NETWORK LISER

Discussion Paper Series

IZA DP No. 18784

July 2026

When AI Does the Work: Does Attribution Shape Meaning and Effort?

Milena Nikolova
University of Groningen,
The Brookings Institution,
GLO, IZA@LISER and Netspar

Viliana Milanova
University of Groningen

Feicheng Wang
University of Groningen,
GLO and IZA@LISER

IZA@LISER NETWORK
ISSN: 2365-9793

When AI Does the Work: Does Attribution Shape Meaning and Effort?*

Abstract

This paper provides the first causal evidence that merely attributing identical creative work to AI rather than to a human affects how much meaning people derive from a task and how much effort they are willing to contribute. We conducted a pre-registered survey experiment in nationally representative samples from the United States (N = 1,511) and the Netherlands (N = 2,117). Participants evaluated identical public health campaign slogans that were randomly attributed either to an AI system or to a human professional, allowing us to isolate the causal effect of AI attribution while holding the creative output constant. AI attribution reduced perceived task meaning modestly and made participants 13% less likely to contribute a slogan of their own, indicating lower voluntary effort. These findings suggest that AI can influence work not only by changing productivity but also by altering the perceived value of human contribution itself.

JEL classification

C91, J01, I30, O33

Keywords

Artificial intelligence (AI), meaning, effort, survey experiment

Corresponding author

Milena Nikolova
m.v.nikolova@rug.nl

* *Acknowledgements*: We thank Marco Angrisani, Viola Angelini, Jana Holthöwer, Evgenia I. Lysova, Jocelyne Ludoph, and Boris Nikolaev for their valuable input during the development of the questionnaire. We also thank Andrew Oswald, Juliette de Wit, participants of the ISQOLS 2025 Annual Conference Session 18E on Technology and Well-being, the 2026 PhD Conference in Groningen, and the *Research Policy* 10th Online Conference for early career researchers for helpful feedback. Jill Darling, Bart Orriens, and Marco Angrisani from the UAS and Stein Jongerius from the LISS panel provided excellent data collection implementation and support. This project has received funding from the Jantina Tammes School of Digital Society, Technology and AI at the University of Groningen and the Institute for Humane Studies at George Mason University (Grant number IHS018859).

1. Introduction

In his 2024 book, Ethan Mollick described the early stages of the Generative Artificial Intelligence (GenAI) era, characterized by chatbots and large language models (LLMs) (Mollick, 2024). For the first time in human history, a technology could generate new content, such as text, images, and video, by repurposing and recombining existing data. Mollick portrayed working with AI as a collaborative process in which humans interact repeatedly with the technology while serving as its stewards. In this "human-in-the-loop" relationship, AI complements human capabilities across a range of writing, analytical, and creative tasks (Boussiouxet al., 2024; Brynjolfsson et al., 2025; Dell'Acqua et al., 2026), many of which overlap with the skill sets of highly educated workers (Eloundou et al., 2024; Felten et al., 2023). As AI increasingly performs tasks once considered uniquely human, it raises important questions about the nature and quality of work. These questions matter because a large body of research shows that meaningful work and job quality influence worker well-being, effort, productivity, and labor supply (Cassar & Meier, 2018; De Neve et al., 2013; De Neve & Ward, 2017; Hendriks & Cnossen, 2026)

Although GenAI now outperforms humans on many tasks, including writing, coding, and even competition-level mathematics (Maslej et al., 2025; Sajadieh et al., 2026), its capabilities remain uneven. AI performs exceptionally well on certain tasks within what Mollick (2024) and Dell'Acqua et al. (2026) describe as a "jagged technological frontier." Outside this frontier, however, AI outputs may be inaccurate, misleading, or otherwise unhelpful. For example, today's AI models can win gold medals in mathematical competitions but cannot reliably tell time (Sajadieh et al., 2026). Consequently, the effective use of AI requires ongoing human oversight, evaluation, and judgment, with humans remaining responsible for assessing and refining AI-generated outputs (Mollick, 2024).

This evolving division of labor raises important questions about the future of knowledge and creative work. Earlier waves of automation primarily mechanized routine tasks (Fernández-Macías et al., 2021) and contributed to a greater *perceived* routinization of work (Fernández-Macías et al., 2023; Nikolova et al., 2025). By contrast, GenAI increasingly participates in activities traditionally regarded as uniquely human, including idea generation, writing, and other forms of creative production. Emerging evidence suggests that human-AI collaboration can raise productivity and improve performance in a variety of knowledge-work settings (Brynjolfsson et al., 2025; Cui et al., 2024; Noy & Zhang, 2023). Yet, as AI generates initial outputs and humans increasingly evaluate, refine, and validate them, workers may become less likely to own the creative process from start to finish. Can they still derive enjoyment, satisfaction, and meaning from their work?

This question is central to debates about the future of work. A growing body of research suggests that technological change is unlikely to generate mass unemployment because any labor-displacing effects are often offset by compensating mechanisms that create new tasks and new forms of work (Hötte et al., 2023). Indeed, even optimistic accounts of AI-driven growth emphasize that production processes continue to depend on human judgment and oversight, at least for the foreseeable future (Jones, 2026). The critical issue, therefore, is not whether people will continue to work, but what role they will take and how they will experience it within the increasingly AI-infused work processes. This question is particularly important because a large body of literature in labor economics shows that perceived meaning at work is a key determinant of worker well-being, labor supply, and effort (Ariely et al., 2008; Cassar & Meier, 2018; Cnossen & Nikolova, 2025; Hendriks & Cnossen, 2026; Nikolova & Cnossen, 2020; Non et al., 2022).

Yet, the literature on the impact of AI on well-being and job quality remains limited. Population-level studies suggest that AI exposure has little effect on job quality. For example, Giuntella et al. (2025) find that occupational AI exposure is generally unrelated to job and life satisfaction and may even improve health satisfaction, although direct work with AI is associated with a small reduction in subjective well-being. Similarly, evidence from Finland shows no relationship between individual AI use intensity and job satisfaction, although more intensive users report slightly higher work engagement (Bryson et al., 2026).

The small experimental literature paints a somewhat more nuanced picture. Noy and Zhang (2023) find that access to ChatGPT increased job satisfaction, but the effect disappeared after two weeks. Using a vignette experiment, Sadeghian and Hassenzahl (2022) find that people perceive collaborating with a human coworker as having higher task significance than collaborating with an AI coworker. In another vignette experiment, Sadeghian et al. (2024) show that meaningfulness is higher when workers remain actively involved in decision-making and interact with AI as teammates rather than merely supervising or monitoring automated systems. Finally, using a vignette-based design, Zhang and Amos (2026) find that experimental subjects perceived work as less meaningful when AI performed the creative component of a task, largely because workers perceived lower autonomy and fewer opportunities for skill utilization.

To our knowledge, this paper is the first to examine whether attributing identical outputs to AI rather than a human affects perceived meaningfulness, effort provision, and trust in the capabilities of the technology in a realistic task with a clear social purpose. While the few existing previous studies provide important insights, they focus primarily on human-AI collaboration arrangements, examining how the presence of an AI coworker, the allocation of tasks between humans and AI, and different forms of human-AI interaction shape perceptions of meaningful work. In contrast, we study a different and increasingly common form of human-AI interaction: the evaluation of AI-generated outputs. Rather than varying the structure of collaboration (Sadeghian et al., 2024) or the allocation of tasks (Sadeghian & Hassenzahl, 2022; Zhang & Amos, 2026), we hold the evaluated output constant and vary only its attributed source, allowing for a cleaner causal analysis. An additional distinct feature of our experiment is that participants actively performed an actual task by evaluating the quality of output, allowing us to observe responses in a behavioral context. By contrast, existing vignette experiments generally rely on hypothetical scenarios in which subjects are asked to evaluate the meaningfulness of an imagined job or task.

In our experiment, participants first learned about a minor public health problem related to insufficient water consumption. We then asked them to help select an effective slogan for a public awareness campaign by evaluating the quality of three slogans. The slogans are identical across the experimental conditions and differ only in their randomly assigned source, which is described as either a human professional or an AI system. This design allows us to isolate the effect of AI attribution from differences in output quality or task content. Before and after the evaluation, participants report their perceptions of the task's meaningfulness. At the end of the survey, they are also invited to provide a slogan of their own.

Moreover, unlike other studies in the literature, our evidence comes from two identical experiments conducted in nationally representative surveys in the United States and the Netherlands. The resulting sample is substantially larger and more representative than those in most previous experimental studies of AI and meaningful work, which have relied on smaller groups of online participants (Sadeghian & Hassenzahl, 2022; Sadeghian et al., 2024; Zhang & Amos, 2026).

We find that attributing identical outputs to AI rather than to a human reduces perceived task meaningfulness by about 0.07 standard deviations, which is a small effect. This estimate is broadly consistent with the emerging population-level studies documenting limited impacts of AI on job satisfaction and well-being outcomes (Giuntella et al., 2025; Bryson et al., 2026). Vignette-based studies of AI and meaningful work (Sadeghian & Hassenzahl, 2022; Zhang & Amos, 2026) point in the same negative direction but report stronger responses, although their effect sizes are not directly comparable due to differences in experimental design.

Importantly, we find that AI-labeling reduces the probability of contributing their own slogan (i.e., our proxy for effort) by 3.4 percentage points, or about 13.1% relative to the sample mean, which is a sizeable effect. This effect is mainly driven by the Dutch sample.[1] The finding suggests that attributing creative work to AI can discourage

[1] The only somewhat related study on effort and AI finds the opposite pattern. Specifically, Abel and Johnson (2025) report that AI labeling increases effort in a follow-up task. The measures across our paper and that of Abel and Johnson (205) are not directly comparable, however: their study captures effort as

individuals from investing effort in the task itself. Such behavioral responses may help explain why the large productivity gains observed in experiments with AI often fail to translate into comparable productivity gains at the aggregate level (Del Rio-Chanona et al., 2025; Filippucci et al., 2024). If workers exert less discretionary effort when AI plays a prominent role, improvements in technological capability may be offset, at least in part, by lower human effort.

In addition to evaluating the slogans and reporting their perceived task meaning, respondents rate how much they trust the author (AI or human, depending on the experimental condition) to generate an effective slogan. Consistent with the literature on AI attribution and evaluations of creative output, we find that respondents evaluate AI-labeled slogans as lower in quality, despite identical slogans across conditions (Abel & Johnson, 2025; Bellaiche et al., 2023; Millet et al., 2023). We also find lower situational trust in the AI author's ability to produce a quality slogan, consistent with the broader literature documenting algorithm aversion (Castelo et al., 2019; Dietvorst et al., 2015), and with evidence that explicit AI disclosure erodes trust in the attributed source (Schilke & Reimann, 2025). These results confirm that the attribution manipulation registered meaningfully with respondents.

Taken together, these findings show that while AI attribution strongly shapes how people evaluate creative output and its source, its impact on the meaning they derive from the task and the effort they are willing to exert is negative but modest. These modest effects are consistent with broader evidence suggesting that AI is not yet deeply integrated into most workers' day-to-day tasks at this stage (Bick et al., 2026). People's sense of task meaning and willingness to contribute may not yet have substantially adjusted to AI's growing presence in creative and knowledge work.

2. Related literature

2.1. AI in the labor market

As with previous major technological breakthroughs (Frey & Osborne, 2017; Küsters & Schneider, 2025; Mokyr et al., 2015), GenAI has raised concerns about large-scale job displacement. However, the literature to date does not yet point to definitive negative job loss effects (Acemoglu et al., 2022; Comunale & Manera, 2024; Georgieff & Hyee, 2022; Hartley et al., 2024). This may be because AI's proliferation implies task reorganization within occupations (Brynjolfsson et al., 2018), or because AI is not yet deployed at scale in the labor market. This is in line with survey evidence from AI-using firms, among which roughly 27 percent reported replacing worker tasks, while only around 5 percent reported employment changes due to AI (Bonney et al., 2024), implying

continuation of an instrumental reading task, whereas we measure effort as voluntary provision of new creative input.

that, at present, AI may more predominantly influence the organization and content of work before it induces any displacement.

Yet even firms that have adopted AI have not seen a deep integration at the worker level. Although 43 percent of US workers and 32 percent of European workers surveyed used GenAI for their jobs in early 2026, GenAI accounted for only 5.2 percent of total US work hours and between 1.5 and 2.8 percent in the European countries surveyed (Bick et al., 2026). This suggests that AI has spread rapidly but is not utilized at scale across entire jobs or work processes.

Task-based frameworks and exposure-mapping studies show that GenAI has significant potential to reach across cognitive and knowledge-intensive occupations (Acemoglu, 2025; Eloundou et al., 2024; Felten et al., 2023). GenAI outperforms humans in structured, language-intensive, or feedback-rich tasks, such as writing, summarizing, coding, customer support, consulting, marketing, and information processing (Noy & Zhang, 2023; Brynjolfsson et al, 2025; Dell'Acqua et al., 2026; Cui et al., 2024; Sajadieh et al., 2026). In these tasks, the use of GenAI can reduce completion time and improve output quality (Brynjolfsson et al., 2025; Cui et al., 2024; Dell'Acqua et al., 2026; Noy & Zhang, 2023), increase the number of outputs workers can produce (Dell'Acqua et al., 2026; Cui et al., 2024), and provide suggestions or feedback that help workers move more quickly through a task (Brynjolfsson et al., 2025).

These benefits, however, are uneven across *workers* and *tasks*. Several studies find that working with GenAI leads to the largest increases in productivity and quality among less-experienced and low-performing workers (Brynjolfsson et al., 2025; Cui et al., 2024; Noy & Zhang, 2023). However, this conclusion is only true for relatively straightforward and concrete tasks, such as coding, brainstorming, or writing text, where AI can augment the capabilities of the worst performers and bring them to the level of average performers. These concrete tasks are within the so-called "jagged AI frontier," in which AI capabilities outstrip those of humans. In such settings, AI can partly substitute for task-specific experience by making guidance, examples, and feedback more accessible (Brynjolfsson et al., 2025; Dell'Acqua et al., 2026; Cui et al., 2024). Yet, AI can lead to lower productivity and quality in complex open-ended tasks that require contextual judgment and real-world complexity, i.e., tasks outside the jagged frontier. For example, in a complex task such as running a business, highly skilled entrepreneurs benefited from AI, while low performers experienced a decline in their business performance (Otis et al., 2024). The difference in how much each benefited from GenAI was not due to different prompting but to choices about which pieces of LLM-generated advice entrepreneurs followed. When workers lack the judgment to assess AI-generated suggestions, performance can decline (Otis et al., 2024). An experiment among student debating teams, which involves a mixture of complex cognitive and social tasks, also confirms this conclusion (Roldán-Monés, 2024). Indeed, even among high-performing, motivated knowledge workers at a top consulting company, GenAI tools only improved performance on tasks at the frontier, not beyond it (Dell'Acqua et al., 2026). For tasks outside the frontier, workers who blindly followed the tools' output were 19 percentage points less likely to produce correct solutions (Dell'Acqua et al., 2026).

