CreativityBot: A Personalized AI Assistant to Support Creativity

Sean W. Kelley

D'Amore-McKim School of Business, Northeastern University, se.kelley@northeastern.edu

Christoph Riedl

D'Amore-McKim School of Business, Northeastern University, c.riedl@northeastern.edu

CCS CONCEPTS • Personalization • Human-AI Collaboration • Large Language Models

Additional Keywords and Phrases: Memory, Attention, Creativity, and Reasoning

1 INTRODUCTION

Generative AI is set to cause significant changes to the way that workers conceptualize and generate novel ideas. Already we are seeing significant gains in labor productivity [1-3], a reallocation of tasks towards evaluation [4], and enhanced rates of scientific discovery [5, 6]. Yet, despite AI's substantial social and economic benefits, achieving human-AI synergy has proven to be especially challenging; human-AI collaborative efforts often fall short of what the best of a human or AI can accomplish alone[7]. When applied to creative tasks, e.g., story writing, AI tools tend to have a homogenizing effect whereby they increase individual, but reduce overall, creativity[8, 9]. A partial explanation may lie in heterogeneity in user attributes and expertise, particularly with regards to how they use AI and how AI responds to their needs. Recent evidence showed that high-skilled workers tend to use AI differently than their lower skilled counterparts [10, 11]. How can AI systems be developed to maximize creative performance while simultaneously providing enhanced assistance? One way forward is to explicitly provide AI tools with relevant personalized information on their human co-workers ranging from demographic characteristics (e.g., age and gender) to psychological attributes (e.g., fluid and emotional intelligence) [12]. By having a comprehensive understanding of the co-worker's strengths and weakness, the AI tool can optimize its responses to enhance alignment with their needs and preferences. Emerging evidence suggests that misaligned AI systems lead to a poorer user experience [13] and fail to establish mutual understanding, a key component of effective collaboration [14]. In this study, we developed a generic ('baseGPT') and personalized ('creativeGPT') AI assistant to help participants develop high quality and creative marketing campaigns. Compared to generic AI tools, personalized AI has the potential to enhance creativity, support more efficient and effective collaboration, and provide a better user experience.

2 METHODS

2.1 Participants

We recruited 252 participants from Prolific Academic, an online worker platform. Participants were based in the U.S. with an approval rate on previous tasks of at least 99%. Participants were compensated with $11.25 for completing the approximately 45-minute study; participants who wrote marketing campaigns judged to be in the top 20% in terms of quality and creativity were eligible for a $2 bonus payment. We selected participants who reported having work experience in design or creative, marketing, and product or product management. Of the 252 participants recruited, 7 were excluded for failing 2 attention checks, 43 for using an AI tool other than the one provided, e.g., ChatGPT, 3 for not engaging with the AI-driven interview chatbot, 20 for engaging using the AI assistant, and 7 for completing the study in under 30 minutes.

After applying these exclusion criteria, a sample of 173 individuals was brought forward for analysis. Participants had a mean age of 38.1 years (SD: 12.3, range: 18-79), 50.8% were female, and a majority had at least a bachelor's degree (65.9%). Regarding industry specific expertise, 26.6% were reported to be marketing experts, 35.3% product experts, and 38.2% were neither marketing nor product experts.

2.2 Demographic, Psychological, and Cognitive Measures

We collected detailed demographic data (i.e., age, gender, highest educational level attained, current employment status, and occupation type) and individual psychological and cognitive traits. Specifically, we measured personality (Big Five Inventory, BFI-44)[15], current emotional state (PANAS-SF)[16], fluid intelligence (Raven's Progressive Matrices 13-item short form)[17], emotional intelligence (Perceiving AI Generated Emotions)[18], and learning goal orientation (Learning Goal Orientation 6-item subscale)[19]. Along with self-reported questionnaires participants completed two validated creativity tasks (Divergent Association Task [DAT] and Alternative Uses Task [AUT]) related to divergent thinking. The DAT measures a person's ability to generate diverse and unrelated words by asking participants to generate "10 words that are as different from each other as possible". While the AUT assesses how well a person can create a novel and imaginative use for two common household objects, e.g., an ice tray and a brick. For a complete description of task instructions and procedures, for both the DAT and AUT, see [20].

