, largely unexplored. To fill this gap, we conducted a pre-registered laboratory experiment in which participants performed a creative task under different incentive schemes (competitive and non-competitive). Following Charness and Grieco (2019), we define creativity as the production of novel and useful ideas and adopt the same open-ended creative writing task used in their study. In particular, participants were asked to write a short composition in response to a broad prompt, without a predefined solution. Crucially, after completing the creative task under an assigned incentive scheme, participants are given the opportunity to choose whether their payment is determined by a competitive or a non-competitive scheme, enabling us to observe sorting into competition.

At the same time, the rapid diffusion of generative AI in both workplace practices and everyday activities (Brynjolfsson et al., 2025; Noy and Zhang, 2023) fundamentally changes the context in which competition in creative tasks takes place, having the potential to reshape performance and, ultimately, gender disparities in labor-market outcomes. Part of the concern relates to unequal access to and adoption of AI tools, with emerging evidence suggesting that women are less likely to use these technologies (Aldasoro et al., 2024; Otis et al., 2024; Carvajal et al., 2025). However, unequal adoption is only one potential source of gender disparities in AI-mediated environments. Recent studies indicate that AI assistance can influence evaluative processes and strategic judgment (Cristofaro et al., 2026), while exposure to AI-generated creative content can shape individuals' self-perceptions (Reich and Teeny, 2026). To examine whether generative AI modifies creative performance and competition entry differently for women and men, participants in our laboratory experiment perform the creative task under two treatments, either with or without access to ChatGPT, while incentive schemes are orthogonally varied.

Following the laboratory experiment, participants' creative performance is assessed by independent external judges recruited through a separate online experiment on Prolific.[1] Judges are blind to all participant characteristics and experimental conditions, and their evaluations determine both participants' monetary rewards and our measure of creative performance.

Our results are threefold. First, we find no evidence of gender differences in creative performance. Access to generative AI improves performance on average, and these gains are

entirely driven by participants who actively engage with the AI assistant rather than by its mere availability. Second, despite performing similarly, women remain significantly less likely than men to enter competition in the absence of AI. This gender gap is no longer statistically significant when AI is available. Finally, we provide evidence consistent with a confidence-based mechanism. We show that underconfidence predicts competition entry and that AI is associated with lower underconfidence, particularly among women. Although our analysis does not establish a formal mediation mechanism, these findings suggest that generative AI may reduce gender differences in willingness to compete primarily by influencing individuals' beliefs about their relative performance rather than by altering objective performance.

More broadly, our findings contribute to the growing debate on the labor-market consequences of generative AI, highlighting creativity as an important dimension through which generative AI may affect gender inequality in the labor market. Indeed, understanding how AI affects workers' decision to enter competition is essential for evaluating its broader implications for gender inequality in the labor market.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the experimental design and procedures and Section 4 introduces the predictions. Section 5 presents the main results, while Section 6 investigates the mechanisms underlying our findings. Section 7 concludes.

2 Literature Review

Since the seminal paper by Niederle and Vesterlund (2007), researchers have investigated gender differences in willingness to compete as one of the reasons for the observed gender gap in the labor market. This literature has been followed by many applications in the laboratory (see Niederle and Vesterlund, 2011 and Markowsky and Beblo, 2022 for a more complete review) and in the field (Delfgaauw et al., 2013; Flory et al., 2015; Hoyer et al., 2020). Using a novel experimental paradigm in which participants choose between alternative compensation schemes when performing an arithmetic task, Niederle and Vesterlund (2007) show that women are significantly less likely than men to enter competition. They attribute this gender gap primarily to differences in confidence and preferences for competition, with men tending to be more overconfident than women. The role of confidence is further supported by Kamas and Preston (2012). Consistent with this view, exposure to female competitive role models has been shown to increase competition entry among high-performing women (Meier et al., 2024). van Veldhuizen (2022) further suggests that gender differences in confidence, together with differences in risk attitudes, are sufficient to explain the observed gender gap in competition entry, the latter mechanism having also been emphasized by Dohmen et al. (2011).

Importantly, recent evidence suggests that the existence and magnitude of gender differences in competition entry depend crucially on the nature of the task being performed. In a meta-analysis, Markowsky and Beblo (2022) show that gender differences in competition entry are substantially smaller – and in some cases disappear – when verbal tasks are considered. However, previous studies have almost exclusively examined structured tasks, such as arithmetic or verbal ability tests (Shurchkov, 2012). This limitation is particularly relevant as creativity is widely regarded as one of the most valuable skills in modern labor markets and a key determinant of future workplace success (Anderson et al., 2014; Katz et al., 2022).

Within economics, creativity has, however, mainly been studied in relation to incentive schemes rather than competition decisions. Specifically, while standard economic theory

1 The use of independent judges to assess creativity is consistent with Amabile's (1982) consensual assessment approach and is commonly adopted in studies using open-ended creative tasks (Bohren et al., 2024; Charness and Grieco, 2019; Doshi and Hauser, 2024). Although our definition of creativity encompasses both novelty and usefulness, we do not ask judges to assess these dimensions separately. Instead, given our focus on overall creative performance, we follow Charness and Grieco (2019) in asking judges to provide an overall evaluation of each composition's creativity.

predicts that stronger incentives should increase effort and improve outcomes (Prendergast, 1999; Lazear, 2000), this might not be as straightforward when considering tasks that require creativity, originality, and problem solving. Performance in these domains is more likely to involve idea generation and open-ended search processes, making it less clear whether stronger monetary incentives necessarily translate into better outcomes. If creative performance relies on enjoyment and autonomy, monetary rewards may crowd out intrinsic motivation (Amabile et al., 1990; Amabile, 1996; Deci et al., 1999). In particular, Charness and Grieco (2019) show that while incentives can increase effort, they do not necessarily translate into higher creative performance, especially when creativity is difficult to evaluate objectively. This finding is in line with a broader body of evidence reviewed by Attanasi et al. (2021), who document the mixed effects of incentives on creativity across economics, psychology, and management.

Another reason why monetary incentives may fail to improve creative performance is that they can generate pressure and anxiety. Ariely et al. (2009) show that high-powered performance-based incentives may reduce performance in cognitively demanding tasks because of "choking under pressure". Consistent with this interpretation, Erat and Gneezy (2016) find that although both piece-rate and tournament incentives induce greater effort relative to a no-incentive benchmark, neither improves creativity. Importantly, competitive incentives appear particularly detrimental for women, whose creative performance declines under tournament schemes. These findings raise broader questions about gender differences in competitive environments, not only in terms of creative performance but also in terms of the willingness to enter competition in the first place, a gap that our paper addresses.

The increasing role of generative AI in creative work and in the labor market as a whole makes this issue particularly timely. Specifically, a recent report by UNESCO, the OECD, and the Inter-American Development Bank highlights both opportunities and risks associated with AI adoption, emphasizing that these technologies may either mitigate or exacerbate existing gender gaps (UNESCO et al., 2022).

Considering the role of generative AI in creative tasks, a growing literature in Economics and Management has focused primarily on two questions: whether generative AI can match or outperform human creativity and whether it enhances human creativity when used as a complementary tool (Charness and Grieco, 2026; Doshi and Hauser, 2024; Bohren et al., 2024). In this respect, Holzner et al. (2025) conduct a meta-analysis of 28 recent studies and show that: i) there is no significant difference in creative performance between GenAI systems and humans; ii) humans collaborating with GenAI significantly outperform individuals working without assistance; iii) however, GenAI reduces the diversity of ideas generated in human–AI collaborations. Overall, their findings suggest that generative AI acts primarily as an augmentative technology that complements, rather than replaces, human creativity.

If AI affects creative performance, it may also affect the incentives to enter competitive environments. This possibility is particularly relevant given the evidence discussed above showing that competition entry depends on both task characteristics and gender. Despite its potentially important implications, to the best of our knowledge, only Mourelatos et al. (2024) investigate how AI might affect men's and women's willingness to compete, finding that providing access to ChatGPT eliminates an eleven-percentage-point gender gap in competition entry that emerges in the absence of AI assistance. However, their analysis builds on the experimental framework of Niederle and Vesterlund (2007) and focuses on a mathematical task, whereas generative AI is expected to have its greatest impact in creative work.

Building on this, our paper makes three main contributions. First, we focus on a creative task rather than a mathematical task, a setting in which complementarities between humans and generative AI are likely to be particularly salient. Second, we investigate both performance and competition choices under alternative incentive schemes. Third, we explicitly examine whether AI reshapes gender differences in creative performance and willingness to compete, and through which mechanisms.

3 Experimental Design

The study consists of two complementary components. First, a laboratory experiment in which participants complete a creative writing task under alternative incentive schemes with or without AI assistance. Second, an online experiment in which independent evaluators assess the creativity of the texts produced in the laboratory. Since participants' payoff depends on these evaluations, final payments are determined only after the survey experiment has been completed[2].

3.1 Laboratory Experiment

The laboratory experiment was divided into three stages. Information about Stages 2 and 3 was disclosed sequentially, with instructions for each stage provided only after completion of the previous one.[3] Final payments consisted of participants' earnings from Stage 3 plus the earnings from one randomly selected stage between Stages 1 and 2.

In Stage 1, participants were asked to complete a creative writing task based on the following prompt, originally proposed by Charness and Grieco (2019): "*If you had the talent to invent things just by thinking of them, what would you create?*" Participants were encouraged to spend no more than 15 minutes on the task and to write approximately 150 words, although both indications were not binding. Participants were informed that their text would be evaluated on a scale from 1 to 10 for creativity by four external judges, randomly selected from a pool of approximately 2,000 Italian-speaking Prolific participants, and that their earnings would depend on the average score assigned by the judges.

We implement a 2×2 experimental design, varying the incentive scheme under which subjects performed the task (Piece Rate vs Tournament) and whether Artificial Intelligence was available or not (AI vs No-AI). The four resulting treatments are denoted *PR*, *T*, *PR&AI*, and *T&AI*.