Taken together, the current body of evidence suggests that GenAI is better understood as a technology that reallocates subtasks between humans and machines to improve productivity and output quality, rather than as a wholesale substitute for knowledge work. While AI increasingly performs the initial stages of cognitive work, including drafting, generating ideas, summarizing information, and producing first-pass outputs, its performance remains highly uneven across tasks. Consequently, productivity gains depend not only on the capabilities of the AI model itself but also on workers' ability to evaluate, interpret, and selectively incorporate AI-generated outputs. In many work settings, the critical human contribution has shifted (or will likely shift) from producing initial outputs to exercising judgment over them. Consequently, GenAI reallocates tasks between humans and machines, with humans increasingly occupying evaluative, supervisory, and decision-making roles. Human judgment, therefore, remains central to realizing the productivity benefits of AI (Mollick, 2024).

2.2. Meaning at work

Task meaning refers to the extent to which a work-related activity is perceived as personally or socially important. A substantial economics literature examines how experimentally varying task significance affects effort, productivity, and reservation wages (see Hendriks & Cnossen, 2026 for an overview).

A substantive concern in this body of literature is whether short experimental interventions conducted outside actual workplaces can meaningfully alter participants' perceptions of task significance (Bäker & Mechtel, 2018; Cnossen & Nikolova, 2025). Existing evidence suggests that they can. Experimental studies typically manipulate the perceived purpose or usefulness of a task, for example, by informing participants that their work will contribute to research, benefit others, or be used in practice, or that it will be ignored, discarded, or serve no meaningful purpose (Ariely et al., 2008; Chandler & Kapelner, 2013; Kesternich et al, 2021; Kosfeld et al., 2017). By manipulating the perceived meaningfulness of work, these experiments show that participants exert more effort, produce higher output, and, in some cases, accept lower compensation when they perceive a task as more significant (Ariely et al., 2008; Bäker & Mechtel, 2018; Non et al., 2022).

Furthermore, task meaningfulness differs from work meaningfulness (Cnossen & Nikolova, 2025). Work meaningfulness is a broader evaluative assessment of one's job and reflects the extent to which work satisfies the fundamental psychological needs for autonomy, competence, and relatedness (Nikolova & Cnossen, 2020). By contrast, task meaningfulness concerns the perceived significance, purpose, or value of a specific activity. While economists have devoted considerable attention to the consequences of task significance for effort and performance, much less is known about how *characteristics of the task itself* shape perceptions of task meaningfulness. Related research has examined how job characteristics influence work meaningfulness (Nikolova & Cnossen, 2020; Cnossen & Nikolova, 2025; Hendriks & Cnossen, 2026) and how technological change affects work meaningfulness (Nikolova et al., 2024). However, these studies focus on broader working conditions and job characteristics rather than on whether attributing a task or its output to a technological system alters the perceived meaning of performing that task, which is the main contribution of the present paper.

2.3. AI, creativity, meaning, and effort

Several experimental studies have examined how labeling identical output as AI- rather than human-generated shapes evaluations of that output and its creator. Across creative and professional contexts, AI attribution consistently produces less favorable assessments of quality, creativity, and competence, even when the underlying work is identical (Abel & Johnson, 2025; Bellaiche et al., 2023; Millet et al., 2023).

For example, Bellaiche et al. (2023) examine whether people evaluate identical artworks differently depending on their perceived creator. Across two small-scale experiments (N = 149 and N = 148), participants rated AI-generated paintings that were randomly labeled as either human-created or AI-created. Human-labeled artworks received significantly higher ratings for beauty and monetary worth, despite being identical to the AI-labeled artworks. The findings suggest that people value not only artistic outputs themselves but also the human involvement underlying their creation. Furthermore, Abel and Johnson (2025) study whether labeling creative writing as AI-generated affects both subjective evaluations and revealed preferences. In an incentivized online experiment with 654 US participants, all respondents read the same short story, which had in fact been generated by GPT-4 but was randomly labeled as either written by a human author or by ChatGPT. Participants then evaluated the story and reported their willingness to pay to read the ending. Consistent with previous research, participants rated the AI-labeled story significantly less favorably on dimensions such as authenticity, creativity, and literary merit. However, the AI labeling had no effect on willingness to pay or time spent reading the story. The authors conclude that although people express a bias against AI-generated creative content, these preferences do not translate into corresponding incentivized behavioral choices.

Similarly, Millet et al. (2023) examine whether AI attribution reduces appreciation of artistic output. Across four experiments, spanning evaluations of music (Study 1, N = 206, the Netherlands), paintings (Study 2, N = 298, UK; Study 3, N = 404, UK), and art posters (Study 4, N = 800, UK), the authors show that artistic works receive less favorable evaluations when attributed to AI rather than to a human creator. Using a similar attribution design to Bellaiche et al. (2023) and Abel and Johnson (2025), the studies vary whether artistic works appear with an AI or human creator label, allowing the authors to isolate the effect of perceived source from the content of the work itself. Across the studies, AI-labeled art elicited less awe and, in the studies where creativity was measured, lower creativity ratings. In the larger preregistered poster experiment, AI attribution also reduced participants' willingness to buy a poster in a hypothetical decision, which differs from Abel and Johnson's (2025) incentivized willingness-to-pay outcome. The authors further show that these responses operate through perceived creativity: participants considered AI-labeled art less creative, which reduced their awe and, in turn, lowered purchase preference. This bias is also stronger among participants who believe that creativity is a uniquely human capacity. Overall, Millet et al. (2023) suggest that negative evaluations of AI-labeled artistic output do not simply reflect judgments about the output itself, but also broader beliefs about human uniqueness in creative domains.

Evidence on effort as an outcome of AI attribution remains scarce. The only exception is Abel and Johnson (2025) mentioned above, who experimentally varied whether an identical short story was labeled as AI-generated or human-written. Participants first reported their willingness to pay and to perform transcription work to read the remainder of the story. Those who proceeded to the second stage then completed a one-minute transcription task. The authors use the number of words transcribed as a revealed-effort measure and find that participants exposed to the AI label transcribed approximately 19% more words than participants who believed the story was human-written. The results suggest that negative evaluations of AI-generated content do not necessarily translate into lower effort or engagement.

Furthermore, a small body of literature examines how human-AI collaboration shapes perceived job quality. Specifically, Noy and Zhang (2023) use a randomized experiment among 444 professionals performing realistic writing tasks and show that access to ChatGPT improves both productivity and output quality while reducing the time required to complete tasks. Importantly, participants with access to ChatGPT also report higher job satisfaction and self-efficacy, but the effects appear temporary. Furthermore, Sadeghian and Hassenzahl (2022) use a vignette experiment with 104 participants to examine collaboration between a human worker and either a human colleague or an AI system in the context of organizing a large company meeting. Participants were assigned to either a more meaningful content-preparation role or a more routine organizational role. Across conditions, collaboration with an AI system was perceived as less motivating and meaningful than collaboration with a human colleague, and participants were more likely to view the AI as a subordinate or tool rather than as a teammate. In a related study, Sadeghian et al. (2024) further show that meaningfulness depends on the structure of the human-AI workflow: participants report higher job meaningfulness when they directly interact with AI, remain involved in decisions and actions, and retain accountability.

Finally, the study most closely related to ours is Zhang and Amos (2026), who use three small-scale vignette experiments among US respondents (Study 1: N = 157; Study 2: N = 166; Study 3: N = 379) to examine how AI involvement in creative work affects perceived work meaningfulness. Drawing on interviews with digital artists, they compare a traditional human-only creative process with two common AI-mediated workflows: AI prompting, in which workers generate outputs by crafting prompts for AI systems, and AI mending, in which workers refine or correct AI-generated outputs. Across all three studies, respondents perceive workflows involving AI as less meaningful than the human-only process. The authors further show that lower perceived autonomy and fewer opportunities for skill utilization explain this relationship, and that the negative effect is strongest among individuals with a more intrinsic orientation toward work.

However, our study significantly differs from that of Zhang and Amos (2026) and uniquely contributes to the literature. First, Zhang and Amos (2026) use vignette experiments to compare hypothetical creative workflows that differ in the extent and form of AI involvement. By contrast, we hold the creative output (i.e., the slogan) constant and vary only its attributed source. This design allows us to identify whether merely labeling identical output as AI-generated rather than human-generated affects perceptions of meaning and effort. Moreover, while Zhang and Amos (2026) focus primarily on perceived meaningfulness, we additionally examine voluntary effort provision, thereby capturing behavioral responses to AI attribution alongside subjective evaluations. We also perform the experiments as part of nationally representative surveys in two countries. To our knowledge, our paper is the first to connect the literature on AI attribution with the literature on meaningful work by examining whether labeling identical outputs as AI-generated rather than human-generated affects perceived task meaningfulness, trust, and effort provision.

3. Conceptual framework

Our conceptual framework builds on the idea that task meaning derives from the perceived significance and purpose of one's contribution to the work process. While prior research has primarily manipulated task significance directly by varying whether a task benefits others or serves a useful purpose (see Section 2.2), we argue that task meaning may also depend on how individuals understand their role within a broader work process. The same task can carry a different meaning depending on whether individuals perceive their contribution as necessary, valued, and consequential.

Drawing on the experimental literature on task meaning (e.g., Ariely et al., 2008; Chandler & Kapelner, 2013; Kosfeld et al., 2017), we propose that AI attribution can influence task meaning by changing how respondents interpret the significance of their role. In our setting, participants evaluate slogans intended for a public health campaign. Although the evaluative task itself remains identical, respondents are informed whether the slogans were created by a human professional or by AI. This information may shape how they understand the importance of their own contribution. While participants perform only the evaluative stage, the meaning they derive from it may depend on what preceded it, specifically, whether the upstream creative work reflects human or AI authorship, given that creative work is traditionally a uniquely human domain, owing to the time, skill, and intentionality it demands. When AI performs that upstream stage instead, the perceived significance of the subsequent evaluative task may shift, even though nothing about participants' own work has changed.

AI attribution may preserve or even enhance task meaning if individuals perceive their evaluative role as a substantive and necessary human contribution, i.e., being the "human in the loop" (Mollick, 2024); specifically, as a form of oversight and quality control that remains consequential regardless of who produced the upstream output. In this case, knowing that AI performed the upstream stage may amplify rather than diminish the felt significance of their own contribution, and hence, meaning.

Alternatively, AI attribution may reduce task meaning if individuals perceive AI as having performed the most valuable or creative component of the process. When AI has performed the upstream creative work, participants may question the significance of their own contribution. Specifically, they may perceive their role as less agentic, reduced to checking and approving AI output rather than exercising meaningful judgment as part of a human-led process, and their sense of connection to a broader purpose may erode accordingly. The perceived significance of the evaluative task may thus diminish not because of anything participants actually do differently, but because of what their work is understood to build upon.

Because both mechanisms are theoretically possible, the overall effect of AI attribution on task meaning remains an open empirical question. We therefore formulate the following two-tailed hypothesis:

H1. Attributing creative output to AI rather than a human professional influences the perceived meaningfulness of the task.

This reasoning also applies to effort provision. We conceptualize effort in this context as a voluntary contribution that requires respondents additional time, attention, and creative thought.[2]

AI attribution may affect effort because it could change how respondents perceive the value and necessity of human contribution. On the one hand, participants may be less willing to contribute a slogan of their own if AI authorship signals that human creative input is less necessary. If a machine is capable of and has already performed the creative portion of the work, individuals may see little point in spending time and effort to come up with a slogan of their own. AI performing the upstream creative stage may also threaten participants' creative identity: if individuals perceive AI as encroaching on a domain they associate with human skill and imagination, they may feel that their own creative contribution is devalued before they even make it, reducing their motivation to contribute. Moreover, knowing that AI rather than a human professional performed the upstream creative work may erode their sense of connection to the task's broader purpose. Since many individuals view creative work as a distinctly human endeavor, AI authorship may diminish the felt significance of any further human contribution, including the voluntary one.

Alternatively, AI attribution may increase effort if participants perceive human input as particularly important as a complement to AI-generated content. In this case, participants may feel motivated to demonstrate that human creativity adds something the machine cannot, or to improve upon the AI-generated content. The knowledge that a machine, rather than a human, performed the creative work may thus heighten, rather than diminish, their sense that human input is valuable and worth providing.

The overall effect of AI attribution on effort is, therefore, ambiguous. As with meaning, we treat the direction of the effect as an open empirical question.

H2. Attributing creative output to AI rather than a human professional affects voluntary effort provision.

4. Experimental Design, Procedures, and Data Cleaning

4.1. Experimental design

Our pre-registered survey experiment employs a between-subjects design to test whether labeling written content as AI-generated influences individuals' perceptions of task meaning and their willingness to exert effort.[3] Participants are randomly assigned to

[2] We operationalize effort as respondents' willingness to go beyond the required evaluative task by contributing a slogan of their own.

[3] We preregistered the experiment in the American Economic Association's registry for randomized controlled trials before the start of the project (RCT ID AEARCTR-0015363). The Research Officer of the Faculty of Economics and Business of the University of Groningen reviewed and approved our project's

one of two experimental conditions through pre-loaded randomization in the survey questionnaire: i) the treatment group, where we present slogans labeled as AI-generated, and ii) a control group, where we attribute the same slogans to a human marketing professional. We present the same three slogans to participants in both groups. The manipulation is therefore purely informational, consisting only of the attributed author ("an Artificial Intelligence (AI)-powered marketing software" versus "a marketing professional"), making this a "conceptual AI experiment" in the Alekseev and Strobel (2026) classification. Random assignment to the treatment and control groups allows us to interpret differences in outcomes as the causal effect of AI attribution. The main advantage of our research design is that it presents a realistic knowledge-work task in a short survey experiment. This makes our approach relatively cost-effective, highly feasible, and replicable across different populations. The main disadvantage is that our study subjects do not interact with AI and our results may not fully capture their actions and responses in the real world (Alekseev & Strobel, 2026).