2.3 Experimental Design

After completing the psychological questionnaires and creativity tasks, participants engaged in an approximately 10 question long conversation with an AI-driven interview chatbot. The goal of this interview was to provide the personalized AI assistant with useful information on a person's professional experience and their work style to enhance collaboration on a marketing task. Specifically, the model was instructed to act as "an AI designed for personalized collaboration on marketing initiatives, aiming to understand your partner's skills and expertise to combine them effectively with your own capabilities". Next, participants were asked to develop a 300–400-word marketing campaign for a fictional lab-grown meat start-up. Each campaign was required to contain 5 distinct components: 1) marketing context & product positioning, 2) target audience and customer segments, 3) campaign messaging & trust building, 4) marketing channel & experience strategy, and 5) success metrics. Participants were also provided with an initial marketing campaign idea to use as either inspiration or a starting point. This task was designed to be sufficiently complex to elicit significant engagement with the AI assistant and require participants to consider how to market an atypical consumer product.

To complete this task, participants were provided with access to either a generic (baseGPT) or personalized (creativeGPT) AI assistant, both using GPT4o. The generic AI assistant was prompted to ask focused questions, build on a person's contributions, and provide constructive feedback matching their patterns of thinking. In contrast, the personalized AI assistant was given a detailed synthesis of a person's creativity – considering the influence of cognitive, psychological, and demographic traits – and a factual summary of the AI-guided collaborative interview (**Figure 1**). Together these two summaries provided the personalized AI tool with substantial knowledge on the user's creative profile, any prior marketing-relevant experience, and an understanding of how they usually collaborate. Both the generic and personalized AI assistants were given access to the company description and initial campaign idea to aid in completing the task. After completing the campaign, participants were asked to rate their experience (on a Likert scale from 1 to 5) with the AI assistant regarding its usefulness, task assistance, likelihood to recommend, trust, and confidence in AI to complete the task on its own. We also specifically asked about the types of higher-level thinking processes that the AI tool assisted

with, e.g., information synthesis, perspective taking etc., with questions adapted from [21]. Lastly, to gauge whether personalization was perceptible, participants provided a binary (yes/no) response to "Did the system seem tailored to your specific needs or preferences?".