In the *Piece Rate treatment* (PR), participants received 0.6 euros multiplied by the average creativity score assigned by external evaluators. In the *Tournament treatment* (T), participants were randomly assigned to groups of four participants composed of two women and two men. Group composition was common knowledge before the beginning of the task. The participant obtaining the highest creativity score within the group received 2.4 euros multiplied by the average creativity score assigned by external evaluators, while all other group members received zero. The payment multiplier was chosen to equalize expected earnings across the two incentive schemes.

Participants assigned to AI treatments (*PR&AI* and *T&AI*) were provided with access to a generative AI assistant implemented through the GPT-4.2 API and integrated directly into the experimental interface. Participants could interact freely with the AI assistant during the task, with no restrictions on the number of prompts submitted. Responses were provided in real time. Participants were informed that all interactions with the AI assistant were recorded for research purposes only and would not be used to train future versions of the model[4].

In Stage 2, after completing the creative task, participants had to choose between the Piece Rate and Tournament compensation schemes. If this stage was selected for payment, earnings depended on i) the compensation scheme chosen and ii) the creativity score obtained in Stage 1. Before making their decision, participants were informed that tournament entrants would be matched with three other participants in gender-balanced groups. Specifically, in treatments T and $T\&AI$, tournament groups were formed using participants from the same experimental session. In contrast, participants in treatments PR and $PR\&AI$ were informed that, should they choose the tournament scheme, they would be

2 The experimental protocol received ethics approval from the Ethical Committee of Ca' Foscari University of Venice (n. 2/2025), and was preregistered prior to data collection. The preregistration is available at `AsPredicted`.
3 The English version of the instructions, originally in Italian, is reported in the online Appendix A.
4 A detailed analysis of communication patterns, prompting strategies, and interaction styles between participants and the AI assistant is provided in Atzori et al. (2026).

matched with three participants from the corresponding tournament treatment (T and $T\&AI$, respectively).

Finally, in Stage 3, we elicited participants' beliefs about their own performance. Participants reported both their expected creativity score – defined as the integer closest to the average rating that their creative writing would receive from the four external evaluators – and their expected rank within their gender-balanced group. In the Tournament treatments, participants were asked to report the rank they expected to obtain in the tournament. In the Piece-Rate treatments, participants were asked the same question, namely the rank they expected to obtain under the tournament ranking, regardless of whether they had chosen tournament or piece-rate compensation in Stage 2. Accurate predictions were rewarded with a payment of 2 euros for each correct answer.

Figure 1 illustrates the timeline of the experimental study.

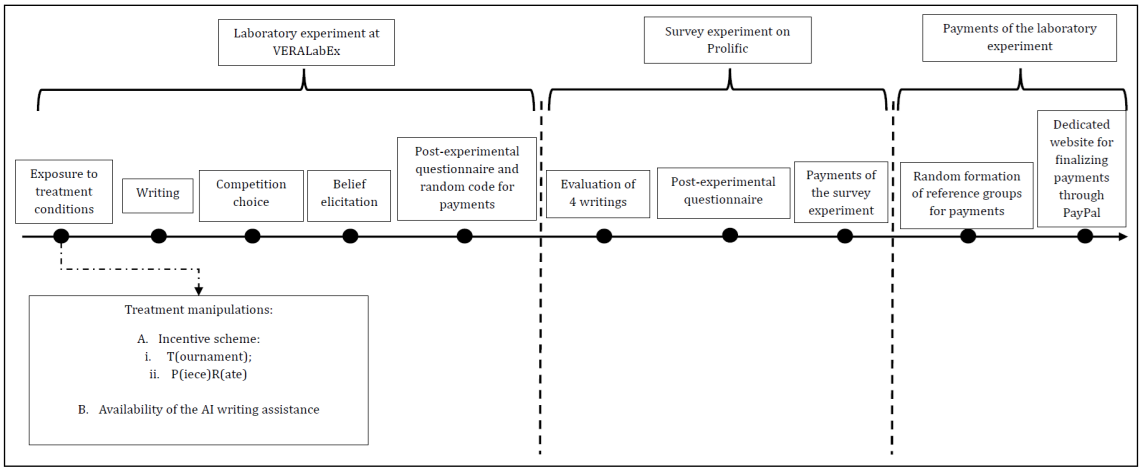

Figure 1 Time flow of the experimental study

We build our experiment on the seminal framework of Niederle and Vesterlund (2007). However, following Erat and Gneezy (2016), we implement a between-subjects design, as repeated exposure to the same creative task would be likely to generate learning, carry-over, and idea-reuse effects, thereby threatening the validity of subsequent measures of creative performance. Participants therefore complete the creative task only once. Competition entry is elicited in Stage 2, in which participants choose between a piece-rate and a tournament compensation scheme based on their performance in Stage 1. This choice represents our measure of willingness to compete and, combined with the experimental variation in AI availability, allows us to examine whether AI affects both creative performance and self-selection into competition.

3.1.1 Experimental Procedures

The laboratory experiment was conducted at VERALabEx (Ca' Foscari University of Venice) and the four treatments considered in this paper involved 249 participants, primarily university students enrolled in economics-related programs. Participation was restricted to native Italian speakers.

Sessions included up to twenty participants and lasted approximately fifty minutes. Sessions in treatments T and $T\&AI$ were perfectly gender-balanced to guarantee the intended gender-balanced composition of tournament groups. Average earnings amounted to 8.17 euros, including a show-up fee of 3 euros.

The experiment was implemented using *oTree* (Chen et al., 2016). Treatment randomization was implemented at the session level, and participants completed comprehension questions before beginning the task.

Following the experiment, participants completed a post-experimental questionnaire collecting socio-demographic characteristics and a set of individual measures commonly associated with creative performance and competition choices. Specifically, we measured their

risk attitude (Dohmen et al., 2011), preference towards competition, and self-esteem (Rosenberg, 1965). We also asked participants to complete the Divergent Association Task (DAT) developed by Olson et al. (2021), which has been shown to correlate strongly with established measures of creativity and is used throughout the analysis as a proxy for creative potential.

Finally, we collected information on participants' reading habits and familiarity with AI technologies.

3.2 Evaluation of Creative Performance - Online experiment

Creative texts written during the laboratory experiment were scored by independent external evaluators recruited through the *Prolific* platform. Participation in the online experiment was restricted to native Italian speakers to ensure linguistic comparability. Each evaluator was randomly assigned four texts to score and received 3 British pounds for completing the task.

Creativity was measured on a scale ranging from 0 to 10. The main performance measure used throughout the paper is the average of the four evaluations assigned to each text.

Importantly, evaluators were blind to both treatment assignment and participant characteristics. They received no information regarding the incentive scheme, access to AI assistance, or the gender of the author. They were informed only of the original writing prompt and of the instructions concerning task duration and expected text length.

3.3 Payment Procedures

After collecting all evaluations, payments were computed according to the rules associated with each incentive scheme. Participants subsequently received an email invitation directing them to a dedicated website where they could retrieve their earnings. Payments were linked to participants through anonymous identification codes generated at the end of the laboratory sessions and were transferred via PayPal by the university administration.

4 Predictions

We develop our predictions on the gender differences in creative performance and willingness to compete in the absence of AI and consequently formulate our hypotheses on how allowing access to Generative AI should contribute to mitigate this gap.[5]

The study of gender differences in creativity has a long-standing tradition in the psychology literature, and even in the absence of a clear consensus, the evidence seems to suggest that females exhibit lower creative performance in comparison to men (Abraham, 2016; Hora et al., 2022). Considering the experimental economics literature, while the gender gap in creativity has received only limited attention (Attanasi et al., 2021), Erat and Gneezy (2016) find results that are consistent with the previous literature, suggesting that women tend to perform worse in creative tasks under competitive pressure.

Allowing individuals to have access to AI in performing creative tasks should attenuate gender differences, precisely because it reduces the gender-specific component of the creative process. It follows that women should feel less competitive pressure when they share the task with an artificial agent. Indeed, the evidence suggests that Generative AI encompasses the gender biases of a prevalently male-dominated society that are inherently present in the real world (Ho et al., 2025). These considerations lead to our first prediction:

Prediction 1: (i) Without AI, women perform worse under competitive incentives (tournament) than under non-competitive schemes (piece-rate), while men's performance remains stable across payment schemes. (ii) The introduction of AI as a support tool is expected to

5 The format of the preregistered hypotheses was edited for expositional reasons, while keeping their content unchanged. Also, Prediction 3 of the preregistration was not investigated in this study.

reduce the gender gap in creative performance by alleviating the negative impact of competition on women's creativity.

In relation to gender differences in willingness to compete, the economics literature discussed above generally points to women's lower willingness to enter competitive environments, although the size and even the existence of this gender gap vary substantially with the nature of the task (Markowsky and Beblo, 2022). A crucial feature of creative tasks, which typically characterize high-skilled labor markets (Charness and Grieco, 2019), is that output evaluation necessarily depends on some form of subjective human judgment, since creativity cannot be assessed mechanically. This distinguishes these tasks from other verbal or non-verbal tasks, for which subjective evaluations do not necessarily apply. Therefore, for creative tasks, gender-biased judgements in determining how one evaluates another subject may play a relevant role (see, for example Bohnet et al., 2015; Milkman et al., 2015; Bordalo et al., 2019). Indeed, there is a well-documented gender bias in associating creative ability with masculine characteristics (Proudfoot et al., 2015; Taylor, 2026). Consistent with this bias, women have been shown to provide lower self-evaluations on creative tasks (Görlitz and Sels, 2025), and this could have a negative impact on their willingness to compete.