4.2. Data and samples

We conducted our survey experiment in the context of two similar, nationally representative surveys. In the US, we fielded the study in February-March 2025 via the Understanding America Study (UAS), a nationally representative, online, probability-based panel administered by the Center for Economic and Social Research at the University of Southern California (USC). The UAS continually interviews about 15,000 respondents and has conducted more than 600 surveys covering a wide range of themes such as health, cognitive assessments, work, and subjective well-being (Kapteyn, et al., 2024). A total of 2,308 respondents were selected, of whom 1,511 completed the survey, yielding an overall response rate of 65.4%. Respondents completed the survey experiment in English.

In the Netherlands, we fielded the study in April 2025 via the Longitudinal Internet Studies for the Social Sciences (LISS) managed by Centerdata at Tilburg University. The LISS is a probability-based panel including about 7,000 respondents from about 5,000 households that complete online questionnaires every month. It frequently features additional survey experimental modules (Das & Knoef, 2019; Scherpenzeel, 2011). LISS selected 2,825 respondents for our experiment, of which 2,117 completed the survey, yielding a 74.9% response rate. Respondents completed the survey experiment in Dutch.

A unique advantage of our study is the use of the LISS and UAS panels, which were both developed by Arie Kapteyn. As such, the two probability-based internet panels share a common design, including aligned sampling procedures, online infrastructure, and data

data management plan (FEB-20250210-15570). The UAS obtained IRB approval from the University of Southern California's IRB Board (IRB00010793). The LISS panel did not require a separate IRB approval.

processing practices. This rare cross-national alignment minimizes methodological heterogeneity and allows for direct cross-country comparisons of results.

4.3. Experimental procedure

We begin the experiment by informing all participants about a minor public health issue related to insufficient water intake. We then tell them that they will evaluate health campaign slogans promoting healthy water consumption and ask participants to rate how meaningful they find the task of evaluating these slogans on a 7-point scale (1 = not meaningful at all, 7 = extremely meaningful). This allows us to collect baseline meaning associated with the task. Before the treatment, respondents reported relatively high levels of task meaning. Average pre-treatment meaning scores were 4.87 in the US and 5.09 in the Netherlands on a 1-7 scale, both above the scale midpoint of 4. These statistics are slightly lower but in line with the pre-treatment mean ratings reported for the task of organizing a meeting and content preparation in Sadeghian and Hassenzahl (2022), who used the same scale.

After rating how meaningful they find the task of evaluating public health campaign slogans, the respondents move to the slogan evaluation part of the survey experiment, where they read slogans attributed to either AI-powered marketing software or a marketing professional, depending on their treatment condition. Each respondent evaluates three slogans by rating them on creativity and persuasiveness on a 5-point scale (1=not creative/persuasive, to 5=very creative/persuasive). We clarify what we mean by creativity and persuasiveness, to ensure that respondents have the same working definition.

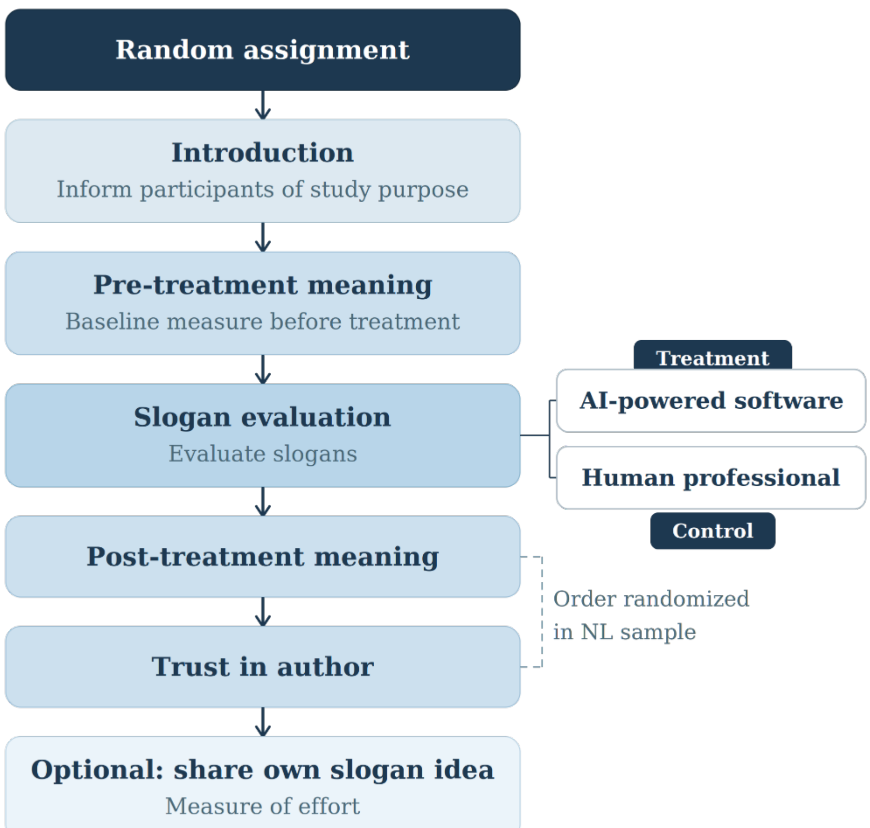
After evaluating the slogans, participants proceed to a post-treatment question assessing once again the meaning they associate with the task. Following this, we ask participants to rate how much they trust an AI-powered marketing software or a marketing professional to write a slogan that is (i) creative and (ii) persuasive on a 5-point scale (1=not at all, to 5=very much). Last, we invite participants to share a slogan idea of their own if they have one. We present the experimental design and question sequence in Figure 1 and the full questionnaires in Appendix A.

The UAS and LISS surveys use the same question sequence, except for the order of the post-treatment meaning and trust questions (Questions 5 and 6a/6b in Appendix A; see also Figure 1). In the UAS survey, respondents first answer the post-treatment meaning question and then report their trust in the author's ability to produce creative and persuasive slogans. In the LISS survey, which was fielded shortly after the UAS survey, we randomize the order of these questions.

Specifically, within both the treatment and control groups in the LISS survey, respondents are randomly assigned to one of two question-order conditions. One group first answers the post-treatment meaning question and then reports their trust in the author's ability to produce a creative and persuasive slogan. The other group first reports their trust in the author's ability to produce a creative and persuasive slogan and then answers the post-treatment meaning question.

While this deviates from our pre-registration plan, we introduce this order randomization because the trust questions explicitly remind respondents of the attributed author of the slogans: AI-powered marketing software in the treatment group and a marketing professional in the control group. Consequently, these questions may make the author attribution more salient. Randomizing the question order in the LISS survey allows us to test whether responses to the meaning question differ depending on whether respondents have already answered the trust questions.

Figure 1. Experimental design and survey question order



Source: Authors' elaboration.

4.4. Outcome variables

We collected baseline and post-treatment measures of task meaning using a 7-point scale ranging from 1 (not meaningful at all) to 7 (extremely meaningful). We also construct a measure of change in task meaning as the difference between post-treatment and pre-treatment responses, which serves as one of our outcome variables. Additionally, we use post-treatment meaning as an alternative measure, controlling for baseline meaning in the corresponding specifications.

We measure effort through an open-ended question inviting respondents to submit a slogan idea of their own if they have one. All three authors of this paper independently reviewed the open-ended responses and coded whether each respondent submitted a genuine slogan suggestion. Responses consisting of irrelevant text, unrelated comments, or other non-slogan content were coded as invalid. We then construct a binary effort indicator equal to one if at least two of the three raters coded the response as a valid slogan, and zero otherwise.

We capture trust in the author's ability to produce creative and persuasive slogans using two separate items, each rated on a 5-point scale from 1 (not at all) to 5 (very much). This measure of trust is situational because it is shaped by the immediate context, task, and information available at a given moment, rather than by stable individual tendencies or accumulated experience with a system (Hoff & Bashir, 2015; Nikolova & Angrisani, 2025). In our setting, it captures respondents' immediate confidence in the attributed source's ability to produce creative and persuasive output after evaluating the slogans. Because people tend to place lower trust in AI or algorithmically attributed sources than in equivalent human ones (Castelo et al., 2019; Dietvorst et al., 2015; Schilke & Reimann, 2025), we use these survey items to assess whether respondents meaningfully distinguished between AI and human authorship, serving as a complementary validation of the treatment manipulation.

Finally, we collect respondents' evaluations of the three slogans. For each slogan, respondents rate its creativity and persuasiveness on separate 5-point scales, ranging from 1 (not creative/persuasive) to 5 (very creative/persuasive). We calculate the average ratings on creativity and persuasiveness across the three slogans for each respondent. These measures allow us to examine how respondents evaluate the quality of identical slogan content labeled either as AI- or human-generated. Similar to the trust measures, we use the quality evaluations as a validation check of our design to assess whether our experiment reproduces the pattern documented in existing studies, namely that AI-labeled work is often rated less favorably than human-labeled work (Abel & Johnson, 2025; Bellaiche et al., 2023; Millet et al., 2023).

4.5. Randomization and balance

Our analysis sample comprises 1,511 respondents in the US and 2,117 in the Netherlands. Table 1 reports the balance on key sociodemographic, psychological, and pre-treatment characteristics between treatment and control groups in both samples, drawing on the background information available in both panels.[4] For each variable, we report group means, standard deviations, and number of observations, as well as total sample counts and mean differences with significance indicators from *t*-tests. We include the pre-treatment values of task meaning alongside standard background variables to verify that the two groups entered the evaluation task with comparable baseline assessments, providing an additional check on the validity of the randomization. For control variables with item non-response, we preserve observations by coding missing values into separate categories where appropriate, which is standard practice in the applied economics literature (e.g., Otrachshenko et al., 2023).

The experimental subjects are broadly similar across treatment conditions. In the US sample, no variable shows a statistically significant difference between the two groups. In the Dutch sample, we observe statistically significant imbalances on four variables: age above the median, tertiary education, marital status, and conscientiousness. Given the large number of balance tests, some statistically significant differences are expected to arise by chance alone. Participants in the treatment group tend to be older, less educated, more likely to be married, and more conscientious than those in the control group. We address these imbalances by including key sociodemographic controls in additional specifications.

[4] For the UAS, demographic variables come from each respondent's most recently completed household survey, available in the December 2024 comprehensive file curated by the UAS team. Personality traits are already computed by the UAS team come from UAS wave 593 (fielded August 2024), with missing values filled in from up to four preceding waves. For LISS, demographic variables correspond to the April 2025 background module, consistent with the panel's practice of aligning background variables to the month of data collection. Personality traits come from LISS Personality wave 16 (fielded May-June 2024), with missing values filled in from the immediately preceding wave. Following Goldberg's Big Five personality scales (Goldberg, 1992), we reverse-code negatively worded items and construct indices for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Imagination by summing the ten items belonging to each trait. We code the emotional-stability items so that higher values correspond to higher Neuroticism, making the measure more comparable to the UAS Neuroticism score. Our data availability statement furnishes further information and links to all datasets.

Table 1: Balance tests, by treatment status

Variable		Control (Human)		Treatment (AI)		Total N	Mean Diff.
		N	Mean/(SD)	N	Mean/(SD)		
Meaning (pre-treatment)	US	748	4.857 (1.690)	763	4.878 (1.632)	1,511	-0.021
	NL	1,035	5.078 (1.466)	1,082	5.104 (1.490)	2,117	-0.025
Age (above median>=55)	US	748	0.492 (0.500)	763	0.529 (0.499)	1,511	-0.038
	NL	1,035	0.535 (0.499)	1,082	0.587 (0.493)	2,117	-0.052**
Male	US	748	0.398 (0.490)	763	0.427 (0.495)	1,511	-0.029
	NL	1,035	0.465 (0.499)	1,082	0.476 (0.500)	2,117	-0.011
Education (Tertiary)	US	748	0.614 (0.487)	763	0.602 (0.490)	1,511	0.012
	NL	1,035	0.476 (0.500)	1,082	0.432 (0.496)	2,117	0.045**
Working	US	748	0.544 (0.498)	763	0.55 (0.498)	1,511	-0.006
	NL	1,034	0.502 (0.500)	1,082	0.496 (0.500)	2,116	0.006
Married	US	748	0.539 (0.499)	763	0.549 (0.498)	1,511	-0.010
	NL	1,035	0.508 (0.500)	1,082	0.558 -0.497	2,117	-0.050**
Household Income (High-income/third tertile)	US	747	0.163 (0.370)	762	0.175 (0.380)	1,509	-0.011
	NL	1035	0.305 (0.461)	1,082	0.298 (0.457)	2,117	0.008
(High) Agreeableness	US	741	0.560 (0.497)	760	0.558 (0.497)	1,501	0.001
	NL	980	0.520 (0.500)	1,030	0.524 (0.500)	2,010	-0.004
(High) Conscientiousness	US	742	0.524 (0.500)	759	0.499 (0.500)	1,501	0.025
	NL	980	0.508 (0.500)	1,030	0.551 (0.498)	2,010	-0.043*
(High) Extroversion	US	742	0.544 (0.498)	760	0.526 (0.500)	1,502	0.018
	NL	980	0.507 (0.500)	1,030	0.493 (0.500)	2,010	0.014
(High) Neuroticism	US	741	0.507 (0.500)	760	0.508 (0.500)	1,501	-0.000
	NL	980	0.520 (0.500)	1,030	0.505 (0.500)	2,010	0.016
(High) Openness	US	740	0.568 (0.496)	759	0.534 (0.499)	1,499	0.034
(High) Imagination	NL	980	0.528 (0.499)	1,030	0.521 (0.500)	2,010	0.006

Notes: This table reports the means and standard deviations of key variables for the treatment and control groups, along with t-tests of differences in means, for both the US and Dutch samples. The treatment group consists of participants who were informed that the slogans were created by AI, while the control group consists of participants who were informed that the slogans were created by a human. *** *p < 0.01*, ** *p < 0.05*, * *p < 0.10*.