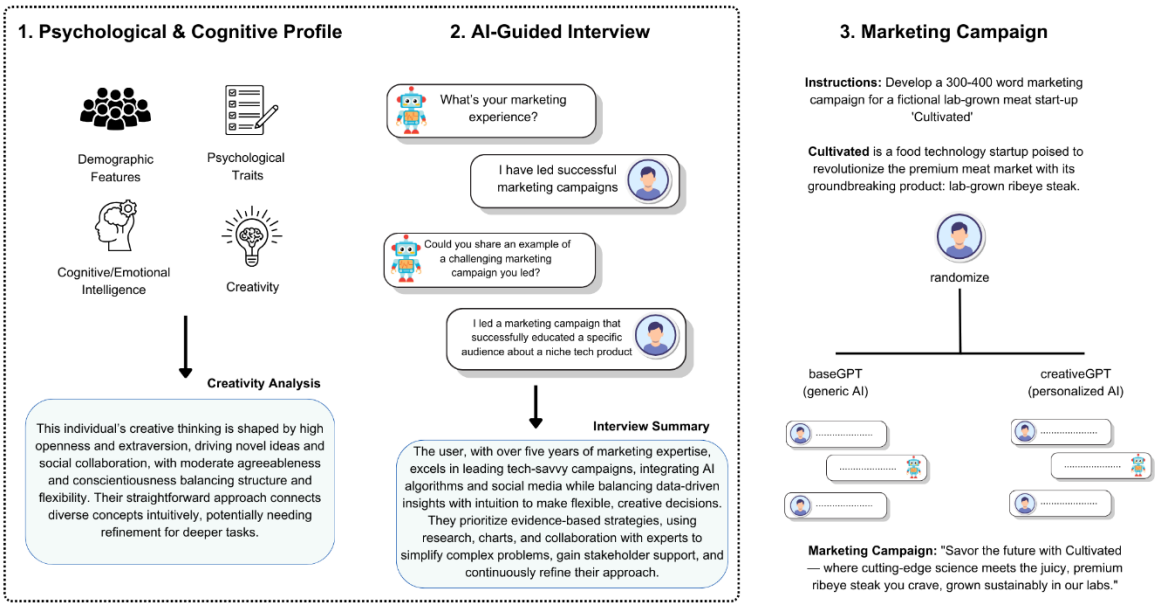

Figure 1: Implementation of AI Personalization and Marketing Task Experimental Design. CreativeGPT was personalized by providing it with a synthesis of a person's creativity, considering their cognitive and psychological traits, and a summary of an AI-guided interview on marketing experience and collaborative work style. After completing stages 1 and 2, participants wrote a marketing campaign for a fictional lab-grown meat start-up and were randomized to work with either baseGPT (generic AI) or creativeGPT (personalized AI).

2.4 Number of Marketing Manager Relevant Skills

In addition to self-reported marketing expertise, we sought to get a more quantitative estimate of a person's marketing experience. To accomplish this, we used O*NET's list of skills for a marketing manager and selected skills rated with an importance of at least 50/100; this resulted in a list of 21 job relevant skills, e.g., active listening and critical thinking. Next, we had GPT4o identify each skill that a person either explicitly mentions or could be reasonably inferred to possess from the AI-guided interview summary. Subsequently, we counted the total number of marketing-relevant skills per person. Participants possessed a mean of 8.9 skills (SD: 2.8, range: 3-15). Compared to people with self-reported marketing expertise, people with neither marketing nor product expertise (β = -1.6, SE = 0.5, p = 0.003) – but not product experts (β = 0.9, SE = 0.50, p = 0.10) – were identified to have significantly fewer marketing relevant skills.

2.5 Assessment of Marketing Campaign Quality and Creativity

We independently evaluated the quality of each marketing campaign using the same criteria described in 2.4. We then ran a general linear regression model to estimate the interaction between treatment conditions and number of marketing skills, controlling for the number of messages sent. While quality is an important dimension to measure, we also considered

the equally important effect of AI assistance on creativity. Here, we operationalized creativity as the proportion of unique concepts or themes within a campaign. For each campaign, we used GPT4o to extract distinctive elements – with a 1-3 description – with regards to market positioning, audience targeting, messaging and narrative, channel strategy, and psychological or emotional triggers. This process helps to ensure that only elements that are task relevant, i.e., marketing related, are identified as a 'true theme'. Every theme was then transformed into text embedding (3072x1 vector) with OpenAI's text-embedding-3-large model. We compared the cosine similarity of a focal theme to every other theme within the same treatment condition. A campaign theme was considered to be unique if the maximum cosine similarity to another theme was below a threshold of 0.75. For every participant, we then counted the proportion of unique campaign concepts and used this as a proxy metric for the campaign's overall creativity and distinctiveness. We repeated this process independently for campaign's produced via baseGPT, creativeGPT, and AI only. Lastly, we compared the proportion of unique campaign concepts, at the same percentile, across the 3 conditions. This comparison enabled us to determine whether a greater proportion of campaigns produced by using creativeGPT are in the top decile of creativity.