As for creative performance, if AI reduces the gender-specific biases associated with competing in a creative task, allowing access to AI may also make tournament entry relatively more attractive to women. We therefore expect access to AI to attenuate the gender gap in willingness to compete. Based on these arguments, we formulate our second prediction:

Prediction 2: (i) Without AI, we expect a gender gap in willingness to compete, with women being less likely than men to select the tournament payment scheme. (ii) The introduction of AI as a support tool is expected to mitigate the gender gap in tournament entry by reducing the tendency of women to shy away from the competitive tournament scheme.

Finally, beyond testing for the existence of gender differences in willingness to compete in creative markets, we explore the role of the possible mechanisms that could be driving this behavior. In this respect, it is important to notice that gender differences may arise either when women rationally anticipate that their actual performance will be lower than that of men in competitive settings, as implied by Prediction 1, or for reasons that are unrelated to actual performance. Within this second category of explanations, the literature has identified different potential mechanisms such as preferences for competition, risk attitude, self-esteem, and underconfidence (Niederle and Vesterlund, 2007; Dohmen et al., 2011; van Veldhuizen, 2022). We therefore attempt to assess the relevance of each of these factors in determining the potential gender gap in decisions to take part in competitions based on creative performance.

5 Results

5.1 Descriptive statistics

Table 1 presents descriptive statistics separately by treatment and gender.

Table 1 Descriptive statistics by treatment

	T	PR	$T\&AI$	$PR\&AI$
Outcome variables				
Creative performance	5.615	5.681	6.281	6.000
	(1.317)	(1.089)	(1.363)	(1.363)
Female	5.865	5.617	6.301	5.934
	(1.281)	(1.102)	(1.266)	(1.312)

	T	PR	$T\&AI$	$PR\&AI$
Male	5.365	5.794	6.261	6.104
	(1.329)	(1.091)	(1.468)	(1.461)
Competition entry	0.404	0.426	0.477	0.387
	(0.495)	(0.500)	(0.502)	(0.491)
Female	0.231	0.333	0.364	0.342
	(0.430)	(0.479)	(0.487)	(0.481)
Male	0.577	0.588	0.591	0.458
	(0.504)	(0.507)	(0.497)	(0.509)
Individual characteristics				
Male	0.500	0.362	0.500	0.387
	(0.505)	(0.486)	(0.503)	(0.491)
Risk attitude	4.173	3.809	4.000	3.661
	(1.530)	(1.952)	(1.516)	(1.536)
Competitive preference	2.981	3.085	3.307	3.048
	(1.093)	(1.195)	(1.010)	(1.179)
Self-esteem	1.885	1.957	1.875	1.806
	(0.646)	(0.588)	(0.622)	(0.649)
Books owned	3.000	2.979	2.909	3.097
	(1.066)	(1.093)	(1.002)	(0.936)
AI exposure	3.577	3.277	3.773	3.484
	(1.334)	(1.410)	(1.239)	(1.446)
DAT score	83.705	82.582	83.518	83.535
	(6.323)	(5.231)	(4.834)	(5.070)
AI-related variables				
AI use			0.818	0.677
			(0.388)	(0.471)
Number of queries			2.625	3.113
			(2.641)	(3.694)
Observations	52	47	88	62

	T	PR	$T\&AI$	$PR\&AI$

Notes. This table reports descriptive statistics (standard deviations in parentheses) by treatment. *Competition entry* is a dummy equal to 1 if the participant chooses the tournament compensation scheme in Stage 2. *Creative performance* is the mean of the four creativity scores assigned to the participant's writing by independent external evaluators in the survey experiment. *Male* is a dummy equal to 1 for male participants. *Risk attitude* is measured on a scale from 1 to 7, with higher values indicating a greater willingness to take risks. *Competitive preference* measures agreement with the statement "I enjoy competition" on a five-point Likert scale, with higher values indicating a stronger preference for competitive environments. *Self-esteem* is constructed from the Rosenberg Self-Esteem Scale (Rosenberg, 1965) and ranges from 1 to 3, with higher values indicating more positive self-evaluations. *DAT score* is the Divergent Association Task (DAT) score (Olson et al., 2021), computed as the average semantic distance (ranging from 0 to 100) between the words provided by the participant, with higher values indicating greater divergent-thinking ability. *Books owned* indicates the number of books owned by the participant at home (either in physical or electronic form) and is measured on a categorical scale: 0 = No books; 1 = Between 1 and 10 books; 2 = Between 11 and 50 books; 3 = Between 51 and 100 books; 4 = More than 100 books. *AI exposure* indicates how often the participant used AI tools in the last 12 months and is measured on a categorical scale: 0 = Never; 1 = Less than once per month; 2 = Once per month; 3 = Once per week; 4 = More than once per week; 5 = More than once per day. *AI use* is a dummy equal to 1 if the participant submitted at least one query to the AI assistant, and *Number of queries* is the total number of submitted prompts.

Creative performance, measured as the average score assigned by independent external evaluators, ranges from 5.62 to 6.28, thus indicating that its value is similar across gender groups within each treatment condition.

Turning to competition choices, the proportion of participants selecting tournament compensation ranges from 38.7% to 47.7% across treatments. Men are descriptively more likely than women to choose tournament compensation in all treatment conditions. While average competition entry varies little across treatments, the gender gap appears less pronounced when AI assistance is available. We investigate this pattern more formally in the regression analysis below.

The lower panels report participants' observable characteristics and AI-related measures used in the empirical analysis. Risk attitudes, competitive preferences, self-esteem, cultural capital, prior exposure to AI technologies, and divergent-thinking ability (DAT score) display broadly similar distributions across treatments.

In the AI treatments, most participants interacted with the AI assistant: AI use equals 81.8% in $T\&AI$ and 67.7% in $PR\&AI$, with average query intensity equal to 2.63 and 3.11 submitted prompts, respectively.

5.2 Creative performance

We first examine creative performance of males and females. Table 2 reports OLS estimates in which the dependent variable is participants' creative performance, measured as the average of the four evaluations assigned by independent external evaluators. All specifications control for participants' DAT score and its squared term. This allows us to compare performance across genders and treatments while holding constant objective ability and allowing for a nonlinear relationship between divergent-thinking ability and creative performance.

In the absence of AI (Columns (1)–(3)), we find no evidence of an overall gender difference in creative performance: in Column (1), the coefficient on *Male* is small and statistically insignificant (-0.090, $p = 0.707$). Looking at the role of incentives, we observe that performance does not differ significantly between Tournament and Piece Rate (column 2), nor does the effect of the incentive scheme differ significantly between women and men (column 3). Turning specifically to women's creative performance, we find no statistically significant difference between the Tournament and Piece-Rate treatments ($p = 0.507$), providing no support for part (i) of Prediction 1.

We next examine the role of AI in Columns (4)–(6). In Column (4), we observe that the coefficient on AI is positive and statistically significant ($p = 0.003$), indicating that participants assigned to AI treatments achieve, on average, creativity scores approximately 0.48 points higher than participants in the non-AI treatments. The overall performance advantage associated with AI availability is not concentrated among one gender, as shown by the small and statistically insignificant coefficient on *Male* (0.011, $p = 0.949$).

Column (5) investigates whether the association between AI and performance differs across treatments by replacing the aggregate AI indicator with treatment-specific indicators and using the Tournament treatment as the reference category. Relative to T, the coefficient on PR is small and statistically insignificant (0.042, $p = 0.862$). By contrast, the coefficient on $T\&AI$ is positive and statistically significant (0.618, $p = 0.006$), whereas the coefficient on $PR\&AI$ is positive but statistically insignificant (0.343, $p = 0.155$). Pairwise comparisons show that performance differs significantly between PR and $T\&AI$ ($p = 0.010$), but not between PR and $PR\&AI$ ($p = 0.207$) or between $T\&AI$ and $PR\&AI$ ($p = 0.225$). Thus, the positive association between AI availability and creative performance appears to be strongest in the Tournament environment, although the evidence does not establish a statistically significant difference between the two AI treatments.

In Column (6), we observe that none of these treatment effects vary by gender as the treatment-by-gender interactions are not statistically significant. Therefore, AI does not generate a statistically different response to the two incentive schemes for either gender.

Table 2 Creative performance

	$AI = 0$			Full sample			$AI = 0$	$AI = 1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Male	−0.090	−0.088	−0.355	0.011	−0.007	−0.385	−0.008	−0.090	0.031
	(0.239)	(0.245)	(0.330)	(0.170)	(0.172)	(0.328)	(0.167)	(0.239)	(0.221)
AI				0.484***					
				(0.161)					
PR		0.017	−0.223		0.042	−0.220			
		(0.250)	(0.336)		(0.239)	(0.331)			
$T\&AI$					0.618***	0.457			
					(0.222)	(0.326)			
$PR\&AI$					0.343	0.069			
					(0.241)	(0.336)			
Male$\times PR$			0.573			0.592			
			(0.479)			(0.475)			
Male$\times T\&AI$						0.331			
						(0.440)			
Male $\times PR\&AI$						0.604			
						(0.483)			
AI use							0.641***		
							(0.168)		
AI no-use							−0.007		
							(0.273)		
Number of queries									0.320***
									(0.076)

Number of queries^2									−0.022***
									(0.005)
DAT score	0.609***	0.608***	0.582***	0.535***	0.525***	0.495***	0.507**	0.609***	0.229
	(0.185)	(0.188)	(0.183)	(0.195)	(0.197)	(0.188)	(0.205)	(0.185)	(0.476)
DAT score^2	−0.004***	−0.004***	−0.003***	−0.003**	−0.003**	−0.003**	−0.003**	−0.004***	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
Constant	−19.825***	−19.783***	−18.643***	−17.048**	−16.662**	−15.314**	−15.956*	−19.825***	−4.550
	(7.126)	(7.203)	(7.007)	(7.869)	(7.905)	(7.568)	(8.292)	(7.126)	(19.625)
F-test	6.24	4.67	4.32	5.23	3.78	3.01	5.42	6.24	5.43
Prob > F	0.001	0.002	0.001	0.001	0.001	0.002	0.000	0.001	0.000
Observations	99	99	99	249	249	249	249	99	150

Notes. This table reports OLS estimates with robust standard errors in parentheses. The dependent variable is the participant's Creative performance, computed as the mean of the four evaluations assigned by independent external evaluators. Columns (1)–(3) use the subsample of participants without AI. Columns (4)–(7) use the full sample, while columns (8) and (9) estimate the determinants of performance separately for the $AI = 0$ and $AI = 1$ subsamples. All specifications control for DAT score and its squared term. In Columns (2), (3), (5), and (6), treatment T and female participants are the reference categories. Column (3) includes the interaction between $Male$ and PR. Column (6) includes the interactions between $Male$ and the treatment indicators. AI use is a dummy equal to one if the participant used the AI tool, whereas AI no-use is a dummy equal to one if no prompt was submitted. Number of queries^2 and DAT score^2 are the squared terms of Number of queries and DAT score, respectively. The other remarks of Table 1 apply. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Our results do not support part (ii) of Prediction 1. Since the competition-induced gender gap anticipated in part (i) of Prediction 1 does not emerge, there is no performance gap for AI to mitigate. Consistently, we find no significant gender differences in performance in either AI treatment.