4.6. Empirical strategy

We rely on Ordinary Least Squares (OLS) regressions to estimate the causal effect of the AI-labelled treatment on our key outcomes. Although effort is binary, we use OLS for ease of interpretation. Because treatment assignment is randomized, OLS yields unbiased estimates of the average causal effect without requiring additional identification assumptions. Our estimating equation is:

$$y_i = \alpha + \beta treat_i + \gamma X_i + \varepsilon_i \tag{1}$$

where y_i is the outcome of interest for respondent i. In separate regressions, we use either change in meaning, post-treatment meaning, or the binary effort indicator. In post-hoc analyses, we also utilize additional dependent variables: i) trust in the author's ability or average slogan evaluations of ii) creativity and iii) persuasiveness; $treat_i$ is an indicator equal to one if respondent i was assigned to the treatment group where participants were informed that the slogans were created by AI and zero otherwise; and X_i is a vector of sociodemographic controls. When the dependent variable is post-treatment meaning, X_i also includes the pre-treatment values of task meaning for each respondent. Because identification follows from random assignment, X_i enters the specification primarily to improve precision and to account for the imbalances observed in the Dutch sample, rather than to satisfy any identification requirement. The parameter of interest is β, which captures the average causal effect of AI attribution relative to human attribution.

We first estimate Equation (1) separately for the US and Dutch samples and then for the pooled sample of both countries. For the US sample, we apply probability weights using the UAS panel weights to account for the sampling design, while we analyze the LISS sample without weights, as these are unavailable and unnecessary by design for the LISS panel. We do not apply weights for the pooled sample estimations. We report robust standard errors throughout.

5. Results

5.1. Effects on task meaning

Tables 2 and 3 present OLS estimates of the effect of the AI label treatment on the primary outcome, perceived task meaning, in the US and Dutch samples, respectively. In both tables, Columns (1)–(3) report between-group differences in post-treatment meaning evaluations, controlling for the pre-treatment values of meaning, and Columns (4)–(6) report between-group differences in within-individual changes in meaning. We calculate these changes as the difference between post- and pre-treatment meaning scores, which accounts for baseline variation and tests whether the treatment altered meaning relative to participants' own initial ratings. We begin with a regression without controls, then add age and gender as exogenous demographic controls, and finally augment the specification with educational attainment, working status, marital status, and household income as additional controls. Table 3 additionally includes Columns (4) and (8), which control for differences in question ordering between the US and Dutch samples, as described in Section 4.3.

Across all specifications in the US sample (Table 2), the estimated treatment effects on perceived task meaning are negative, small in magnitude, and statistically insignificant, ranging from -0.081 to -0.084 for post-treatment meaning and from -0.100 to -0.104 for the change in meaning. We observe a similar pattern in the Dutch sample (Table 3), with the treatment effects ranging from -0.061 to -0.070 for post-treatment meaning and from -0.069 to -0.076 for the change in meaning, all statistically insignificant across the specifications.

Table 2: The effect of exposure to AI-labeled output on meaning, US Sample

	Meaning (post-treatment)			Meaning Δ		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.081	-0.081	-0.084	-0.100	-0.100	-0.104
	(0.079)	(0.079)	(0.079)	(0.086)	(0.086)	(0.086)
Basic controls	No	Yes	Yes	No	Yes	Yes
Full controls	No	No	Yes	No	No	Yes
Mean (control group)	4.703	4.703	4.703	-0.132	-0.132	-0.132
R^2	0.518	0.518	0.522	0.002	0.002	0.010
Observations	1,510	1,510	1,508	1,510	1,510	1,508

Notes: All specifications use OLS. The treatment indicator takes the value 1 if the participant evaluated slogans labeled as AI-generated and 0 if the slogans were labeled as human-generated. In Columns (1) to (3), the dependent variable is the post-treatment meaning score, measured on a 1–7 scale. These specifications include the pre-treatment meaning score as a control to account for baseline meaning ratings. In Columns (4) to (6), the dependent variable is meaning change, defined as the difference between post-treatment and pre-treatment meaning ratings. Basic controls include age and gender. Full controls additionally include education, employment status, marital status, and household income. The number of observations differs across specifications because models with the full set of controls exclude two respondents with missing household income data. All specifications use survey weights. Robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 3: The effect of exposure to AI-labeled output on meaning, Dutch Sample

	Meaning (post-treatment)				Meaning Δ			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.070	-0.065	-0.061	-0.061	-0.076	-0.069	-0.070	-0.070
	(0.053)	(0.053)	(0.053)	(0.053)	(0.057)	(0.057)	(0.057)	(0.057)
Order dummy				-0.026				-0.034
				(0.054)				(0.058)
Basic controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Full controls	No	No	Yes	Yes	No	No	Yes	Yes
Mean (control group)	4.606	4.606	4.606	4.606	-0.472	-0.472	-0.471	-0.471
R^2	0.396	0.397	0.399	0.399	0.001	0.004	0.006	0.006
Observations	2,117	2,117	2,116	2,116	2,117	2,117	2,116	2,116

Notes: All specifications use OLS. The treatment indicator takes the value 1 if the participant evaluated slogans labeled as AI-generated and 0 if the slogans were labeled as human-generated. In Columns (1) to (4), the dependent variable is the post-treatment meaning score, measured on a 1–7 scale. These specifications include the pre-treatment meaning score as a control to account for baseline meaning ratings. In Columns (5) to (8), the dependent variable is meaning change, defined as the difference between post-treatment and pre-treatment meaning ratings. Basic controls, included in Columns (2) to (4) and (6) to (8), include age and gender. Full controls in Columns (3), (4), (7), and (8) additionally include education, employment status, marital status, and household income. Columns (4) and (8) include an additional control for question order. The number of observations differs across specifications because full-control models exclude one respondent with missing values on working status. Robust standard errors are reported in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10.

While none of the coefficient estimates in Tables 2 and 3 are statistically significant, they share the same negative sign and similar magnitudes, with the US sample yielding somewhat larger coefficients. To formally assess whether the treatment operated differently in the two countries, we examine the equality of the treatment coefficients (Paternoster et al., 1998) and do not find a statistically significant difference for either post-treatment meaning (z =-0.352, p=0.725) or the change in meaning (z =-0.326, p=0.744). This suggests that the country-specific estimates are statistically indistinguishable from one another.

Additionally, we estimate an interaction model based on the pooled sample, where we interact the treatment indicator with the Dutch sample dummy. The sample dummy captures the average difference in outcomes as well as the cultural, and institutional differences between the two countries. In addition, we include a question order dummy to account for any possible differences arising from the different question order of the post-meaning question for half of the Dutch sample.[5] Consequently, the treatment

[5] Adding a control for the question order in the Dutch sample does not affect the estimated treatment effect in Table 3 on either post-treatment meaning or the change in meaning (see Table B1). Nevertheless, the pooled specifications in this section include an order dummy to account for the small questionnaire design difference across samples.

coefficient captures the estimated effect for the US sample, while the interaction term captures whether the Dutch treatment effect differs from that of the US.

Table 4: The effect of exposure to AI-labeled output on meaning, pooled results with an interaction control

	Meaning (post-treatment)	Meaning Δ
	(1)	(2)
Treatment	-0.096	-0.104
	(0.061)	(0.066)
Treatment*Dutch sample	0.033	0.030
	(0.081)	(0.087)
Dutch sample	-0.223***	-0.302***
	(0.058)	(0.062)
Full controls	Yes	Yes
R-squared	0.446	0.015
Observations	3,625	3,625

Notes: All specifications are estimated using OLS with the pooled US and Dutch sample. The treatment indicator takes the value 1 if the participant evaluated slogans labeled as AI-generated and 0 if the slogans were labeled as human-generated. The Dutch sample indicator takes the value 1 for respondents in the Dutch sample and 0 for respondents in the US sample. All specifications include the full list of controls: age, gender, education, employment status, marital status, and household income. Robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4 demonstrates that the treatment effect does not differ systematically between the two countries.[6] Therefore, we pool the two samples and estimate a common treatment effect, thereby increasing statistical power and improving the precision of our estimates.

[6] Tables 4 and 5 contain one additional observation relative to Tables 2 and 3. This is because we apply weights in the US regressions where one respondent has a weight of zero and is excluded from the weighted analyses in Tables 2. The pooled analyses in Tables 4 and 5 do not use survey weights and, therefore, include this respondent.

Table 5: The effect of exposure to AI-labeled output on meaning, pooled sample

	Meaning (post-treatment)		Meaning Δ	
	(1)	(2)	(3)	(4)
Treatment	-0.077*	-0.077*	-0.087**	-0.086**
	(0.040)	(0.040)	(0.043)	(0.043)
Order dummy		-0.028		-0.035
		(0.053)		(0.057)
Full controls	Yes	Yes	Yes	Yes
R-squared	0.446	0.446	0.015	0.015
Observations	3,625	3,625	3,625	3,625

Notes: All specifications are estimated using OLS on the pooled US and Dutch sample. The treatment indicator takes the value 1 if the participant evaluated slogans labeled as AI-generated and 0 if the slogans were labeled as human-generated. Columns (1) and (2) use post-treatment meaning as the dependent variable, measured on a 1–7 scale. Columns (3) and (4) use the change in meaning as the dependent variable, defined as post-treatment meaning minus pre-treatment meaning. All specifications include the full set of controls: age, gender, education, employment status, marital status, and household income. In columns (1) and (2), pre-treatment meaning is additionally included as a control. Columns (2) and (4) add an order dummy to control for variation in question order across survey versions. Robust standard errors are reported in parentheses. *** *p < 0.01, ** p < 0.05, * p < 0.10.*

In the pooled-sample models of Table 5, the estimated treatment effect on meaning remains negative but becomes marginally statistically significant. This suggests that labeling the slogans as AI-generated reduces the respondent's perception of the task's significance. To assess the magnitude of the effect, we express the treatment estimates in standard deviation units by dividing the treatment coefficient by the pooled sample standard deviation of the outcome. For post-treatment meaning, the coefficient of -0.077 in Table 5, Model (1), corresponds to an effect size of -0.047 standard deviations (-0.077/1.625). For meaning change, the coefficient of -0.087 in Model (3) corresponds to an effect size of -0.066 standard deviations (-0.087/1.310). These effect sizes are small according to conventional benchmarks. The small negative effect of AI attribution on perceived task meaning is consistent with the mechanism outlined in Section 3, whereby AI authorship during the creative stage may diminish participants' perceived role and felt significance in contributing to the task.

Furthermore, our results dovetail with the emerging experimental literature on AI and meaningful work. Sadeghian and Hassenzahl (2022) and Zhang and Amos (2026) both find that people evaluate work involving AI as less meaningful than comparable human-centered work, especially when AI performs creative or decision-relevant parts of the task. Directly comparing effect sizes is not straightforward, however, as these studies rely on vignette-based designs with small convenience samples and report different effect-size metrics. Nevertheless, their raw mean differences and reported ANOVA effect sizes suggest potentially larger effects on perceived meaningfulness than we find here.

The modest magnitude of our estimates aligns more closely with the emerging population-level evidence documenting limited associations between AI exposure and workers' subjective well-being and job satisfaction (Bryson et al., 2026; Giuntella et al., 2025). Unlike those studies, our experiment does not examine actual AI use or workplace exposure; instead, it isolates the psychological effect of attributing a creative task to AI, so the designs and outcomes are not directly comparable.

Our results are also broadly consistent with evidence on AI's limited reach in the workplace (Bick et al., 2026). As outlined in Section 2.1, GenAI remains partially and unevenly integrated into everyday work, concentrated in specific tasks, and accounts for a small share of total work hours. In other words, although AI has spread rapidly, it has not yet become a routine substitute for or complement to human creative contribution. Its capacity to reframe the perceived significance of work may therefore remain limited for now.

5.2. Effects on effort

We examine the effect of the AI label treatment on our secondary outcome, effort, which we measure as a dummy variable indicating whether participants provided a valid slogan suggestion of their own at the end of the survey.

Table 6 details the country-specific estimates of the treatment effect on effort. In the US sample, the treatment coefficients are small and not statistically significant across all specifications, suggesting that AI attribution did not influence participants' willingness to contribute their own slogans. In the Dutch sample, the treatment coefficients are consistently negative and statistically significant, ranging from approximately −0.044 to −0.046, indicating that exposure to the AI label leads to a robust reduction in voluntary effort.

Table 6: The effect of exposure to AI-labeled output on effort, US and Dutch samples

	Effort, US sample			Effort, Dutch sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	-0.006	-0.007	-0.005	-0.046**	-0.046**	-0.044**	-0.044**
	(0.029)	(0.028)	(0.028)	(0.019)	(0.019)	(0.019)	(0.019)
Order dummy							0.018
							(0.019)
Basic controls	No	Yes	Yes	No	Yes	Yes	Yes
Full controls	No	No	Yes	No	No	Yes	Yes
Mean (control group)	0.260	0.260	0.260	0.269	0.269	0.269	0.269
R^2	0.000	0.012	0.032	0.003	0.007	0.020	0.021
Observations	1,510	1,510	1,508	2,117	2,117	2,116	2,116

Notes: All specifications are estimated using OLS. The treatment indicator takes the value 1 if the participant evaluated slogans labeled as AI-generated and 0 if the slogans were labeled as human-generated. Across all columns, the dependent variable, effort, is a dummy variable that takes the value 1 if the respondent provided a slogan idea of their own and 0 otherwise. Columns (1) to (3) present the results for the US sample and columns (4) to (7) for the Dutch sample. Basic controls (columns 2-3 and 5-7) include age and gender. Full controls (columns 3, 6 and 7) additionally include education, employment status, marital status, and household income. Column 7 includes an additional control for the difference in the order of questions between the US and Dutch samples. Specifications for the US sample are estimated using survey weights. Robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

While the point estimates differ in magnitude across the two countries, the difference is not statistically significant ($z = 1.150$, $p = 0.248$). To assess the common impact, we again pool the two samples (Table 7). The pooled coefficient estimate on the treatment is robust to the inclusion of the order dummy and statistically significant at the 5% level, suggesting that AI attribution reduces the probability of providing a slogan by 3.4 percentage points. In other words, respondents exposed to AI-generated slogans exert less effort than respondents in the human-authorship condition. Relative to the pooled sample mean (0.259), this corresponds to a reduction of about 13.1%. Overall, the effort results are consistent with the potential negative channel outlined in Section 3, whereby AI authorship of the upstream creative stage reduces participants' perceived motivation to contribute their own creative input. The negative effect is also driven mainly by Dutch respondents. The absence of a significant effect in our US sample likely reflects a combination of lower statistical power and a smaller effect size rather than a true null effect.