3 RESULTS

3.1 CreativeGPT Increases Both Marketing Campaign Quality and Creativity

Compared to people with access to baseGPT, participants who used creativeGPT produced marketing campaigns that were rated as significantly higher in quality (β= 1.95, SE = 0.70, p = 0.006) (**Figure 2a**). While participants who sent more messages to their AI assistant (β = 0.08, SE = 0.03, p = 0.003) and those who had more marketing-related skills (β = 0.21, SE = 0.05, $p < 0.001$) generated campaigns significantly higher in quality. Interestingly, there was a significant interaction effect between treatment and skills (β = -0.16, SE = 0.08, p = 0.04) such that the most skilled participants with access to creativeGPT produced campaigns that were slightly lower in quality. Notably, creativeGPT was particularly helpful for low skilled people – such that it substantially reduced the gap in performance relative to those with more skills. After establishing a significant net benefit to using a personalized AI assistant, we sought to assess creativeGPT's impact on creativity. The distribution of unique concepts for both baseGPT (D = 0.62, $p < 0.001$) and creativeGPT (D = 0.51, $p < 0.001$) was significantly different compared to AI only (**Figure 2b**). Although there was no overall significant difference in the distribution of unique campaign concepts between baseGPT and creativeGPT (D = 0.14, p = 0.23), people in the creativeGPT condition produced campaigns rated in the top 10% of creativity approximately twice as frequently as those in baseGPT (12.8% creativeGPT vs. 5.7% baseGPT). Meanwhile, the top 1% campaigns in the AI only condition would rank below the 50% percentile of creativity in baseGPT and creativeGPT, demonstrating substantial homogeneity among AI only generated campaigns.

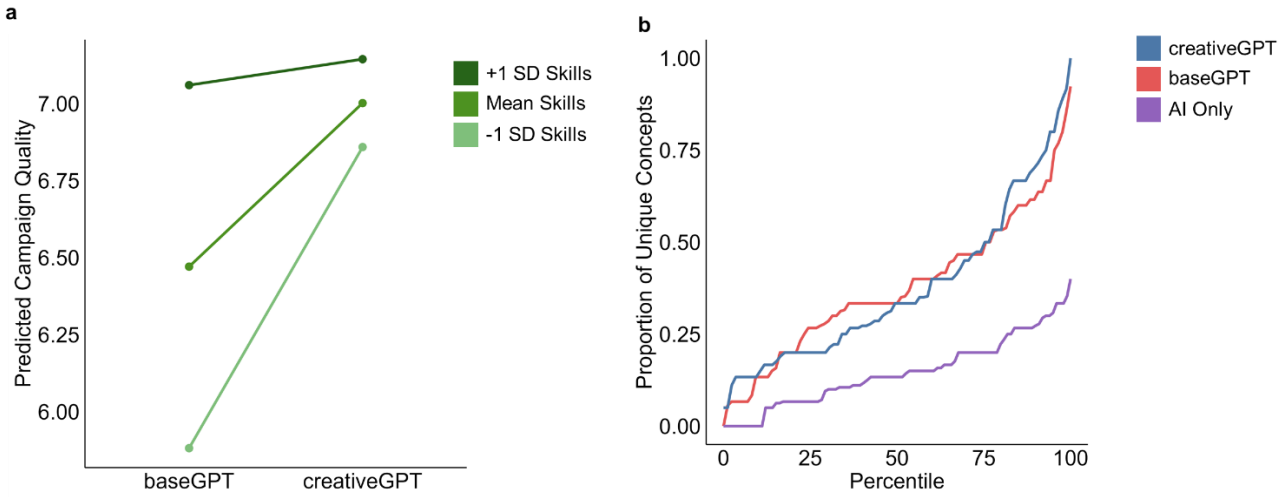

Figure 2: Marketing Campaign Quality and Creativity. A) Interaction between treatment conditions (baseGPT or creativeGPT) and number of marketing related skills (below, at, or above the average) controlling for number of messages sent to the AI assistant. There was a significant main effect of creativeGPT ($\beta = 1.9$, SE = 0.7, $p = 0.006$), number of skills ($\beta = 0.21$, SE = 0.05, $p < 0.001$), and an interaction between treatment condition and number of skills ($\beta = -0.16$, SE = 0.08, $p = 0.04$). B) Proportion of unique concepts identified in the marketing campaigns, i.e., a proxy for campaign creativity, within baseGPT, creativeGPT, and AI only.