Taken together, these findings yield our first main result:

> **Result 1.** We find no evidence of gender differences in creative performance, either with or without AI. Incentive schemes do not affect women and men differently in either condition. AI availability, however, is associated with higher creative performance, with the largest treatment-specific difference arising between the PR and $T\&AI$ treatments.

Finally, columns (7)–(9) provide additional evidence on the determinants of performance and AI-related behavior. These latter specifications are complements to the main analysis, since participants' interaction with the AI assistant is examined in greater detail in the companion paper (Atzori et al., 2026). Column (7) distinguishes between participants who actively used the AI assistant, defined as submitting at least one prompt, and participants who had access to AI but did not use it. Relative to participants without access to AI, active AI users achieve significantly higher creativity scores. The coefficient on $AI\ use$ is 0.641 ($p < 0.001$), whereas the coefficient on $AI\ no\text{-}use$ is close to zero and statistically insignificant (-0.007, $p = 0.980$). Thus, the positive association between AI availability and creative performance is concentrated among participants who actively engage with the technology.

Columns (8) and (9) examine the determinants of performance separately in the no-AI and AI subsamples. In Column (8), as in all previous ones, the coefficient on DAT is positive and statistically significant, whereas the coefficient on its squared term is negative and significant, indicating a concave relationship between DAT scores and creative performance, with diminishing returns to divergent-thinking ability. Differently, in Column (9), which focuses on participants with access to AI, neither the DAT score ($p = 0.631$) nor its squared term ($p = 0.659$) is statistically significant, with the coefficients on the number of submitted quer-

ies and its squared term suggesting a positive but concave relationship between interaction intensity and creative performance. These additional results suggest that, once AI is available, performance may depend less directly on baseline divergent-thinking ability and more on participants' interaction with the AI assistant[6].

5.3 Gender differences in competition entry

We now examine participants' decision to enter competition. Table 3 reports the average marginal effects of a set of Probit regressions where the dependent variable is equal to one if the participant chooses the tournament compensation scheme in Stage 2 and is equal to zero otherwise.

6 As robustness checks, we estimate additional specifications of columns (4), (5), (6), (7), (8), and (9) including risk attitude, competitive preference, self-esteem, books owned, and prior AI exposure. These additional controls contribute little explanatory power, while the estimated effects of AI availability, AI use, AI non-use, number of queries, number of queries^2, DAT score, and DAT score^2 remain qualitatively unchanged. Results from these robustness checks are provided in Appendix B.

Table 3 Gender differences in competition entry

	$AI = 0$			
	(1)	(2)	(3)	(4)
Gender difference	0.225***	0.279***		
	(0.056)	(0.082)		
Gender difference in T			0.346***	0.255**
			(0.128)	(0.122)
Gender difference in PR			0.255*	0.125
			(0.147)	(0.141)
Gender difference in $T\&AI$			0.227**	0.077
			(0.104)	(0.097)
Gender difference in $PR\&AI$			0.116	−0.063
			(0.128)	(0.118)
Risk attitude				0.053***
				(0.020)
Competitive preference				0.094***
				(0.028)
Self-esteem				0.123***
				(0.045)
Books owned				−0.008
				(0.027)
AI exposure				−0.046**
				(0.020)
Wald χ^2	13.43	8.58	15.89	58.25
Prob $> \chi^2$	0.000	0.003	0.026	0.000
Observations	249	99	249	249

Notes. This table reports average marginal effects from Probit models with robust standard errors in parentheses. The dependent variable equals one if the participant chooses tournament compensation in Stage 2. Column (2) restricts the sample to treatments without AI availability. All other notes to Table 1 apply. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Column (1) provides an aggregate comparison of willingness to compete between men and women, pooling observations across the four experimental treatments. Male participants are approximately 22.5 percentage points more likely than female participants to choose tournament compensation ($p < 0.001$). This result contrasts with the evidence on creative performance presented in the previous section: while we find no significant overall difference between men and women in creative performance, a substantial gender difference emerges in willingness to compete.

This finding is consistent with the extensive literature documenting gender differences in competition entry (Niederle and Vesterlund, 2007). However, unlike most previous studies, which examine competition in mathematical, analytical, or other structured tasks, our ex-

periment focuses on a creative task. The persistence of a gender gap in this setting suggests that gender differences in willingness to compete extend beyond the mechanical and quantitative environments that have been mainly investigated in the literature.

Column (2) provides clear support for part (i) of Prediction 2. In the absence of AI, men are 27.9 percentage points more likely than women to enter competition ($p = 0.001$). This specification also provides a natural benchmark for the existing literature on gender differences in competition entry, which has largely studied environments without AI assistance.

To examine part (ii) of Prediction 2, which concerns whether AI availability mitigates the gender gap in tournament entry, we extend our previous analysis by distinguishing between the four experimental treatments. While part (i) focuses on the overall gender gap in the absence of AI, part (ii) concerns how this gap changes with AI availability. Our experimental design allows us to investigate this question separately for each incentive scheme, comparing gender differences in competition entry with and without AI while holding the incentive scheme constant.

We proceed in two steps. First, in Column (3), we estimate treatment-specific gender differences without additional controls. Second, in Column (4), we assess whether these differences persist after controlling for individual characteristics that are likely to influence the decision to enter competition.

The estimates in Column (3) reveal three main patterns. First, the largest gender difference in willingness to compete is observed in the T treatment, where men are 34.6 percentage points more likely than women to enter competition ($p = 0.007$). Second, a similar difference is observed also in the PR treatment and in the $T\&AI$ treatment, although the estimated magnitudes become progressively smaller. Third, in the $PR\&AI$ treatment, the estimated gender difference is smaller and no longer statistically significant. In particular, the estimated gender gap decreases from 34.6 percentage points in T to 22.7 percentage points in $T\&AI$. Thus, the point estimates move in the direction predicted by part (ii) of Prediction 2.

In column (4), we examine whether this pattern is robust to controlling for risk attitude, competitive preference, self-esteem, cultural capital, and prior exposure to AI technologies. We observe that a statistically significant gender gap remains only in the T treatment: male participants remain approximately 25.5 percentage points more likely than female participants to enter competition ($p = 0.036$). By contrast, no statistically significant gender differences are observed in PR ($p = 0.376$), $T\&AI$ ($p = 0.425$) or $PR\&AI$ ($p = 0.593$). The comparison between the T and $T\&AI$ treatments is particularly informative since these two environments share the same compensation structure and differ only in the availability of AI assistance. While a statistically significant gender difference is observed in the Tournament treatment without AI, the corresponding estimate is smaller and no longer statistically significant when AI is available. Hence, even after accounting for individual characteristics associated with competition entry, the evidence remains consistent with the pattern predicted by part (ii) of Prediction 2.

> **Result 2.** In the absence of AI, men are significantly more likely than women to enter competition. The gender gap in competition entry is smaller when AI is available, with the largest reduction observed when comparing T with $T\&AI$.

Finally, we briefly discuss the average marginal effects of the individual characteristics associated with competition entry, reported in the lower panel of column (4). We observe that a one-unit increase in risk tolerance increases the probability of selecting tournament compensation by approximately 5.3 percentage points ($p = 0.008$). This result is consistent with standard economic intuition, as tournament schemes involve greater payoff uncertainty and may therefore be relatively more attractive to individuals who are more comfortable with risk.

Competitive preference is also associated with entry decisions. Participants reporting a stronger preference for competition are approximately 9.4 percentage points more likely to choose tournament compensation ($p = 0.001$). Participants with higher self-esteem are approximately 12.3 percentage points more likely to compete ($p = 0.007$), suggesting that self-

perceptions and confidence-related dimensions play an important role in shaping willingness to enter competitive environments.

Prior exposure to AI technologies is negatively associated with competition entry. Participants with greater familiarity with AI are approximately 4.6 percentage points less likely to choose the tournament compensation scheme ($p = 0.023$). While the underlying mechanism is beyond the scope of the present analysis, this finding is consistent with the idea that greater familiarity with AI may reduce the perceived returns to one's own ability in competitive environments. By contrast, cultural capital, as measured by books owned, does not significantly predict competition entry ($p = 0.773$).

Treatment assignment is orthogonal to these observable characteristics (balance regressions are provided in Appendix B), suggesting that the treatment differences documented above are unlikely to reflect systematic compositional differences across experimental conditions.

6 Underconfidence as a potential mechanism

The previous sections highlight two main findings. First, men and women perform similarly in the creative task. Second, a statistically significant gender difference in competition entry emerges in the tournament treatment. Since this difference cannot be explained by performance, we next examine whether men and women differ in their beliefs about their relative performance. Together with risk preferences, confidence about relative performance has indeed been shown to contribute to gender differences in competitiveness (Lozano and Reuben, 2022; van Veldhuizen, 2022).