This finding differs from Abel and Johnson (2025), who find that participants in the AI-label condition transcribe about 19% more words in a follow-up transcription task, which the authors interpret as greater revealed effort. However, the greater revealed effort in Abel and Johnson (2025) pertains to a selected subset of participants who chose to exert effort to reach the end of the story. As such, it is not directly comparable to our full-sample measure.

Table 7: The effect of exposure to AI-labeled output on effort, pooled sample

	Effort	
	(1)	(2)
Treatment	-0.034**	-0.034**
	(0.014)	(0.014)
Order dummy		0.018
		(0.019)
Full controls	Yes	Yes
R-squared	0.018	0.019
Observations	3,625	3,625

Notes: Both specifications are estimated using OLS on the pooled US and Dutch sample. The treatment indicator takes the value 1 if the participant evaluated slogans labeled as AI-generated and 0 if the slogans were labeled as human-generated. The dependent variable, effort, is a dummy variable taking the value 1 if the respondent provided a slogan idea of their own and 0 otherwise. Both specifications include the full set of controls: age, gender, education, employment status, marital status, and household income. Column (2) additionally includes an order dummy to control for variation in question order across survey versions. Robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

5.3. Treatment salience

In addition to meaning and effort, we examine four further outcomes: respondents' trust in the attributed author's ability to produce creative and persuasive slogans, and their evaluations of the slogans' quality in terms of creativity and persuasiveness.[7] This additional analysis provides an important validation check of our experimental design. Prior research documents a general tendency to trust algorithmic outputs less than equivalent human-generated ones, particularly for subjective tasks (Castelo et al., 2019; Dietvorst et al., 2015), and experimental evidence links explicit AI disclosure to reduced trust in the attributed source (Schilke & Reimann, 2025). Additionally, experimental evidence shows that AI labels can affect evaluations of otherwise identical creative outputs, with AI-labeled work often rated less favorably than human-labeled work (Abel & Johnson, 2025; Bellaiche et al., 2023; Millet et al., 2023). Trust and quality assessments of the slogans, therefore, allow us to ascertain the validity of the experimental manipulation. Specifically, if the AI label lowers trust and the quality assessment, as the literature predicts, then the modest effects we find on perceived task meaning and effort are less likely to reflect failed experimental manipulation. Instead, they are more likely to capture genuine perceptions and behavioral reactions.

[7] Section 4.4 and Appendix A details how we elicited these questions in the survey.

Table 8 reports the regression estimates for trust in the attributed author's creative and persuasive ability in Columns (1) and (2), and the average evaluation scores on slogan creativity and persuasiveness in Columns (3) and (4), across the US sample (Panel A), the Dutch sample (Panel B), and the pooled sample (Panel C). Across both trust outcomes, AI attribution has a strong negative and statistically significant effect in all three panels. In the country-specific models, AI attribution reduces trust by approximately 0.48 to 0.50 points in the US sample and by about 0.30 points in the Dutch sample, measured on a 1-to-5 scale. The pooled estimates similarly show a drop of about 0.38 points in trust in both the author's creative and persuasive ability in response to the AI label, corresponding to about -0.37 standard deviations.

Columns (3) and (4) of Table 8 show that AI attribution also affects respondents' evaluations of the slogan quality. In the US sample, the coefficients for slogan creativity and persuasiveness are negative but not statistically significant. In the Dutch sample, AI attribution significantly reduces both slogan creativity ratings by 0.11 points and slogan persuasiveness ratings by 0.08 points. The pooled estimates similarly show that AI attribution reduces average creativity and persuasiveness ratings by about 0.09. In standardized terms, these pooled effects correspond to approximately −0.11 standard deviations. While smaller than the effect sizes of other papers, our results are consistent with the evidence that AI-labeled work receives less favorable evaluations (Abel & Johnson, 2025; Bellaiche et al., 2023; Millet et al., 2023). Overall, Table 8 shows that respondents clearly registered the AI attribution treatment, supporting the validity of the experimental design and the interpretation of the modest effects on task meaning and effort.

Table 8: The effects of exposure to AI-labeled output on trust and slogan quality evaluation, US, Dutch, and pooled samples

	Trust, creative slogan	Trust, persuasive slogan	Evaluation, slogan creativity	Evaluation, slogan persuasiveness
	(1)	(2)	(3)	(4)
Panel A: US sample				
Treatment	-0.475***	-0.499***	-0.055	-0.088
	(0.076)	(0.077)	(0.062)	(0.064)
Full controls	Yes	Yes	Yes	Yes
Mean (control group)	3.218	3.174	3.061	2.903
R^2	0.056	0.069	0.026	0.031
Observations	1,508	1,508	1,508	1,508
Panel B: Dutch sample				
Treatment	-0.306***	-0.303***	-0.111***	-0.082**
	(0.042)	(0.042)	(0.035)	(0.034)
Full controls	Yes	Yes	Yes	Yes
Mean (control group)	3.091	3.080	2.911	2.813
R^2	0.033	0.034	0.017	0.017
Observations	2,116	2,116	2,116	2,116
Panel C: Pooled sample				
Treatment	-0.384***	-0.376***	-0.094***	-0.088***
	(0.034)	(0.034)	(0.028)	(0.028)
Full controls	Yes	Yes	Yes	Yes
R^2	0.041	0.042	0.026	0.022
Observations	3,625	3,625	3,625	3,625

Notes: All specifications are estimated using OLS. The treatment indicator takes the value 1 if the participant evaluated slogans labeled as AI-generated and 0 if the slogans were labeled as human-generated. Panel A reports results for the US sample, Panel B reports results for the Dutch sample, and Panel C reports results for the pooled sample. In columns (1) and (2), the dependent variables measure trust in the author's ability to produce creative and persuasive slogans, respectively, both measured on a 1–5 scale. In columns (3) and (4), the dependent variables are the average creativity and persuasiveness ratings across the three evaluated slogans, respectively, with each slogan rated on a 1–5 scale. Full controls include age, gender, education, employment status, marital status, and household income. Pooled specifications additionally include a Dutch sample dummy. US specifications and corresponding control-group means use survey weights. Dutch and pooled specifications are unweighted. Robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

5.4. Robustness checks

We perform two robustness checks to verify that the main results do not simply reflect differences in respondent engagement. To this end, we control for survey completion time and self-reported interest in the survey topic. We do not include the question-order dummy in these checks because it has a negligible effect on the results, as shown before.

5.4.1. Survey duration

First, we examine whether the main treatment effects on meaning and effort remain stable after accounting for the time respondents spent completing the survey. Survey duration serves as a proxy for respondents' engagement and attentiveness. Respondents who completed the survey quickly may have processed the attribution manipulation less carefully or engaged with the outcome measures less thoroughly by clicking through quickly. This is more relevant to meaning than to effort, given that we ask respondents to rate meaning pre- and post-treatment, and less engaged respondents may have selected similar values on these items without genuinely reflecting on them, thereby attenuating the treatment effect. The impact of survey duration on effort is, in part, mechanical, as those creating their own slogans need more time to complete the survey. We examine whether AI-label treatment still affects respondents' effort conditional on the same duration.[8]

Survey duration is measured in minutes. Raw completion times are highly dispersed in the pooled sample (mean: 122.8 minutes; SD: 1,313.0 minutes), reflecting respondents who started the survey but completed it much later, sometimes on a different day. We therefore winsorize duration at the 5th and 95th percentiles before including it as a control variable.

Table 9 shows that including survey duration leaves the treatment effects on both outcomes largely unchanged. For meaning, the treatment coefficient is virtually identical to that in the main specification. For effort, the coefficient estimate on the treatment variable changes slightly from −0.034 to −0.029 but remains statistically significant. The coefficient on survey duration is small and statistically insignificant in the meaning regression (Column (1) of Table 9), suggesting that completion time does not confound the meaning outcome. As expected, survey duration is positively and significantly associated with effort: an additional survey minute is associated with a 5.2 percentage point higher probability of providing a slogan idea, suggesting that duration captures respondents' engagement with the open-ended task. Overall, the results indicate that the main treatment effects are not driven by differences in survey completion time.

[8] We do not include a survey duration control in the main specifications, because it may itself be affected by the treatment. For example, respondents assigned to the AI condition might move through the survey faster if they feel less invested in the evaluation task.

Table 9: The effects of exposure to AI-labeled output on meaning and effort, with a duration control

	Meaning Δ	Effort
	(1)	(2)
Treatment	-0.086*	-0.029**
	(0.043)	(0.014)
Survey duration	0.008	0.052***
	(0.008)	(0.003)
Full controls	Yes	Yes
R-squared	0.016	0.122
Observations	3,622	3,622

Notes: Both specifications are estimated using OLS. The treatment indicator takes the value 1 if the participant evaluated slogans labeled as AI-generated and 0 if the slogans were labeled as human-generated. In column (1), the dependent variable is meaning change, defined as the difference between post-treatment and pre-treatment meaning ratings. In column (2), the dependent variable is effort, measured as a dummy taking the value of 1 if the respondent provided a slogan idea of their own and 0 otherwise. Both specifications include the full set of controls: age, gender, education, employment status, marital status, and household income. Both specifications also include a Dutch sample dummy. Survey duration is measured in minutes and is winsorized at the 5th and 95th percentiles to limit the influence of implausibly short or long completion times. The number of observations is slightly lower in the specifications with a duration control because survey duration is missing for 3 respondents who otherwise have complete outcome and control data: two in the US sample and one in the Dutch sample. Robust standard errors are reported in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10.

5.4.2. Survey topic interest

We also account for respondents' reported interest in the survey or topic. Participants who found the survey topic more interesting may report higher perceived meaning and be more willing to invest effort in the voluntary slogan-provision task.[9] The questions regarding interest in the topic differ slightly between the two samples: US respondents were asked how interesting they found the *interview*, while Dutch respondents were asked whether they found the *topic* interesting. We harmonize both into a common 1-to-5 Likert-type measure, coded such that higher values indicate greater interest, and include it as a linear control variable.

Table 10 shows that the treatment effects remain stable after controlling for survey interest. The treatment coefficients are virtually unchanged from the main specification in Tables 5 and 7 for both meaning change and effort. Therefore, the main results are

[9] Similar to survey duration, reported interest is measured post-treatment. Therefore, respondents' interest ratings may also partly reflect their response to the AI attribution itself. We therefore do not include this variable in the primary specifications but only as a robustness check.

unlikely to be driven by differences in respondents' reported survey or topic interest. Survey topic interest is positively and significantly associated with both outcomes: respondents who reported greater interest in the survey or topic also reported higher meaning change and were more likely to provide a slogan idea.

Table 10: The effects of exposure to AI-labeled output on meaning and effort, with a survey interest control

	Meaning Δ	Effort
	(1)	(2)
Treatment	-0.084*	-0.033**
	(0.043)	(0.014)
Survey/topic interest	0.136***	0.080***
	(0.023)	(0.007)
Full controls	Yes	Yes
R-squared	0.027	0.055
Observations	3,623	3,623

Notes: Both specifications are estimated using OLS. The treatment indicator takes the value 1 if the participant evaluated slogans labeled as AI-generated and 0 if the slogans were labeled as human-generated. In column (1), the dependent variable is meaning change, defined as the difference between post-treatment and pre-treatment meaning ratings. In column (2), the dependent variable is effort, measured as a dummy taking the value of 1 if the respondent provided a slogan idea of their own and 0 otherwise. Both specifications include the full set of controls: age, gender, education, employment status, marital status, and household income. Both specifications also include a Dutch sample dummy. Survey/topic interest is a harmonized Likert-type measure coded from 1 to 5, with higher values indicating greater reported interest, and is entered linearly. The number of observations is slightly lower in specifications controlling for survey topic interest because this variable is missing for 2 respondents: one in the US sample and one in the Dutch sample. Robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

5.5. Heterogeneity

5.5.1. Socio-demographic characteristics

We next examine whether the effects of AI attribution vary across key socio-demographic characteristics in the pooled sample. This heterogeneity analysis is motivated by the technology acceptance literature, which identifies age, gender, and experience as relevant moderators of individuals' responses to new technologies (Venkatesh et al., 2003). Recent evidence on AI adoption points in a similar direction. LLM adoption differs by gender, age, and technology-related education, with men and individuals with technology-related education more likely to adopt these tools (Angrisani et al., 2026; Draxler et al., 2023). Gender gaps in GenAI adoption correlate with differences in perceived societal risks (Stephany & Duszynski, 2026). Women are less likely to use ChatGPT for work, whereas younger and less experienced workers adopt it more widely (Humlum & Vestergaard, 2024). Similarly, women and older individuals express lower trust in AI than men and younger individuals (Nikolova & Angrisani, 2025). Together, these findings suggest that exposure to, familiarity with, and concerns about AI may differ across sociodemographic groups.

We therefore test whether age, gender, tertiary education, and employment status moderate the treatment effects on meaning and effort. The results displayed in Table 11 provide little evidence that the treatment effects differ systematically across sociodemographic groups.[10] Males and those with tertiary education were more likely to provide their own slogan, but none of the treatment interaction terms are statistically significant for either meaning change or effort, suggesting that the negative effects of AI attribution are not concentrated among specific demographic groups. The main treatment coefficients remain generally negative across specifications.

[10] Tables B2 and B3 report the results for the US and Dutch samples, respectively.