3.2 Participants Self-Reported that CreativeGPT is More Useful, Provides Better Assistance, and Generally Supports Higher-Level Thinking Processes

Along with measures of campaign quality and creativity, we measured people's perception of and feedback on the AI assistants. Participants reported that creativeGPT provided significantly more assistance ($\beta = 0.46$, SE = 0.18, $p = 0.01$) and better feedback ($\beta = 0.43$, SE = 0.16, $p = 0.008$) on the campaign (**Figure 3a**). Not only were those participants more likely to recommend creativeGPT ($\beta = 0.54$, SE = 0.17, $p = 0.002$), but they also had more confidence ($\beta = 0.38$, SE = 0.12, $p = 0.002$) and trust ($\beta = 0.35$, SE = 0.17, $p = 0.04$) in the tool than people who used baseGPT. Along with holistic value perceptions, participants were asked to specifically comment on the type of cognitive assistance, e.g., perspective taking, that AI assistant provided them with (**Figure 3b**). Although there was no significant difference in the aggregate number of cognitive supports endorsed ($\beta = 0.38$, SE = 0.33, $p = 0.26$), people who used creativeGPT were significantly more likely to report they received help with idea brainstorming (84.9% creativeGPT vs. 69.0% baseGPT; $\chi^2 = 5.3$ (df =1), $p = 0.02$). While not a significant difference compared to baseGPT ($\chi^2 = 1.1$ (df = 1), $p = 0.28$), only 2.3% of people in the creativeGPT condition reported getting no cognitive assistance after using the tool.

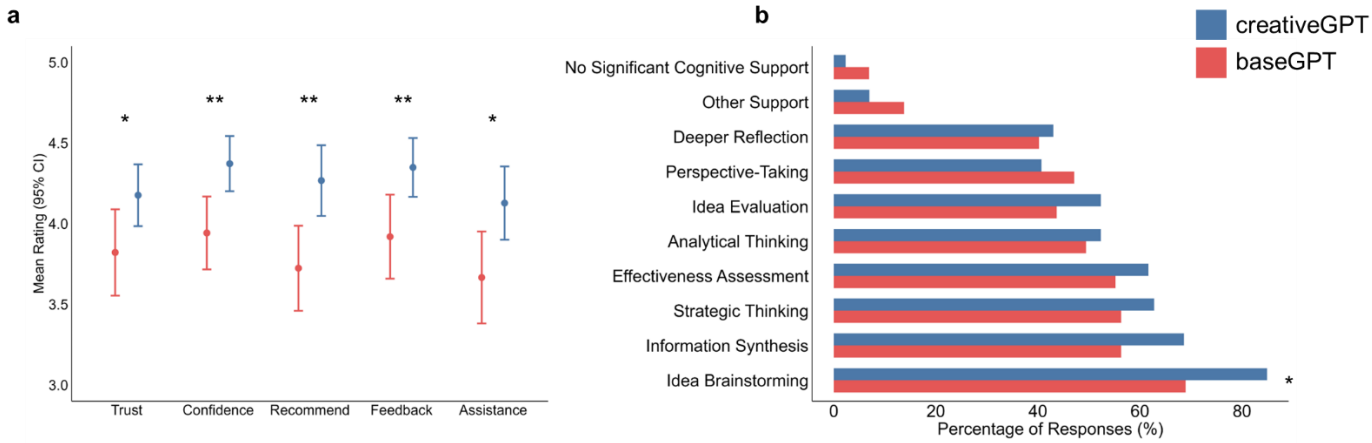

Figure 3: Self-Reported Feedback on Utility of an AI Assistant on Completing a Marketing Campaign (a) and Support with Higher Level Thinking Processes (b) by treatment condition (either baseGPT or creativeGPT). For each type of feedback, participants who used creativeGPT reported significantly higher ratings. CreativeGPT participants generally reported enhanced cognitive support across a wide variety of categories, although there was only a significant effect observed for brainstorming ($p < 0.05$). *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4 DISCUSSION