In Stage 3 of the laboratory experiment, participants were asked to report the rank they expected to obtain in a gender-balanced tournament group, ranging from first to fourth.[7] Their reports were incentivized: participants received €2 for each correct prediction. To construct our measure of *Underconfidence*, we compare each participant's subjective rank belief with a benchmark rank obtained from $1,000$ simulations of randomly generated gender-balanced comparison groups. Specifically, for each participant, we record the rank obtained in each simulated tournament and define the benchmark rank as the modal rank across these simulations. *Underconfidence* is then measured as the difference between the participant's subjective rank belief and this modal rank. Positive values indicate greater underconfidence, as they correspond to participants expecting to obtain a lower rank than the one suggested by the simulations. Negative values indicate that participants expect to obtain a higher rank than the modal simulated rank, while a value of zero indicates that the two ranks coincide.

Table 4 investigates whether underconfidence contributes to explaining gender differences in competition entry. Since self-esteem is potentially correlated with underconfidence, in Column (1) of Table 4 we report the last column of Table 3, and in Column (2) we replace self-esteem with underconfidence. This allows us to compare the explanatory power of underconfidence, which is an experimental variable, with respect to self-esteem, which is a survey variable.

7 Participants in the Piece-Rate treatments were asked to report the rank they expected to obtain under the tournament ranking, regardless of whether they had chosen tournament or piece-rate compensation in Stage 2.

Table 4 Underconfidence and competition entry

	Self-esteem	Underconfidence
	(1)	(2)
Gender difference in T	0.255**	0.204*
	(0.122)	(0.120)
Gender difference in PR	0.125	0.146
	(0.141)	(0.138)
Gender difference in $T\&AI$	0.077	0.096
	(0.097)	(0.094)
Gender difference in $PR\&AI$	−0.063	−0.046
	(0.118)	(0.118)
Confidence measures		
Self-esteem	0.123***	
	(0.045)	
Underconfidence		−0.082***
		(0.020)
Additional controls		
Risk attitude	0.053***	0.052***
	(0.020)	(0.020)
Competitive preference	0.094***	0.090***
	(0.028)	(0.028)
Books owned	−0.008	0.002
	(0.027)	(0.027)
AI exposure	−0.046**	−0.039*
	(0.020)	(0.020)
Wald χ^2	58.25	72.72
Prob > χ^2	0.000	0.000
Observations	249	249

Notes. This table reports average marginal effects from Probit models of competition entry, with robust standard errors in parentheses. The dependent variable equals one if the participant chooses tournament compensation in Stage 2. Column (1) reproduces the baseline specification reported in Table 3, including treatment-by-gender interactions, risk attitude, competitive preference, self-esteem, books owned, and prior AI exposure. Column (2) replaces self-esteem with the underconfidence measure. Underconfidence is defined as the difference between the participant's subjective rank belief and the modal rank obtained from 1,000 simulations of randomly generated gender-balanced comparison groups. Positive values indicate greater underconfidence. The other remarks of Table 3 apply. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

The coefficient of underconfidence is negative and highly statistically significant ($p < 0.001$). A one-unit increase in underconfidence is associated with an 8.2 percentage point lower probability of entering competition. Importantly, introducing underconfidence reduces the estimated gender difference after controlling for competitive preferences, risk attitude, and prior exposure to AI: in the Tournament treatment, the coefficient on male decreases from 25.5 to 20.4 percentage points and becomes only marginally significant ($p = 0.088$), while it is not statistically significant when considering the other treatments.

> **Result 3.** Underconfidence is associated with competition entry. Individuals who underestimate their relative performance are significantly less likely to compete. Once differences in confidence calibration are taken into account, the gender gap in tournament entry is substantially reduced.

6.1 Does AI reduce underconfidence?

These findings indicate that (wrong) beliefs about relative performance are closely associated with competition decisions. Since accounting for underconfidence reduces the estimated gender difference in tournament entry, we next examine whether AI assistance affects participants' beliefs about expected performance and rank.

Table 5 reports ordered probit estimates in which the dependent variable is underconfidence. The explanatory variables include AI availability, self-esteem, and measures of interaction intensity.

Table 5 Determinants of underconfidence

	Average AI effect	Gender-specific AI effects
	(1)	(2)
Male	−0.187	−0.223
	(0.135)	(0.216)
Self-esteem	−0.208**	−0.211**
	(0.106)	(0.106)
Number of queries filled	0.173***	0.173***
	(0.050)	(0.050)
Number of queries filled^2	−0.010***	−0.010***
	(0.003)	(0.003)
AI effects		
AI	−0.381**	
	(0.167)	
AI (women)		−0.407**
		(0.195)
AI (men)		−0.347
		(0.241)
Wald χ^2	20.81	20.84
Prob > χ^2	0.001	0.002
Observations	249	249

Notes. This table reports ordered probit coefficients, with robust standard errors in parentheses. The dependent variable is the underconfidence measure. Column (1) estimates the average effect of AI availability on underconfidence. Column (2) allows the effect of AI to vary by gender. Both specifications additionally control for gender, self-esteem, and the intensity of AI use, measured by the number of submitted queries and its squared term. The reported effects for women and men are obtained from post-estimation linear combinations. The other remarks of Tables 3 and 4 apply. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Column (1) estimates the average relationship between AI availability and underconfidence. Access to AI is associated with significantly lower underconfidence ($p = 0.022$), implying that participants assigned to AI treatments report beliefs that are more closely aligned with their predicted relative performance. Self-esteem is also negatively associated with underconfidence ($p = 0.050$), indicating that individuals with more positive self-perceptions tend to hold beliefs that are closer to their actual ranking. Finally, interaction intensity exhibits a concave relationship with underconfidence, as indicated by the positive coefficient on the number of submitted queries and the negative coefficient on its squared term. This pattern mirrors the non-linear relationship we have already observed for creative performance. Specifically, as shown in Table 2, performance initially improves with the intensity of AI interaction, although at a decreasing rate. The present results suggest that underconfidence exhibits a similar pattern: moderate interaction with AI is associated with lower underconfidence, whereas additional interaction yields progressively smaller improvements. These findings are in line with the evidence reported in Atzori et al. (2026).

Given that underconfidence contributes to explaining the observed gender gap in competition entry, in Column (2) we examine whether the association between AI and underconfidence differs by gender. The estimated AI effect is negative and statistically significant among women ($p = 0.037$), whereas the corresponding estimate for men is also negative but statistically insignificant ($p = 0.149$).

> **Result 4.** Access to AI is associated with lower underconfidence, particularly among women. Together with the relationship between underconfidence and competition entry, this pattern is consistent with a potential mechanism through which AI may attenuate the gender gap in willingness to compete.

Taken together, these findings suggest that the reduction in the gender gap in competition entry is unlikely to be explained by gender differences in creative performance. Instead, our evidence points to a confidence-based mechanism. Although our analysis does not establish a formal mediation mechanism, AI is associated with lower underconfidence, particularly among women, and underconfidence is a strong predictor of competition entry.
LUCA

7 AI, creative performance, and selection into competition

The results presented above show that access to AI reduces the gender gap in willingness to compete, with the largest reduction observed when comparing the two treatments in which participants perform the creative task under tournament incentives, T and $T\&AI$). Focusing on these treatments, we next provide descriptive evidence on whether this change in participation is accompanied by a change in the quality composition of individuals who enter competition. Holding the incentive scheme constant allows us to isolate the role of AI availability and examine whether the reduction in the gender gap is accompanied by a change in the performance profile of tournament entrants.

Figure 2 reports kernel-density estimates of creative-performance scores separately by gender, treatment, and competition choice. The left panel refers to participants who do not select tournament compensation, whereas the right panel refers to those who enter competition. The figure illustrates the positive association between AI availability and creative performance already documented in Section 5. In particular, the distributions shift to the right when moving from the Tournament treatment to $T\&AI$. This shift is visible for both women and men, and appears particularly pronounced among participants who do not enter competition.

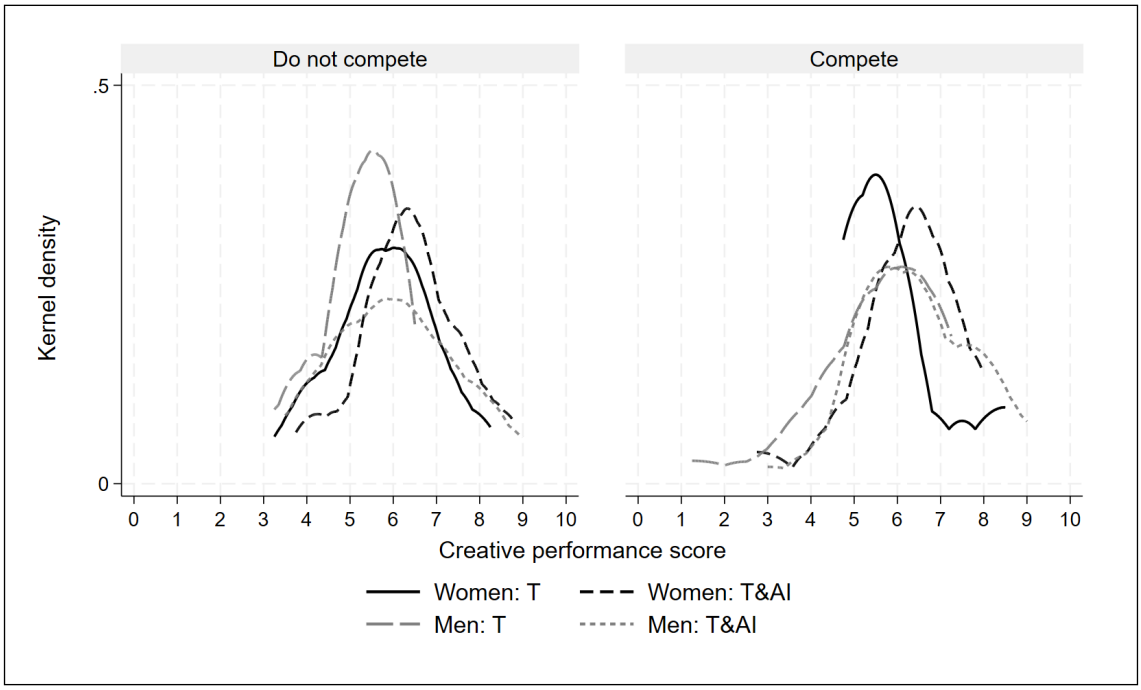

Figure 2 Creative performance distributions by gender and competition choice

Among women who select tournament compensation, the distribution also appears to shift moderately to the right when AI is available. The corresponding average score increases from 6.000 in T treatment to 6.188 in $T\&AI$, while the median increases from 5.500 to 6.375. However, the two distributions remain substantially overlapping. Moreover, the difference is not statistically significant, and the small number of female tournament entrants in the no-AI treatment makes this comparison imprecise. We therefore interpret the graphical evidence as descriptive rather than as evidence of a statistically significant change in the quality composition of female competitors.