Table 11: Heterogeneity analysis by socio-demographics, pooled sample

	Meaning Δ				Effort			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.041	-0.100*	-0.087	-0.150**	-0.051**	-0.030	-0.048**	-0.023
	(0.063)	(0.060)	(0.066)	(0.065)	(0.021)	(0.019)	(0.020)	(0.021)
Treatment*Older	-0.085				0.031			
	(0.087)				(0.029)			
Older	-0.074				0.025			
	(0.064)				(0.022)			
Treatment*Male		0.031				-0.011		
		(0.086)				(0.029)		
Male		0.034				0.061***		
		(0.060)				(0.022)		
Treatment*Tertiary Education			0.000				0.026	
			(0.087)				(0.029)	
Tertiary Education			-0.096				0.068***	
			(0.062)				(0.021)	
Treatment*Working				0.123				-0.023
				(0.087)				(0.029)
Working status				-0.088				0.030
				(0.065)				(0.023)
R-squared	0.016	0.015	0.015	0.016	0.019	0.019	0.019	0.019
Observations	3,625	3,625	3,625	3,625	3,625	3,625	3,625	3,625

Notes: All specifications are estimated using OLS on the pooled sample. The treatment indicator equals 1 if the participant evaluated slogans labeled as AI-generated and 0 if the slogans were labeled as human-generated. Columns (1) to (4) use meaning change as the dependent variable, defined as the difference between post-treatment and pre-treatment meaning ratings. Columns (5) to (8) use effort as the dependent variable, measured as a dummy equal to 1 if the respondent provided a valid slogan idea of their own and 0 otherwise. Each specification includes an interaction between the treatment indicator and one binary sociodemographic moderator. The moderators are age above the median, male, tertiary education, and working status. Age above the median equals 1 if the respondent is older than 55 and 0 otherwise. Male equals 1 for male respondents and 0 otherwise. Tertiary education equals 1 if the respondent has completed tertiary education and 0 otherwise. Working status equals 1 if the respondent is currently working and 0 otherwise. All specifications include the full set of controls and a Dutch sample indicator. Robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

5.5.2. Personality traits

As a second heterogeneity check, we examine whether responses to AI attribution exposure vary by personality traits. Recent studies suggest that attitudes toward AI may differ not only by sociodemographic characteristics, but also by individual psychological characteristics. Stein et al. (2024) and Grassini et al. (2025) show that personality traits may shape how individuals evaluate AI-related technologies, including whether they perceive them as novel, trustworthy, uncertain, or threatening. We therefore test whether the treatment effects on meaning and effort differ across respondents with different personality traits.

We conduct this analysis separately by sample because the available personality measures are not fully harmonized across the two panels. The UAS includes Big Five measures of Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness, whereas the LISS panel uses Goldberg's Big Five framework (1992), with Imagination replacing Openness as the fifth trait. We therefore avoid pooling these specifications and instead estimate country-specific interaction models.

Each specification interacts the treatment indicator with a binary indicator for high levels of a given personality trait, defined as being at or above the sample-specific median. Overall, we find limited evidence that personality traits systematically moderate responses to AI attribution. In the US sample, none of the interaction terms are statistically significant for either meaning change or effort.

In the Dutch sample, we observe one notable exception, whereby the interaction between treatment and high conscientiousness is positive and statistically significant for meaning change. In this specification, the treatment coefficient indicates that AI attribution lowers meaning change by 0.219 points among respondents below the conscientiousness median, while the positive interaction suggests that this negative effect is offset among respondents with higher conscientiousness. For respondents above the median in conscientiousness, the estimated treatment effect is close to zero and not statistically significant. This pattern is consistent with the idea that more conscientious individuals may continue to find the slogan evaluation task meaningful despite the AI attribution, whereas less conscientious respondents appear more sensitive to the AI label. We find no statistically significant personality-based heterogeneity for effort in the Dutch sample.

Table 12: Heterogeneity analysis by personality traits, US sample

	Meaning Δ					Effort				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment	-0.048	-0.007	-0.101	-0.026	-0.041	-0.005	-0.026	0.032	0.020	-0.035
	(0.128)	(0.117)	(0.124)	(0.131)	(0.134)	(0.037)	(0.039)	(0.040)	(0.046)	(0.038)
Treatment*High Agreeableness	-0.113					0.005				
	(0.173)					(0.056)				
High Agreeableness	0.118					0.062				
	(0.124)					(0.040)				
Treatment*High Conscientiousness		-0.212					0.044			
		(0.174)					(0.057)			
High Conscientiousness		0.173					-0.021			
		(0.121)					(0.040)			
Treatment*High Extroversion			0.003					-0.063		
			(0.172)					(0.056)		
High Extroversion			0.077					0.081**		
			(0.120)					(0.040)		
Treatment*High Neuroticism				-0.143					-0.043	
				(0.174)					(0.058)	
High Neuroticism				0.222*					-0.029	
				(0.124)					(0.041)	
Treatment*High Openness					-0.115					0.066
					(0.174)					(0.056)
High Openness					0.106					0.062
					(0.125)					(0.040)
R-squared	0.011	0.013	0.011	0.013	0.011	0.037	0.032	0.036	0.035	0.045
Observations	1,504	1,504	1,505	1,504	1,502	1,504	1,504	1,505	1,504	1,502

Notes: All specifications are estimated using OLS. The treatment indicator equals 1 if the participant evaluated slogans labeled as AI-generated and 0 if labeled as human-generated. Columns (1) - (5) use the change in meaning (post-treatment minus pre-treatment) as the dependent variable; Columns (6) – (10) use a dummy equal to 1 if the respondent provided their own slogan idea. Each specification includes an interaction between the treatment and a moderator. Moderators are binary indicators for high levels (at or above the median) of the Big Five personality traits: Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness. All specifications are estimated using survey weights. The number of observations differs slightly across specifications because, even after filling in missing personality measures from previous waves where possible, some respondents still have missing values for particular personality traits and are therefore excluded from analyses. Robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 13: Heterogeneity analysis by personality traits, Dutch sample

	Meaning Δ					Effort				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment	-0.167**	-0.219***	-0.044	-0.166*	-0.059	-0.027	-0.035	-0.021	-0.072**	-0.028
	(0.078)	(0.084)	(0.081)	(0.081)	(0.083)	(0.026)	(0.027)	(0.026)	(0.028)	(0.025)
Treatment*High Agreeableness	0.172					-0.043				
	(0.116)					(0.038)				
High Agreeableness	0.216***					0.106***				
	(0.081)					(0.028)				
Treatment*High Conscientiousness		0.264**					-0.027			
		(0.116)					(0.038)			
High Conscientiousness		-0.106					0.039			
		(0.082)					(0.028)			
Treatment*High Extraversion			-0.069					-0.053		
			(0.117)					(0.038)		
High Extraversion			0.001					0.090***		
			(0.082)					(0.028)		
Treatment*High Neuroticism				0.170					0.046	
				(0.116)					(0.038)	
High Neuroticism				-0.058					-0.054*	
				(0.084)					(0.029)	
Treatment*High Imagination					-0.035					-0.041
					(0.117)					(0.038)
High Imagination					-0.100					0.105***
					(0.083)					(0.029)
R-squared	0.007	0.007	0.005	0.005	0.006	0.030	0.022	0.027	0.023	0.030
Observations	2,009	2,009	2,009	2,009	2,009	2,009	2,009	2,009	2,009	2,009

Notes: All specifications are estimated using OLS. The treatment indicator equals 1 if the participant evaluated slogans labeled as AI-generated and 0 if labeled as human-generated. Columns (1) - (5) use the change in meaning (post-treatment minus pre-treatment) as the dependent variable; Columns (6) – (10) use a dummy equal to 1 if the respondent provided their own slogan idea. Each specification includes an interaction between the treatment and a moderator. Moderators are binary indicators of high levels (at or above the median) of the five personality traits in Goldberg's Big Five framework: Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Imagination. The number of observations differs slightly across specifications because, even after filling in missing personality measures from the previous wave where possible, some respondents still have missing values for particular personality traits and are therefore excluded from the analyses. Robust standard errors are reported in parentheses. *** *p < 0.01,* ** *p < 0.05,* * *p < 0.10.*

6. Discussion and Conclusion

This study provides the first experimental evidence on how AI attribution affects task meaningfulness and voluntary effort. It also examines its effects on trust in AI and perceptions of creativity and persuasiveness. To this end, we rely on survey experiments conducted as part of two nationally representative studies in the United States (N = 1,511) and the Netherlands (N = 2,117). Participants evaluated identical slogans promoting healthy water intake, randomly attributed to either AI-powered marketing software or a human marketing professional. This design allows us to attribute any differences in responses purely to the AI or human label rather than to the quality of the output itself, and to capture both subjective perceptions of meaning and voluntary behavioral effort.

The results show that AI attribution has a modest negative effect on task meaning. In the pooled sample of US and Dutch respondents, AI attribution reduces perceived task meaning by about 0.07 standard deviations, which is a rather small effect. This effect is consistent with the explanation that AI involvement in the upstream creative stage leads participants to perceive their own evaluative role in the task as less significant, even though the task itself and output reviewed remain unchanged.

We also find that AI attribution reduced voluntary effort. Specifically, participants exposed to the AI label were 3.4 percentage points less likely to provide an original slogan of their own than the human-labeled group, or about 13.1% relative to the mean. This effect is mainly driven by Dutch respondents and is also consistent with the explanation that AI authorship of the creative component reduces participants' motivation to provide their own creative input.

Alongside these main outcomes, we also find that AI attribution strongly reduces participants' trust in the author's ability to produce high-quality slogans and lowers their evaluations of slogan quality, in line with existing findings in the literature. These robust effects on trust and output quality validate the effectiveness of the experimental design by confirming that the AI label registered meaningfully with respondents. These findings further indicate that the modest effects on meaning and effort reflect genuine subjective and behavioral responses rather than a failed experimental manipulation. Finally, we find little evidence that the treatment effects on task meaning and effort provision vary systematically across gender, age, education, employment status, or most personality traits. The main results are also robust to controlling for survey duration and respondents' self-reported interest in the survey topic.

Our paper points to several opportune avenues for future research. First, future studies should examine work meaningfulness and task meaning in real workplace settings, where researchers can observe how repeated use of AI interacts with job characteristics such as autonomy, competence, relatedness, and organizational mission (Cassar & Meier, 2018; Cnossen & Nikolova, 2025). Second, future work should identify the mechanisms through which AI attribution affects task meaning and effort. Our design cannot distinguish whether these effects reflect concerns about job replacement, reduced task ownership, diminished opportunities for self-expression, or other psychological mechanisms. Third, future research should examine how organizational practices shape these effects. Organizations may mitigate negative responses by framing AI as a collaborative tool, increasing workers' control over its use, improving transparency about AI's role in task completion, investing in AI training, or strengthening workers' participation in decisions about AI adoption through works councils.

Our results suggest that, at the current stage of AI integration, individuals' sense of task meaning and willingness to exert creative effort decline only modestly when creative work is attributed to AI. This aligns with broader labor market evidence suggesting that, despite AI's potential to reshape work processes, its effects on job quality and subjective work experience remain limited (e.g., Bryson et al., 2026; Guintella et al., 2025) given current adoption levels (Bick et al., 2026).

Nevertheless, the relatively large decline in creative effort points to a possible explanation for the micro-macro productivity paradox. Although experiments often find sizeable productivity gains from AI at the individual level, these gains have yet to translate into comparable improvements in aggregate productivity (Del Rio Chanona et al., 2025; Filippucci et al., 2024). One reason may be that AI changes not only workers' capabilities but also their motivation. If workers invest less effort because they perceive their contribution as less meaningful or less attributable to themselves, potential productivity gains from AI may remain partly unrealized. This interpretation accords with recent evidence that AI-generated time savings often translate into task reallocation or reduced effort rather than higher output (Del Rio-Chanona et al., 2025).

Our findings raise important questions about the future of meaningful work. As AI becomes embedded in everyday work and increasingly performs creative and cognitive tasks, will workers continue to derive meaning from evaluating, refining, and supervising AI-generated output, or will AI gradually erode feelings of ownership, creativity, and purpose? Will the new tasks and job titles that arise from the AI integration in the workplace allow for the expression of human agency and purpose, or will they make work more monotonous and degrading? These questions echo Mollick's (2024) argument that AI will fundamentally reshape the creative process. Our evidence suggests that such changes remain modest today, but whether they grow as AI becomes a more pervasive collaborator at work remains an important question for future research.

Data availability statement. All datasets and their associated documentation are freely available for download from the LISS and UAS data archives, after creating and signing a user contract.

Center for Economic and Social Research. (2025). *UAS 686: Water slogans*. Understanding America Study, University of Southern California. Link: https://uasdata.usc.edu/survey/UAS+686

Center for Economic and Social Research. (2024). *Understanding America Study Comprehensive File (Release 1224)*. Understanding America Study, University of Southern California. The latest comprehensive dataset is available at: https://uasdata.usc.edu/page/Comprehensive+File+And+Panel+Dataset

Centerdata. (2025). *Machines and Meaning: Work Meaningfulness in the Age of AI (L_MachineMeaning_1.0p)*. Tilburg University. https://www.dataarchive.lissdata.nl/study-units/view/1719

Centerdata. (2024). *Wave 16: LISS Core Study Personality (Project 7.16)*. LISS panel, Tilburg University. https://doi.org/10.57990/c63p-1250

Centerdata. (2023). *Wave 15: LISS Core Study Personality (Project 7.15)*. LISS panel, Tilburg University. https://doi.org/10.57990/rxs0-n703

Declaration of GenAI and AI-assisted technologies in the writing process. During the preparation of this work, the authors used Grammarly and ChatGPT version 5.5 to edit parts of the paper. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

Abel, M., & Johnson, R. (2025). *AI bias for creative writing: Subjective assessment versus willingness to pay* (IZA Discussion Paper No. 17646). IZA Institute of Labor Economics. https://docs.iza.org/dp17646.pdf

Acemoglu, D. (2025). The simple macroeconomics of AI. *Economic Policy*, *40*(121), 13-58. https://doi.org/10.1093/epolic/eiae042

Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics*, *40*(S1), S293-S340. https://doi.org/10.1086/718327

Alekseev, A., & Strobel, C. (2026). A taxonomy of AI experiments. *Journal of Behavioral and Experimental Economics, 121*, Article 102525. https://doi.org/10.1016/j.socec.2026.102525

Angrisani, M., Casanova, M., Fast, N. J., Narang, J., & Schroeder, J. (2026). Gaps in large language model awareness, usage, and perceptions in the United States: Evidence from a nationally representative longitudinal survey. *PNAS Nexus, 5*(2), Article pgag007. https://doi.org/10.1093/pnasnexus/pgag007

Ariely, D., Kamenica, E., & Prelec, D. (2008). Man's search for meaning: The case of Legos. *Journal of Economic Behavior & Organization, 67*(3-4), 671-677. https://doi.org/10.1016/j.jebo.2008.01.004

Bäker, A., & Mechtel, M. (2018). The role of task meaning on output in groups: Experimental evidence. *Managerial and Decision Economics, 39*(2), 131-141. https://doi.org/10.1002/mde.2875

Bellaiche, L., Shahi, R., Turpin, M. H., Ragnhildstveit, A., Sprockett, S., Barr, N., & Seli, P. (2023). Humans versus AI: Whether and why we prefer human-created compared to AI-created artwork. *Cognitive Research: Principles and Implications*, *8*, Article 42. https://doi.org/10.1186/s41235-023-00499-6