In this study, we presented a novel framework for personalizing AI assistance using theory-driven measures previously shown to be associated with creative performance and an AI-guided interview to optimize task collaboration. We found that a personalized AI assistant significantly increased the quality of marketing campaigns for a fictional start-up company relative to a more general support tool. Personalization was especially helpful for people with fewer marketing relevant skills; those participants who were already highly skilled did not gain much additional assistance from personalization. While both AI assistants offered broad cognitive support, users perceived the personalized version as particularly effective in augmenting higher-level thinking. Despite recent evidence that AI leads to homogenization, we found that people collaborating with both generic and personalized assistants produced more diverse sets of marketing ideas than those generated solely by AI. Participants working with personalized assistants created a higher proportion of campaigns ranked in the top decline of creativity. Our results show that providing AI systems with detailed information regarding their human co-worker leads to enhanced quality, creativity, and a better user experience. Personalized AI holds promise for enhancing human-AI collaboration by calibrating AI to be attuned to a person's strengths and weakness.

REFERENCES

1. Brynjolfsson, E., D. Li, and L.R. Raymond, *Generative AI at work*. 2023, National Bureau of Economic Research.
2. Dell'Acqua, F., et al., *Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality.* Harvard Business School Technology & Operations Mgt. Unit Working Paper, 2023(24-013).
3. Merali, A., *Scaling laws for economic productivity: Experimental evidence in llm-assisted translation.* arXiv preprint arXiv:2409.02391, 2024.
4. Toner-Rodgers, A., *Artificial Intelligence, Scientific Discovery, and Product Innovation.* 2024.
5. Sourati, J. and J.A. Evans, *Accelerating science with human-aware artificial intelligence.* Nature human behaviour, 2023. **7**(10): p. 1682-1696.

6. Si, C., D. Yang, and T. Hashimoto, *Can llms generate novel research ideas? a large-scale human study with 100+ nlp researchers.* arXiv preprint arXiv:2409.04109, 2024.
7. Vaccaro, M., A. Almaatouq, and T. Malone, *When combinations of humans and AI are useful: A systematic review and meta-analysis.* Nature Human Behaviour, 2024: p. 1-11.
8. Anderson, B.R., J.H. Shah, and M. Kreminski. *Evaluating Creativity Support Tools via Homogenization Analysis*. in *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 2024.
9. Doshi, A.R. and O.P. Hauser, *Generative AI enhances individual creativity but reduces the collective diversity of novel content.* Science Advances, 2024. **10**(28): p. eadn5290.
10. Riedl, C. and E. Bogert, *Effects of AI Feedback on Learning, the Skill Gap, and Intellectual Diversity.* arXiv preprint arXiv:2409.18660, 2024.
11. Otis, N., et al., *The uneven impact of generative AI on entrepreneurial performance.* Available at SSRN 4671369, 2023.
12. Matz, S., et al., *The potential of generative AI for personalized persuasion at scale.* Scientific Reports, 2024. **14**(1): p. 4692.
13. Palta, S., et al., *Speaking the Right Language: The Impact of Expertise Alignment in User-AI Interactions.* arXiv preprint arXiv:2502.18685, 2025.
14. Shaikh, O., et al., *Navigating Rifts in Human-LLM Grounding: Study and Benchmark.* arXiv preprint arXiv:2503.13975, 2025.
15. John, O.P., E.M. Donahue, and R.L. Kentle, *Big five inventory.* Journal of personality and social psychology, 1991.
16. Thompson, E.R., *Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS).* Journal of cross-cultural psychology, 2007. **38**(2): p. 227-242.
17. Raven, J., *Raven progressive matrices*, in *Handbook of nonverbal assessment*. 2003, Springer. p. 223-237.
18. Weidmann, B. and Y. Xu, *PAGE: A Modern Measure of Emotion Perception for Teamwork and Management Research.* arXiv preprint arXiv:2410.03704, 2024.
19. VandeWalle, D., *Development and validation of a work domain goal orientation instrument.* Educational and psychological measurement, 1997. **57**(6): p. 995-1015.
20. Olson, J.A., et al., *Naming unrelated words predicts creativity.* Proceedings of the National Academy of Sciences, 2021. **118**(25): p. e2022340118.
21. Lee, H.-P.H., et al., *The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers.* 2025.