Table 6 reports the corresponding descriptive statistics for creative performance and most likely ranking. The comparison between participants who enter and those who do not suggests that the relationship between performance and competition choice differs by gender. In the absence of AI, women who enter competition have a slightly higher average creative-performance score than women who do not enter, although the difference is small (6.000 versus 5.825). When AI is available, this difference disappears and becomes slightly negative: women who enter competition have an average score of 6.188, compared with 6.366 among women who do not enter. Thus, access to AI does not appear to strengthen positive performance-based selection among female competitors.

Table 6 Creative Performance and Ranking by Treatment, Gender, and Competition Choice

	Do not compete		Compete	
	T	T&AI	T	T&AI
Panel A: Creative performance score				
Women	5.825	6.366	6.000	6.188
	(1.293)	(1.252)	(1.351)	(1.324)
	[5.875]	[6.250]	[5.500]	[6.375]
Men	5.182	6.014	5.500	6.433
	(0.936)	(1.561)	(1.575)	(1.405)
	[5.500]	[6.000]	[5.750]	[6.000]
Panel B: Benchmark rank				
Women	2.167	2.464	2.167	2.375
	(0.983)	(0.999)	(0.983)	(1.088)
	[2.000]	[2.000]	[2.000]	[2.000]
Men	2.818	2.611	2.533	2.577
	(0.982)	(1.145)	(1.187)	(1.206)
	[3.000]	[3.000]	[3.000]	[3.000]
Observations	20	28	6	16
	11	18	15	26

Notes: Means are reported first, standard deviations in parentheses, and medians in square brackets. The creative-performance score is the average score assigned by the external evaluators. The benchmark rank is the modal rank obtained from 1,000 Monte Carlo simulations of randomly generated gender-balanced comparison groups; see Table 4 for details. Lower values indicate better relative performance. The first row of observations refers to women and the second row to men.

The pattern is different for men. In both treatments, men who enter competition have higher average creative-performance scores than men who do not. The difference is 0.318 points in the absence of AI and 0.419 points when AI is available. This suggests that male competition entry remains positively selected on creative performance, whereas the relationship between performance and competition entry is considerably weaker among women.

The ranking measures provide a similar picture. Among women who enter competition, the average most likely ranking changes from 2.167 in T to 2.375 in $T\&AI$, while the median remains equal to 2. Among men, the corresponding averages are 2.533 and 2.577, with a median of 3 in both treatments. Since lower values indicate better relative performance, these figures do not suggest that AI improves the relative ranking of participants who enter competition. Instead, they are consistent with the interpretation that AI increases participation without substantially changing the relative performance profile of tournament entrants.

Taken together, the evidence does not indicate that the reduction in the gender gap in competition entry is driven by a systematic change in the quality composition of competitors. AI improves creative performance overall, but it does not appear to make competition entry more strongly selected on performance, particularly among women. This finding complements the evidence on underconfidence presented above. The reduction in the gender gap in tournament entry appears to be primarily behavioral: AI is associated with lower underconfidence, especially among women, and underconfidence predicts competition entry. At the same time, the additional descriptive evidence suggests that AI broadens women's participation in competitive environments without selectively attracting only those with

higher objective creative performance. Since the present analysis is descriptive and based on relatively small subgroups, these conclusions should be interpreted as suggestive rather than definitive.

8 Conclusion

"AI is the next Industrial Revolution"[8]. This claim, frequently echoed by technology leaders and the popular press, has become a pervasive narrative in the current public debate, affecting expectations and educational choices among university students[9]. Reflecting this growing interest, the scientific literature has increasingly focused on providing empirical evidence on the economic consequences of generative AI by identifying the specific mechanisms through which it affects workers, firms, and labor markets. At the same time, despite decades of research, gender inequalities in labor-market outcomes remain one of the most debated issues in economics.

Bringing together these two research agendas – the economic consequences of generative AI and persistent gender inequalities – raises a fundamental question: how will AI affect existing gender inequalities? This question is particularly relevant in creative work, as creativity is recognized as a key determinant of success in modern labor markets where generative AI is expected to have a great impact.

Our study contributes to this debate by examining whether access to ChatGPT affects creative performance and competition entry, and whether these effects differ between women and men, thereby providing new evidence on one of the potential channels through which generative AI may shape future labor-market inequalities. Specifically, we conducted a preregistered laboratory experiment in which participants first completed a creative writing task under either piece-rate or tournament incentives, with or without access to ChatGPT. Immediately after completing the task, they chose the compensation scheme under which they wished their performance to be evaluated, allowing us to measure their willingness to compete. To preserve incentive compatibility, one of the two stages was randomly selected for payment. This design allows us to isolate the effect of generative AI on creative performance and competition entry, and to examine whether these effects differ between women and men.

Our findings yield three main conclusions. First, we find no evidence of gender differences in creative performance, regardless of whether participants have access to AI. Second, access to AI increases creative performance, but these gains are driven by active engagement with the AI assistant rather than by its mere availability. Third, women remain less likely than men to enter competitive environments in the absence of AI, but this gender gap disappears when participants have access to AI. The benefits of AI are not automatic, however. Performance gains emerge only when individuals actively engage with the technology, emphasizing that effective human-AI collaboration, rather than simple access, is the key driver of improved outcomes. Finally, our evidence suggests that underconfidence is an important mechanism underlying this result, being associated with competition entry. AI is associated with lower underconfidence, particularly among women, and accounting for underconfidence substantially attenuates the remaining gender gap in tournament participation.

While our experiment focuses on a specific creative task performed in a laboratory setting, future work might examine whether similar effects arise in real-world creative occupations, such as managerial, entrepreneurial, or design-related work, where generative AI is increasingly integrated into everyday decision-making. It will also be important to investigate whether the changes in confidence and competition entry documented here persist over time and translate into educational choices, job applications, promotion decisions, and other

8 Hinton, G. (2025, December 11). AI Is the Next Industrial Revolution. TIME. Retrieved on 29.07.2026.
9 Marken, S. (2026, April 1). College Students Weigh AI's Impact on Majors and Careers. Gallup survey. Retrieved on 29.07.2026.

career outcomes. Such evidence will be essential for evaluating whether the widespread adoption of generative AI contributes to narrowing – or widening – existing gender inequalities in the labor market.

References

Abraham, A. (2016). Gender and creativity: An overview of psychological and neuroscientific literature. *Brain Imaging and Behavior 10*(2), 609–18.

Aldasoro, I., O. Armantier, S. Doerr, L. Gambacorta, and T. Oliviero (2024). The gen AI gender gap. *Economics Letters 241*, 111814.

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology 43*(5), 997–1013.

Amabile, T. M. (1996). *Creativity in Context*. Westview Press.

Amabile, T. M., P. Goldfarb, and S. C. Brackfield (1990). Social influences on creativity: Evaluation, coaction, and surveillance. *Creativity Research Journal 3*(1), 6–21.

Anderson, N., K. Potočnik, and J. Zhou (2014). Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework. *Journal of Management 40*(5), 1297–1333.

Ariely, D., U. Gneezy, G. Loewenstein, and N. Mazar (2009, 04). Large stakes and big mistakes. *The Review of Economic Studies 76*(2), 451–469.

Attanasi, G., M. Chessa, S. Gil-Gallen, and P. Llerena (2021). A survey on experimental elicitation of creativity in economics. *Revue d'Économie Industrielle 174*, 273–324.

Atzori, F., L. Corazzini, V. Maggian, F. Pavesi, and M. Scotti (2026). Human-AI interaction in creative tasks: An experimental investigation. Working Paper 16/2026, Department of Economics, Ca' Foscari University of Venice.

Blau, F. D. and L. M. Kahn (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature 55*(3), 789–865.

Bohnet, I., A. van Geen, and M. Bazerman (2015). When performance trumps gender bias: Joint vs. separate evaluation. *Management Science 62*(5), 1225–1234.

Bohren, N., R. Hakimov, and R. Lalive (2024). Creative and strategic capabilities of generative AI: Evidence from large-scale experiments. Technical report, IZA Institute of Labor Economics.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2019). Beliefs about gender. *American Economic Review 109*(3), 739–73.

Brynjolfsson, E., D. Li, and L. Raymond (2025). Generative AI at work. *The Quarterly Journal of Economics 140*(2), 889–942.

Buser, T., M. Niederle, and H. Oosterbeek (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics 129*(3), 1409–1447.

Buser, T., M. Niederle, and H. Oosterbeek (2026, May). Can competitiveness predict education and labor market outcomes? evidence from incentivized choice and survey measures. *The Review of Economics and Statistics 108*(3), 755–773.