Bick, A., Blandin, A., Deming, D. J., Fuchs-Schündeln, N., & Jessen, J. (2026). *Mind the gap: AI adoption in Europe and the US* (NBER Working Paper No. 34995). National Bureau of Economic Research. https://www.nber.org/papers/w34995

Bonney, K., Breaux, C., Buffington, C., Dinlersoz, E., Foster, L., Goldschlag, N., Haltiwanger, J., Kroff, Z., & Savage, K. (2024). *Tracking firm use of AI in real time: A snapshot from the Business Trends and Outlook Survey* (NBER Working Paper No. 32319). National Bureau of Economic Research. https://www.nber.org/papers/w32319

Boussioux, L., Lane, J. N., Zhang, M., Jacimovic, V., & Lakhani, K. R. (2024). The crowdless future? GenAI and creative problem-solving. *Organization Science, 35*(5), 1589-1607. https://doi.org/10.1287/orsc.2023.18430

Brynjolfsson, E., Li, D., & Raymond, L. (2025). GenAI at Work. *The Quarterly Journal of Economics, 140*(2), 889-942. https://doi.org/10.1093/qje/qjae044

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What Can Machines Learn, and What Does It Mean for Occupations and the Economy? *AEA Papers and Proceedings, 108*, 43–47. https://doi.org/10.1257/pandp.20181019

Bryson, A., Kauhanen, A., & Rouvinen, P. (2026). *AI and worker well-being: Evidence from a nationally representative study* (IZA Discussion Paper No. 18540). IZA Institute of Labor Economics. https://www.iza.org/publications/dp/18540/ai-and-worker-well-being-evidence-from-a-nationally-representative-study

Cassar, L., & Meier, S. (2018). Nonmonetary Incentives and the Implications of Work as a Source of Meaning. *Journal of Economic Perspectives, 32*(3), 215-238. https://doi.org/10.1257/jep.32.3.215
Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-Dependent Algorithm Aversion. *Journal of Marketing Research*, *56*(5), 809–825. https://doi.org/10.1177/0022243719851788
Chandler, D., & Kapelner, A. (2013). Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior & Organization, 90*, 123-133. https://doi.org/10.1016/j.jebo.2013.03.003
Cnossen, F., & Nikolova, M. (2025). Work meaningfulness and effort. *Journal of Behavioral and Experimental Economics, 119*, Article 102460. https://doi.org/10.1016/j.socec.2025.102460
Comunale, M., & Manera, A. (2024). *The economic impacts and the regulation of AI: A review of the academic literature and policy actions* (IMF Working Paper No. 2024/065). International Monetary Fund. https://doi.org/10.5089/9798400271663.001
Cui, K. Z., Demirer, M., Jaffe, S., Musolff, L., Peng, S., & Salz, T. (2024). *The productivity effects of GenAI: Evidence from a field experiment with GitHub Copilot*. An MIT Exploration of Generative AI, March. https://doi.org/10.21428/e4baedd9.3ad85f1c
Das, M., & Knoef, M. (2019). Experimental and longitudinal data for scientific and policy research: Open access to data collected in the Longitudinal Internet Studies for the Social Sciences (LISS) panel. In N. Crato & P. Paruolo (Eds.), *Data-driven policy impact evaluation: How access to microdata is transforming policy design* (pp. 131–146). Springer. https://doi.org/10.1007/978-3-319-78461-8_9
De Neve, J.-E., Diener, E., Tay, L., & Xuereb, C. (2013). The objective benefits of subjective well-being. In J. F. Helliwell, R. Layard, & J. Sachs (Eds.), *World Happiness Report 2013* (pp. 54–79). Sustainable Development Solutions Network.
De Neve, J.-E., & Ward, G. W. (2017). Happiness at work. In J. F. Helliwell, R. Layard, & J. Sachs (Eds.), *World Happiness Report 2017* (pp. 144–177). Sustainable Development Solutions Network.
Del Rio-Chanona, R. M., Ernst, E., Merola, R., Samaan, D., & Teutloff, O. (2025). AI and jobs. A review of theory, estimates, and evidence. arXiv preprint arXiv:2509.15265.
Dell'Acqua, F., McFowland, E., Mollick, E., Lifshitz, H., Kellogg, K. C., Rajendran, S., . . . Lakhani, K. R. (2026). Navigating the jagged technological frontier: Field experimental evidence of the effects of Artificial Intelligence on knowledge worker productivity and quality. *Organization Science, 37*(2), 403-423. https://doi.org/10.1287/orsc.2025.21838
Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033
Draxler, F., Buschek, D., Tavast, M., Hämäläinen, P., Schmidt, A., Kulshrestha, J., & Welsch, R. (2023). *Gender, Age, and Technology Education Influence the Adoption and Appropriation of LLMs* (arXiv:2310.06556). arXiv. https://doi.org/10.48550/arXiv.2310.06556

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. *Science, 384*(6702), 1306-1308. https://doi.org/10.1126/science.adj0998

Felten, E. W., Raj, M., & Seamans, R. (2023). *Occupational heterogeneity in exposure to generative AI* (SSRN Scholarly Paper No. 4414065). Social Science Research Network. https://doi.org/10.2139/ssrn.4414065

Fernández-Macías, E., Bisello, M., Peruffo, E., & Rinaldi, R. (2023). Routinization of work processes, de-routinization of job structures. *Socio-Economic Review, 21*(3), 1773-1794. https://doi.org/10.1093/ser/mwac044

Fernández-Macías, E., Klenert, D., & Anton, J.-I. (2021). Not so disruptive yet? Characteristics, distribution and determinants of robots in Europe. *Structural Change and Economic Dynamics, 58*, 76-89. https://doi.org/10.1016/j.strueco.2021.03.010

Filippucci, F., P. Gal and M. Schief (2024), Miracle or myth? Assessing the macroeconomic productivity gains from Artificial Intelligence", *OECD Artificial Intelligence Papers*, No. 29, OECD Publishing, Paris, https://doi.org/10.1787/b524a072-en.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change, 114*, 254-280. https://doi.org/10.1016/j.techfore.2016.08.019

Georgieff, A., & Hyee, R. (2022). Artificial intelligence and employment: New cross-country evidence. *Frontiers in Artificial Intelligence*, *5*, Article 832736. https://doi.org/10.3389/frai.2022.832736

Giuntella, O., Konig, J., & Stella, L. (2025). Artificial intelligence and the wellbeing of workers. *Scientific Reports, 15*(1), Article 20087. https://doi.org/10.3389/frai.2022.832736

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*(1), 26–42. https://doi.org/10.1037/1040-3590.4.1.26

Grassini, S., Thorp, S. O., Ree, A. S., Sevic, A., & Cipriani, E. (2025). Distinct predictors of positive attitudes toward artificial intelligence and general technology: Big five traits, gender, and age. *Behaviour & Information Technology*. Advance online publication. https://doi.org/10.1080/0144929X.2025.2598623

Hartley, J., Jolevski, F., Melo, V., & Moore, B. (2024). *The labor market effects of generative artificial intelligence* (SSRN Scholarly Paper No. 5136877). Social Science Research Network. https://doi.org/10.2139/ssrn.5136877

Hendriks, M., & Cnossen, F. (2026). The economics of meaningful work: A scoping review. In M. Nikolova (Ed.), *Work Meaning and Motivation: Interdisciplinary Perspectives on Work Well-being* (pp. 151–184). Springer. https://doi.org/10.1007/978-3-032-23826-9_7

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Hötte, K., Somers, M., & Theodorakopoulos, A. (2023). Technology and jobs: A systematic literature review. *Technological Forecasting and Social Change, 194*, Article 122750. https://doi.org/10.1016/j.techfore.2023.122750

Humlum, A., & Vestergaard, E. (2024). *The Adoption of ChatGPT* (SSRN Scholarly Paper No. 4807516). Social Science Research Network. https://doi.org/10.2139/ssrn.4807516

Jones, C. I. (2026). *AI and our economic future* (NBER Working Paper No. 34779). National Bureau of Economic Research. https://www.nber.org/papers/w34779

Kapteyn, A., Angrisani, M., Darling, J., & Gutsche, T. (2024). The Understanding America Study (UAS). *BMJ open, 14*(10), Article e088183. https://doi.org/10.1136/bmjopen-2024-088183

Kesternich, I., Schumacher, H., Siflinger, B., & Schwarz, S. (2021). Money or meaning? Labor supply responses to work meaning of employed and unemployed individuals. *European Economic Review, 137*, Article 103786. https://doi.org/10.1016/j.euroecorev.2021.103786

Kosfeld, M., Neckermann, S., & Yang, X. (2017). The effects of financial and recognition incentives across work contexts: The role of meaning. *Economic Inquiry, 55*(1), 237-247. https://doi.org/10.1111/ecin.12350

Küsters, A., & Schneider, B. (2025). *What is technological unemployment?* (Oxford Economic and Social History Working Paper No. 218). University of Oxford. https://ora.ox.ac.uk/objects/uuid:839abc6e-1c7a-48f2-aff6-0fad792b44f9

Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., Capstick, E., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., Walsh, T., Hamrah, A., Santarlasci, L., … Oak, S. (2025). *The AI Index 2025 annual report*. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University. https://hai.stanford.edu/ai-index/2025-ai-index-report

Millet, K., Buehler, F., Du, G., & Kokkoris, M. D. (2023). Defending humankind: Anthropocentric bias in the appreciation of AI art. *Computers in Human Behavior*, *143*, Article 107707. https://doi.org/10.1016/j.chb.2023.107707

Mokyr, J., Vickers, C., & Ziebarth, N. L. (2015). The History of Technological Anxiety and the Future of Economic Growth: Is This Time Different? *Journal of Economic Perspectives, 29*(3), 31-50. https://doi.org/10.1257/jep.29.3.31

Mollick, E. (2024). *Co-intelligence: Living and working with AI*. Portfolio/Penguin.

Nikolova, M., & Angrisani, M. (2025). The impact of learning about AI advancements on trust. Technology in Society, 83, Article 102958. https://doi.org/10.1016/j.techsoc.2025.102958

Nikolova, M., & Cnossen, F. (2020). What makes work meaningful and why economists should care about it. *Labour economics, 65*, Article 101847. https://doi.org/10.1016/j.labeco.2020.101847

Nikolova, M., Cnossen, F., & Nikolaev, B. (2024). Robots, meaning, and self-determination. *Research Policy, 53*(5), Article 104987. https://doi.org/10.1016/j.respol.2024.104987

Nikolova, M., Lepinteur, A., & Cnossen, F. (2025). Just another cog in the machine? A worker-level view of robotization and tasks. *Economica*, *92*(368), 1101–1148. https://doi.org/10.1111/ecca.70006

Non, A., Rohde, I., de Grip, A., & Dohmen, T. (2022). Mission of the company, prosocial attitudes and job preferences: A discrete choice experiment. *Labour Economics*, 74, Article 102087. https://doi.org/10.1016/j.labeco.2021.102087

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187-192. https://doi.org/10.1126/science.adh2586

Otis, N., Clarke, R., Delecourt, S., Holtz, D., & Koning, R. (2024). *The uneven impact of generative AI on entrepreneurial performance* (SSRN Scholarly Paper No. 4671369). Social Science Research Network. https://doi.org/10.2139/ssrn.4671369

Otrachshenko, V., Nikolova, M., & Popova, O. (2023). Double-edged sword: Persistent effects of Communist regime affiliations on well-being and preferences. *Journal of Population Economics, 36*(3), 1139–1185. https://doi.org/10.1007/s00148-022-00930-0

Roldán-Monés, A. (2024). *When GenAI increases inequality: Evidence from a university debating competition* (EsadeEcPol Working Paper). Esade Center for Economic Policy. https://www.esade.edu/ecpol/wp-content/uploads/2019/09/2409-ChatGPTRoldan_ecpol.pdf

Sadeghian, S., & Hassenzahl, M. (2022). The "artificial" colleague: Evaluation of work satisfaction in collaboration with non-human coworkers. In *Proceedings of the 27th International Conference on Intelligent User Interfaces* (pp. 27–35). Association for Computing Machinery. https://doi.org/10.1145/3490099.3511128

Sadeghian, S., Uhde, A., & Hassenzahl, M. (2024). The soul of work: Evaluation of job meaningfulness and accountability in human-AI collaboration. *Proceedings of the ACM on Human-Computer Interaction, 8*(CSCW1), Article 38. https://doi.org/10.1145/3637407

Sajadieh, S., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Santarlasci, L., Pava, J., Maslej, N., Altman, R., Brynjolfsson, E., Brodley, C., Clark, J., Dignum, V., Kumar, V., Landay, J., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., … Weld, D. (2026). *The AI Index 2026 annual report*. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University. https://hai.stanford.edu/ai-index/2026-ai-index-report

Scherpenzeel, A. C. (2011). Data collection in a probability-based Internet panel: How the LISS panel was built and how it can be used. *Bulletin de Méthodologie Sociologique, 109*(1), 56–61. https://doi.org/10.1177/0759106310387713

Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*, *188*, Article 104405. https://doi.org/10.1016/j.obhdp.2025.104405

Stein, J.-P., Messingschlager, T., Gnambs, T., Hutmacher, F., & Appel, M. (2024). Attitudes towards AI: Measurement and associations with personality. *Scientific Reports*, *14*, Article 2909. https://doi.org/10.1038/s41598-024-53335-2

Stephany, F., & Duszynski, J. (2026). *Women Worry, Men Adopt: How Gendered Perceptions Shape the Use of GenAI* (arXiv:2601.03880). arXiv. https://doi.org/10.48550/arXiv.2601.03880

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward A Unified View. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Zhang, L., & Amos, C. (2026). Artificial intelligence and work meaningfulness. *Behaviour & Information Technology*. Advance online publication. https://doi.org/10.1080/0144929X.2026.2615039

Appendix A. Survey Questionnaires

English questionnaire

Introduction Text (Shown to All Respondents)

More than half of the US population drinks less than the recommended three 16-oz glasses of water per day (Source: Civicscience.com, 2023). Not drinking enough water can cause dehydration, which can lead to fatigue, reduced concentration, and negative effects on overall health.

In this survey, we ask you to help evaluate health campaign slogans such as “Drink water, feel better!” that aim to promote healthy water drinking habits. Your responses will help us understand how different slogans can encourage individuals to drink more water and improve their health.

Task Meaningfulness (Pre-Treatment)

Q1. How meaningful do you find the task of evaluating a slogan for a public health campaign promoting healthier water drinking habits? *Scale: 1 (Not meaningful at all) – 7 (Extremely meaningful)*

Experimental Manipulation

Respondents were shown the following instruction (source attribution varied by condition):

Please review the following slogans that encourage drinking more water. Each was created by **[an Artificial Intelligence (AI)-powered marketing software / a marketing professional]**. Please evaluate each slogan and rate it on creativity and persuasiveness.

• By creativity, we mean a slogan that is unique, original and innovative.
• By persuasiveness, we mean a slogan that is convincing and effective in encouraging drinking more water.

Slogan Evaluations

Q2. Slogan 1: “Don’t wait, hydrate!”

Q2a. To what extent do you find this slogan creative? *Scale: 1 (Not creative at all) – 5 (Very creative)*

Q2b. To what extent do you find this slogan persuasive? *Scale: 1 (Not persuasive at all) – 5 (Very persuasive)*

Q3. Slogan 2: "Pure and true, water is the best for you!"

Q3a. To what extent do you find this slogan creative? *Scale: 1 (Not creative at all) – 5 (Very creative)*

Q3b. To what extent do you find this slogan persuasive? *Scale: 1 (Not persuasive at all) – 5 (Very persuasive)*

Q4. Slogan 3: "Hydrate now, feel wow!"

Q4a. To what extent do you find this slogan creative? *Scale: 1 (Not creative at all) – 5 (Very creative)*

Q4b. To what extent do you find this slogan persuasive? *Scale: 1 (Not persuasive at all) – 5 (Very persuasive)*

<u>Task Meaningfulness (Post-Evaluation)</u>

Q5. How meaningful did you find the task of evaluating a slogan for a public health campaign promoting healthier water drinking habits? *Scale: 1 (Not meaningful at all) – 7 (Extremely meaningful)*

<u>Trust in Content Source</u>

After viewing the slogans, respondents rated their trust in the assigned content source.

Q6a. How much do you trust **[AI-powered marketing software / a marketing professional]** to write a creative slogan for a public health campaign? *Scale: 1 (Not at all) – 5 (Very much)*

Q6b. How much do you trust **[AI-powered marketing software / a marketing professional]** to write a persuasive slogan for a public health campaign? *Scale: 1 (Not at all) – 5 (Very much)*

<u>Open-Ended Response (=Effort)</u>

Q7. Do you have a slogan idea of your own? If so, please share it below. *(Open-ended response)*

Dutch questionnaire

Introduction Text (Shown to All Respondents)

Twee derde van de Nederlanders drinkt minder dan de aanbevolen 1,5 liter water per dag (Bron: Gezondheidsnet.nl, 2022). Niet genoeg water drinken kan leiden tot uitdroging, wat ons moe en minder gefocust kan maken en onze algehele gezondheid kan beïnvloeden.

In deze enquête vragen we u om gezondheidsvoorlichtingsslogans te beoordelen, zoals "Drink meer water!", die bedoeld zijn om gezond drinkgedrag te bevorderen. Uw antwoorden helpen ons te begrijpen hoe verschillende slogans mensen kunnen aanmoedigen om meer water te drinken en gezonder te leven.

Task Meaningfulness (Pre-Treatment)

Q1. Hoe zinvol vindt u de taak om een slogan te beoordelen voor een publieke gezondheidscampagne die het drinken van voldoende water aanmoedigt? *Schaal: 1 (Helemaal niet zinvol) – 7 (Heel erg zinvol)*

Experimental Manipulation

Respondenten kregen de volgende instructie te zien (bronvermelding varieerde per conditie):

Beoordeel de volgende slogans die mensen aanmoedigen om meer water te drinken. Elke slogan is gemaakt door **[een AI-gestuurd marketingprogramma / een marketingprofessional]**. Beoordeel elke slogan op creativiteit en overtuigingskracht.

• Met *creativiteit* bedoelen we dat een slogan uniek, origineel en innovatief is.

• Met *overtuigingskracht* bedoelen we dat een slogan aansprekend en effectief is in het aanmoedigen om meer water te drinken.

Slogan Evaluations

Q2. Slogan 1: "Hydrateer en presteer"

Q2a. In hoeverre vindt u deze slogan creatief? Schaal**:** 1 (Helemaal niet creatief) – 5 (Heel erg creatief)

Q2b. In hoeverre vindt u deze slogan overtuigend? *Schaal:* *1 (Helemaal niet overtuigend) – 5 (Heel erg overtuigend)*

Q3. Slogan 2: "Denk aan later, drink water"

Q3a. In hoeverre vindt u deze slogan creatief? *Schaal:* *1 (Helemaal niet creatief) – 5 (Heel erg creatief)*

Q3b. In hoeverre vindt u deze slogan overtuigend? *Schaal:* *1 (Helemaal niet overtuigend) – 5 (Heel erg overtuigend)*

Q4. Slogan 3: "Drink op tijd, blijf fit in je lijf"

Q4a. In hoeverre vindt u deze slogan creatief? *Schaal:* *1 (Helemaal niet creatief) – 5 (Heel erg creatief)*

Q4b. In hoeverre vindt u deze slogan overtuigend? *Schaal:* *1 (Helemaal niet overtuigend) – 5 (Heel erg overtuigend)*

Task Meaningfulness (Post-Evaluation)

Q5. Hoe zinvol vond u de taak om een slogan te beoordelen voor een publieke gezondheidscampagne die het drinken van voldoende water aanmoedigt? *Schaal:* *1 (Helemaal niet zinvol) – 7 (Heel erg zinvol)*

Trust in Content Source

Na het zien van de slogans beoordeelden respondenten hun vertrouwen in de toegewezen bron.

Q6a. Hoeveel vertrouwen heeft u in **[een AI-gestuurd marketingprogramma / een marketingprofessional]** om een creatieve slogan te schrijven voor een publieke gezondheidscampagne? *Schaal:* *1 (Helemaal geen vertrouwen) – 5 (Volledig vertrouwen)*

Q6b. Hoeveel vertrouwen heeft u in **[een AI-gestuurd marketingprogramma / een marketingprofessional]** om een overtuigende slogan te schrijven voor een publieke gezondheidscampagne? *Schaal:* *1 (Helemaal geen vertrouwen) – 5 (Volledig vertrouwen)*

Open-Ended Response (= Effort)

Q7. Hebt u zelf een idee voor een slogan? Zo ja, deel deze hieronder. (Open antwoord)

Appendix B. Supplementary Tables

Commentary related to Table B1.

The US and Dutch versions of the questionnaire differed slightly. In the Dutch survey, we introduced a second randomization that varied the order of the post-treatment meaning question and the two trust questions. Because the trust questions explicitly restate the attributed authorship of the slogans (see Appendix A), asking them before the post-treatment meaning question may reinforce the salience of the original AI attribution treatment. In the US survey, respondents always saw the post-treatment meaning question immediately after the treatment, followed by the trust questions and then the open-ended slogan generation task, our measure of effort. In the Dutch survey, half of the respondents saw the same sequence as in the US survey, while the other half saw the trust questions before the post-treatment meaning question. This second randomization allows us to examine whether restating the AI or human attribution in the trust questions affected subsequent responses by increasing the salience of the original treatment. In this subsection, we use this design feature to assess potential question-order effects.

Table B1 reports the split-sample estimates of the treatment effect separately by question order in the Dutch sample. Panel A presents estimates for respondents who saw the same question order as in the US survey, while Panel B presents estimates for respondents who received the opposite question order. The table also reports the difference between the treatment coefficients across the two subsamples, together with a *p*-value from a coefficient-difference test following Paternoster et al. (1998). This test assesses whether the estimated treatment effect differs statistically between the same-order and opposite-order subsamples.

The results in Table B1 provide no evidence that the treatment effect differs significantly by question order. For change in meaning, the estimated treatment effects remain statistically insignificant in both subsamples, and the coefficient-difference tests indicate no difference. For effort, the treatment estimates are negative in both subsamples, but statistically significant only in the opposite-order subsample. Descriptively, this suggests that the negative effort effect is more pronounced when the trust questions appear earlier in the survey, with the post-treatment meaning question placed between the trust questions and the effort measure. However, the formal equality of coefficients tests show that the two estimates are not statistically different from each other. Thus, the split-sample estimates do not provide formal evidence that question order influences the treatment effect on either change in meaning or effort.

Table B1: Dutch sample split-sample estimates, by question order

	Meaning Δ		Effort	
	(1)	(2)	(3)	(4)
Panel A: Sample split, *same* order				
Treatment	-0.044	-0.042	-0.033	-0.033
	(-0.082)	(-0.082)	(-0.027)	(-0.027)
Basic controls	Yes	Yes	Yes	Yes
Full controls	No	Yes	No	Yes
R-squared	0.003	0.009	0.006	0.021
Observations	1,058	1,057	1,058	1,057
Panel B: Sample split, *opposite* order				
Treatment	-0.096	-0.100	-0.060**	-0.053**
	(-0.080)	(-0.080)	(-0.026)	(-0.026)
Basic controls	Yes	Yes	Yes	Yes
Full controls	No	Yes	No	Yes
R-squared	0.006	0.007	0.008	0.021
Observations	1,059	1,059	1,059	1,059
Difference in treatment coefficients	0.052	0.058	0.028	0.021
p-value, coefficient equality	0.652	0.614	0.458	0.578

Notes: The table presents split-sample estimates of the effect of AI attribution on change in meaning and effort by question order in the Dutch sample. Panel A reports the estimates for respondents who received the same question order as in the US sample, while Panel B reports the estimates for respondents who received the opposite question order. Meaning change is measured as the difference between post-treatment and pre-treatment task meaning. Effort is a dummy variable equal to one if the respondent provided a valid slogan idea, and zero otherwise. "Difference in treatment coefficients" reports the difference between the treatment coefficient in the same-order subsample and the treatment coefficient in the opposite-order subsample. The corresponding *p*-value tests equality of the two treatment coefficients using the coefficient-difference test following Paternoster et al. (1998), which divides the difference between the coefficients by the square root of the sum of their squared standard errors. Basic controls (Columns (1) and (3)) include age and gender. Full controls (Columns (2) and (4)) additionally include education, employment status, marital status, and household income. Robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B2: Heterogeneity analysis by key sociodemographic characteristics, US sample

	Meaning Δ				Effort			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.137	-0.120	-0.130	-0.145	-0.018	-0.025	-0.004	0.004
	(0.118)	(0.126)	(0.141)	(0.140)	(0.036)	(0.037)	(0.039)	(0.069)
Treatment*Older	0.084				0.033			
	(0.170)				(0.057)			
Older	-0.062				0.082*			
	(0.129)				(0.043)			
Treatment*Male		0.034				0.040		
		(0.175)				(0.056)		
Male		0.066				0.018		
		(0.125)				(0.041)		
Treatment*Tertiary			0.057				-0.002	
			(0.173)				(0.056)	
Tertiary			-0.188				0.106***	
			(0.120)				(0.040)	
Treatment*Working				0.239				-0.002
				(0.252)				(0.057)
Working				-0.112				0.036
				(0.205)				(0.042)
R-squared	0.011	0.010	0.011	0.011	0.032	0.032	0.032	0.032
Observations	1,508	1,508	1,508	1,508	1,508	1,508	1,508	1,508

Notes: All specifications are estimated using OLS. The treatment indicator takes the value 1 if the participant evaluated slogans labeled as AI-generated and 0 if the slogans were labeled as human-generated. In Columns (1) – (4), the dependent variable is the change in scores between the post-treatment and pre-treatment evaluations of meaning; In columns (5) – (8), the dependent variable is a dummy taking the value of 1 if the respondent provided a slogan idea of their own and 0 otherwise. In columns (1) and (5), the moderator is a dummy variable taking the value of 1 if the respondent is above the age of 55 (median) and 0 otherwise; in columns (2) and (6) – gender dummy taking the value of 1 if the respondent is male and 0 if female or other; in columns (3) and (7) – dummy variable if the respondent has completed tertiary education (bachelor or more) with a diploma, and 0 otherwise; and in columns (4) and (8) – a dummy variable taking the value of 1 if the respondent is presently working and 0 otherwise. All specifications are estimated using survey weights. Robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B3: Heterogeneity analysis by key sociodemographic characteristics, Dutch sample

	Meaning Δ				Effort			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.011	-0.095	-0.064	-0.089	-0.069**	-0.037	-0.056**	-0.024
	(0.086)	(0.081)	(0.079)	(0.083)	(0.028)	(0.025)	(0.024)	(0.026)
Treatment*Older	-0.105				0.044			
	(0.115)				(0.037)			
Older	-0.099				-0.019			
	(0.085)				(0.029)			
Treatment*Male		0.054				-0.014		
		(0.115)				(0.038)		
Male		0.011				0.057**		
		(0.079)				(0.028)		
Treatment*Tertiary			-0.014				0.026	
			(0.115)				(0.038)	
Tertiary			-0.076				0.085***	
			(0.081)				(0.028)	
Treatment*Working				0.039				-0.040
				(0.115)				(0.037)
Working				-0.077				0.020
				(0.084)				(0.029)
R-squared	0.006	0.006	0.006	0.006	0.021	0.020	0.020	0.021
Observations	2,116	2,116	2,116	2,116	2,116	2,116	2,116	2,116

Notes: All specifications are estimated using OLS. The treatment indicator takes the value 1 if the participant evaluated slogans labeled as AI-generated and 0 if the slogans were labeled as human-generated. In Columns (1) – (4), the dependent variable is the change in scores between the post-treatment and pre-treatment evaluations of meaning; In columns (5) – (8), the dependent variable is a dummy taking the value of 1 if the respondent provided a slogan idea of their own and 0 otherwise. In columns (1) and (5), the moderator is a dummy variable taking the value of 1 if the respondent is above the age of 55 (median) and 0 otherwise; in columns (2) and (6) – gender dummy taking the value of 1 if the respondent is male and 0 if female or other; in columns (3) and (7) – dummy variable if the respondent has completed tertiary education (bachelor or more) with a diploma, and 0 otherwise; and in columns (4) and (8) – a dummy variable taking the value of 1 if the respondent is presently working and 0 otherwise. Robust standard errors are reported in parentheses. *** *p < 0.01,* ** *p < 0.05,* * *p < 0.10.*