Carvajal, D., C. Franco, and S. Isaksson (2025). Will artificial intelligence get in the way of achieving gender equality? Discussion Paper 03, Department of Economics, Norwegian School of Economics. https://ssrn.com/abstract=4759218.

Charness, G. and D. Grieco (2019). Creativity and incentives. *Journal of the European Economic Association 17*(2), 454–496.

Charness, G. and D. Grieco (2026). Creativity and AI. *The Economic Journal*, forthcoming.

Chen, D. L., M. Schonger, and C. Wickens (2016). oTree — An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance 9*, 88–97.

Cristofaro, M., P. L. Giardino, and M.-S. Augier (2026). Generative AI and cognitive biases in strategic decision-making. *Strategic Organization*.

Deci, E. L., R. Koestner, and R. M. Ryan (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin 125*(6), 627–668.

Delfgaauw, J., R. Dur, J. Sol, and W. Verbeke (2013). Tournament incentives in the field: Gender differences in the workplace. *Journal of Labor Economics 31*(2), 305–326.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*(3), 522–550.

Doshi, A. R. and O. P. Hauser (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances 10*(28), eadn5290.

Erat, S. and U. Gneezy (2016). Incentives for creativity. *Experimental Economics 19*(2), 269–280.

Flory, J. A., A. Leibbrandt, and J. A. List (2015). Do competitive workplaces deter female workers? a large-scale natural field experiment on job-entry decisions. *The Review of Economic Studies 82*(1), 122–155.

Görlitz, K. and T. Sels (2025). Gender differences in performance evaluations. *Journal of Behavioral and Experimental Economics 119*, 102459.

Ho, J. Q. H., A. Hartanto, A. Koh, and N. M. Majeed (2025). Gender biases within artificial intelligence and ChatGPT: Evidence, sources of biases and solutions. *Computers in Human Behavior: Artificial Humans 4*, 100145.

Holzner, N., S. Maier, and S. Feuerriegel (2025). Generative AI and creativity: a systematic literature review and meta-analysis. *arXiv:2505.17241*.

Hora, S., K. L. Badura, G. J. Lemoine, and E. Grijalva (2022). A meta-analytic examination of the gender difference in creative performance. *Journal of Applied Psychology 107*(11), 1926–50.

Hoyer, B., T. van Huizen, L. Keijzer, S. Rezaei, S. Rosenkranz, and B. Westbrock (2020). Gender, competitiveness, and task difficulty: Evidence from the field. *Labour Economics 64*, 101815.

Kamas, L. and A. Preston (2012). The importance of being confident; gender, career choice, and willingness to compete. *Journal of Economic Behavior and Organization 83*(1), 82–97.

Katz, J. H., T. C. Mann, X. Shen, J. A. Goncalo, and M. J. Ferguson (2022). Implicit impressions of creative people: Creativity evaluation in a stigmatized domain. *Organizational Behavior and Human Decision Processes 169*, 104116.
Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review 90*(5), 1346–1361.
Lozano, L. and E. Reuben (2022). Measuring preferences for competition: Technical report.
Markowsky, E. and M. Beblo (2022). When do we observe a gender gap in competition entry? a meta-analysis of the experimental literature. *Journal of Economic Behavior and Organization 198*, 139–163.
Meier, K., A. Niessen-Ruenzi, and S. Ruenzi (2024). The impact of role models on women's self-selection into competitive environments. *The Quarterly Journal of Finance 14*(02), 2440001.
Milkman, K. L., M. Akinola, and D. Chugh (2015). What happens before? a field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *The Journal of Applied Psychology 100*(6), 1678–1712.
Mourelatos, E., P. Zervas, D. Lagios, and G. Tzimas (2024). Can AI bridge the gender gap in competitiveness? *GLO Discussion Paper Series* (1404).
Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics 122*(3), 1067–1101.
Niederle, M. and L. Vesterlund (2011). Gender and competition. *Annual Review of Economics 3*(Volume 3, 2011), 601–630.
Noy, S. and W. Zhang (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science 381*(6654), 187–192.
Olson, J. A., J. Nahas, D. Chmoulevitch, S. J. Cropper, and M. E. Webb (2021). Naming unrelated words predicts creativity. *Proceedings of the National Academy of Sciences 118*(25), e2022340118.
Otis, N. G., S. Delecourt, K. Cranney, and R. Koning (2024). Global evidence on gender gaps and generative AI. Technical report, Harvard Business School, Boston, MA.
Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature 37*(1), 7–63.
Proudfoot, D., A. C. Kay, and C. Z. Koval (2015). A gender bias in the attribution of creativity: Archival and experimental evidence for the perceived association between masculinity and creative thinking. *Psychological science 26*(11), 1751–61.
Reich, T. and J. D. Teeny (2026). Does artificial intelligence cause artificial confidence? generative artificial intelligence as an emerging social referent. *Journal of Personality and Social Psychology 130*(3), 452–464.
Reuben, E., M. Wiswall, and B. Zafar (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal 127*(604), 2153–2186.
Rosenberg, M. (1965). The measurement of self-esteem, society and the adolescent self-image. *Princeton*, 16–36.
Shurchkov, O. (2012). Under pressure: Gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association 10*(5), 1189–1213.
Taylor, C. L. (2026). Gender bias in creativity: A process model for understanding the gender gap in creative achievement. *Perspectives on Psychological Science 21*(1), 3–24. PMID: 40808246.
UNESCO, OECD, and IDB (2022). The effects of AI on the working lives of women. Technical report, UNESCO, Paris.
van Veldhuizen, R. (2022). Gender differences in tournament choices: Risk preferences, overconfidence, or competitiveness? *Journal of the European Economic Association 20*(4), 1595–1618.

Online Appendix. Competition in Creative Tasks, Gender Gaps and the Mitigating Role of AI: Experimental Evidence

This online appendix contains supplementary material referenced in the main paper. Specifically, it includes: (i) the experimental instructions for both the laboratory and survey experiments (Appendix A); and (ii) the robustness checks discussed in the main paper (Appendix B).

A Experimental Instructions

The instructions were originally written in Italian. The instructions for the laboratory experiment were distributed at the beginning of each phase and read aloud. Instructions that apply only to the AI treatments are shown in blue, those that apply only to the Tournament treatments in red, and those that apply only to the Piece Rate treatments in green.
The instructions for the survey experiment were provided on the first page of the Qualtrics platform.

A.1 Instructions for the Laboratory Experiment

General Instructions

Thank you for participating in this experiment! During the experiment, you are not allowed to speak or communicate in any way with other participants. For any questions, click on the lifebuoy-shaped button at the top right of the screen; we will answer your questions privately. The following rules are the same for all participants.

A total of 20 people are participating in this experiment. The choices you make during the experiment are anonymous: other participants will not know your choices either during or after the experiment. Similarly, you will not be informed of the identity of the other participants, nor will you know the choices they made during the experiment.

The experiment consists of 3 consecutive phases: Phase 1, Phase 2, and Phase 3. You will shortly receive the instructions for Phase 1. The instructions for the remaining two phases will be provided at the beginning of the corresponding phase.

Although you will participate in three phases, your payment at the end of the experimental session will be determined by adding your earnings from Phase 3 to those obtained in only one of the first two phases. Specifically, at the end of the experiment, we will randomly select which of the first two phases will be used to determine your final earnings by flipping a coin.

In addition to the earnings you obtain during the experiment, you will receive € 3 for arriving on time. Your final earnings will be credited to your PayPal account.

Before your earnings are credited to your PayPal account, you will receive an email inviting you to access a webpage where you will find information about your results in the experiment and your earnings.

Instructions for Phase 1

What do you have to do in Phase 1?

In Phase 1, you are asked to write a creative composition based on the following prompt:

> "If you had the talent to invent things just by thinking of them, what would you create?"

We suggest that you do not exceed 150 words and do not spend more than 15 minutes completing your composition. These suggestions are purely indicative. Exceeding 150 words or spending more than 15 minutes will have no effect on the final outcome of Phase 1. To assist you, the computer will display the remaining time (in minutes) as well as a word count.

[Only in the AI treatments] On the screen where you are asked to write your composition, you will find a box that allows you to access a generative artificial intelligence application (ChatGPT). Using this box, you may send any requests you wish, without restrictions on their form or number. The possibility to use ChatGPT will end when you complete your composition and proceed to the next phase.

If you decide to send requests to ChatGPT, we suggest including indications about the theme or context of your creative composition.

The questions you send to ChatGPT and the corresponding responses will not be saved nor used to train or improve OpenAI models. Data sent through this interface will not be accessible to OpenAI developers, in accordance with the platform's Data Usage Policy. For further information, please contact the researchers involved in this experiment.

When writing your composition *[Only in the AI treatments]* and sending requests to ChatGPT, please avoid including personal information or any elements that could reveal your identity.

Your creative composition will receive 4 independent and anonymous evaluations from external evaluators. At the end of this session, your composition will be sent to 4 evaluators randomly selected from approximately 2000 Italian-speaking individuals (male or female) registered on the Prolific platform.

Each evaluator will be asked to carefully read your composition and answer the following question:

> "In general, from 0 to 10, where 0 is the minimum and 10 is the maximum, how do you rate the level of creativity of this composition?"

You will not receive any information about the identity of the evaluators. Similarly, evaluators will not receive any information about your identity or about the rules determining earnings. The only information provided to evaluators will be the prompt of the composition.

How are earnings determined in Phase 1?

[only in the Tournament Treatments] At the beginning of Phase 1, the computer will randomly assign you, with equal probability, to a group of 4 participants (2 males and 2 females). You will not know the identity or the composition of the other group members.

Your earnings in Phase 1 will be determined by a competitive scheme comparing the average rating of your composition with those of the other 3 group members. Specifically, for each composition, the arithmetic mean of the 4 evaluations will be calculated. Based on these averages, a ranking will be established.

If your composition receives the highest average rating in the group, your earnings (in euros) will be:

$$2.4 \times \text{Average rating}$$

Otherwise, you will receive no earnings. In case of a tie, the winner will be randomly determined among those with the same average rating.

[only in the Piece Rate Treatments] Your earnings in Phase 1 will be determined according to a scheme based exclusively on the evaluations of your creative composition. Your earnings (in euros) will be:

$$0.6 \times \text{Average rating}$$

If Phase 1 is selected for payment, the webpage you access before receiving payment will show:

Your average rating

[only in the Tournament Treatments] Whether your composition ranked first in your group

Instructions for Phase 2

What do you have to do and how are earnings determined in Phase 2?

In Phase 2, you must choose between two payment schemes for your composition written in Phase 1:

- A non-competitive scheme based exclusively on the average rating of your creative composition *[only in the Piece Rate Treatments]*, as described above. Specifically, the arithmetic mean of the 4 ratings assigned by the evaluators will be calculated. Your earnings (in euros) will then be determined as follows: Your earnings will be:

$$0.6 \times \text{Average rating}$$

- A competitive scheme that compares the average rating of your creative composition with those of 3 other people *[only in the Tournament Treatments]*, as described above. Specifically, each group consists of 4 participants: 2 men and 2 women. *[only in the Piece Rate Treatments]* who previously took part in this experiment. Specifically, the arithmetic mean of the 4 ratings assigned by the evaluators will be calculated for each composition. The compositions will then be ranked within the group according to their average ratings. If your composition receives the highest average rating in the group, your earnings (in euros) will be:

$$2.4 \times \text{Average rating}$$

Otherwise, you will receive no earnings. In case of a tie, the winner will be randomly selected with equal probability from among the group members with the same average rating.

[only in the Piece Rate Treatments] Who are the 3 other people you compete against under the competitive scheme? If you choose the competitive scheme to determine your earnings in this phase, the computer will randomly assign you to a group of 4 people, consisting of 2 men and 2 women. You will not know the identities of the other 3 group members or see their creative compositions. The only information available to you about them is that: (i) they took part in an experimental session previously conducted in this laboratory; (ii) they wrote their compositions knowing that their earnings would be determined according to the same competitive scheme; (iii) they were asked to write a composition based on the same prompt described above; and (iv) each of their compositions will be rated by 4 external evaluators according to the same rules described above.

Your earnings in Phase 2 will depend on the scheme you choose.

If Phase 2 is selected for payment, the webpage will show:

- Your average rating
- If you chose the competitive scheme, whether your composition ranked first

Instructions for Phase 3

What do you have to do in Phase 3?

In Phase 3, you must make 2 guesses:

- You have to guess the rank of your creative composition within the group formed under the competitive scheme described in Phase 2. In other words, you must predict whether, based on the average rating assigned by the evaluators, your composition will rank first, second, third, or fourth in your group.
- You have to guess the integer (from 0 to 10) closest to your average rating assigned to your creative composition by the 4 evaluators.

How are earnings determined in Phase 3?

For each correct guess, you will earn € 2.

The webpage you access before receiving payment will provide information on the correctness of your guesses.

A.2 Instructions for the Survey Experiment

Instructions

Thank you for participating in this experiment! During the experiment, you will be asked to make simple choices and complete a short questionnaire. All the choices you make during the experiment and the information you provide in the questionnaire will remain completely anonymous and will be analyzed by the researchers involved in the study in aggregate form for scientific research purposes.

The expected duration of the experiment, including the time required to complete the questionnaire, is approximately 15 minutes. For participating in this experiment, you will receive a payment of £3 (British pounds). Payment will be made through the Prolific platform. For this reason, we ask you to keep your Prolific ID code at hand. Payment will only be granted if you complete the experiment and the final questionnaire in full.

For further information about the experimental study, please contact the researchers involved in the research project:

Luca Corazzini, University of Milano-Bicocca, luca.corazzini@unimib.it;
Valeria Maggian, Ca' Foscari University of Venice, valeria.maggian@unive.it;

Filippo Pavesi, LIUC - Carlo Cattaneo University, fpavesi@liuc.it;
Massimo Scotti, LIUC - Carlo Cattaneo University, mscotti@liuc.it;
Federico Atzori, University of Milano-Bicocca, federico.atzori@unimib.it

Experimental Instructions

What do you have to do?

You are asked to evaluate the level of creativity of 4 creative compositions produced by participants in a laboratory experiment conducted at the VeraLabEx experimental economics laboratory of Ca' Foscari University of Venice.

In particular, participants in the laboratory experiment were asked to write a creative composition based on the following prompt:

> "If you had the talent to invent things just by thinking of them, what would you create?"

For each creative composition presented to you, you are asked to provide your evaluation based on the following question:

> "In general, from 0 to 10, where 0 is the minimum and 10 is the maximum, how do you rate the level of creativity of this composition?"

There are no right or wrong answers to the evaluation task. We ask you to carefully read each composition and provide the evaluation that best reflects your opinion of its level of creativity.

In addition to providing your evaluation, for each of the 4 compositions you are asked to indicate 3 keywords that, in your opinion, best describe the content of the composition.

Additional Information on the Creative Compositions and the Study

Participants in the laboratory experiment were advised not to exceed 150 words and not to spend more than 15 minutes completing their composition. However, they were informed that exceeding 150 words or spending more than 15 minutes would have no effect on the outcome of their participation in the experiment.

As indicated in the previous instructions, participants in the laboratory experiment will not receive any information about your identity. Similarly, you will not receive any information about the identity of the 4 participants whose compositions you will evaluate.

B Robustness checks

Table 1 Creative performance: robustness checks

	Full sample				$AI = 0$	$AI = 1$
	(1)	(2)	(3)	(4)	(5)	(6)
Male	−0.022	−0.032	−0.398	−0.044	−0.147	0.057
	(0.178)	(0.180)	(0.335)	(0.173)	(0.256)	(0.243)
AI	0.504***					
	(0.160)					
PR		0.069	−0.185			
		(0.242)	(0.334)			
$T\&AI$		0.657***	0.531*			

		(0.219)	(0.321)			
$PR\&AI$		0.371	0.073			
		(0.240)	(0.341)			
Male$\times PR$			0.582			
			(0.472)			
Male$\times T\&AI$			0.263			
			(0.431)			
Male$\times PR\&AI$			0.665			
			(0.490)			
AI use				0.678***		
				(0.166)		
AI no-use				−0.018		
				(0.274)		
Number of queries						0.317***
						(0.077)
Number of queries^2						−0.022***
						(0.005)
DAT score	0.542**	0.532**	0.503**	0.515**	0.679***	0.180
	(0.212)	(0.214)	(0.207)	(0.227)	(0.170)	(0.485)
DAT score^2	−0.003**	−0.003**	−0.003**	−0.003**	−0.004***	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
Risk attitude	0.038	0.038	0.038	0.036	0.148*	−0.080
	(0.065)	(0.065)	(0.064)	(0.063)	(0.083)	(0.094)
Competitive preference	−0.095	−0.103	−0.104	−0.094	−0.154	−0.012
	(0.090)	(0.089)	(0.090)	(0.088)	(0.110)	(0.138)
Self-esteem	0.123	0.117	0.098	0.150	0.088	0.119
	(0.135)	(0.134)	(0.133)	(0.128)	(0.180)	(0.173)
Books owned	0.073	0.082	0.098	0.103	0.045	0.126
	(0.076)	(0.078)	(0.082)	(0.075)	(0.105)	(0.108)
AI exposure	0.051	0.046	0.049	0.035	−0.052	0.100
	(0.066)	(0.067)	(0.068)	(0.066)	(0.100)	(0.089)
Constant	−17.804**	−17.388**	−16.113*	−16.761*	−22.745***	−3.250
	(8.612)	(8.661)	(8.364)	(9.239)	(6.556)	(19.973)
F-test	2.68	2.47	2.17	3.51	3.98	3.28
Prob > F	0.006	0.006	0.009	0.000	0.001	0.001

Observations	249	249	249	249	99	150

Notes. This table reports OLS estimates with robust standard errors in parentheses. The dependent variable is the participant's Creative performance, computed as the mean of the four evaluations assigned by independent external evaluators. All specifications include the controls used in the corresponding columns of Table 2 in the main paper and additionally control for risk attitude, competitive preference, self-esteem, books owned, and AI exposure. The other remarks of Table 2 in the paper apply. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 2 Balance of individual characteristics across treatments

	Risk attitude	Competitive preference	Self-esteem	Books owned	AI exposure
	(1)	(2)	(3)	(4)	(5)
PR	−0.229	0.085	0.132	−0.021	−0.229
	(0.229)	(0.218)	(0.223)	(0.217)	(0.206)
$T\&AI$	−0.115	0.294*	−0.018	−0.091	0.184
	(0.168)	(0.175)	(0.202)	(0.182)	(0.186)
$PR\&AI$	−0.330*	0.067	−0.143	0.097	−0.076
	(0.181)	(0.200)	(0.221)	(0.190)	(0.207)
Constant				3.000***	
				(0.148)	
Wald χ^2/F-test	3.69	3.55	1.63	0.46	5.47
Prob > χ^2/F	0.297	0.315	0.653	0.709	0.140
Observations	249	249	249	249	249

Notes. This table reports robust ordered probit estimates in columns (1)–(3) and (5), and OLS estimates with robust standard errors in column (4). The dependent variables are, respectively, risk attitude, competitive preference, self-esteem, books owned, and AI exposure, as measured in the post-experimental questionnaire. The purpose of these estimations is to test whether individual characteristics are balanced across treatment groups. Treatment T is the reference category in all specifications. The table reports the Wald χ^2 statistic for the ordered probit models and the F statistic for the OLS model. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.