When AI Says No: Investigating the Creative Power of Dissent

David Bouter
LIACS
Leiden University
d.r.bouter@umail.leidenuniv.nl

Rob Saunders
LIACS
Leiden University
r.saunders@liacs.leidenuniv.nl

Abstract

In recent years, co-creative AI has emerged as a novel approach to facilitate ideation tasks. Contemporary AI systems tend to exhibit conforming behavior. For example, well known chat bots such as ChatGPT will almost exclusively provide affirmative feedback on ideas, regardless of their quality. In human collaboration, having dissenting ideation partners often enhances creativity by fostering divergent thinking and preventing groupthink. This study examines whether the same applies to AI, investigating whether interaction with a dissenting co-creative AI leads to more creative story-writing outcomes and how it affects adoptability. Participants engaged in a co-creative story-writing task, after which expert evaluators assessed the creativity of the written stories using the consensual assessment technique. The likelihood of adoption was examined using the Unified Theory of Acceptance and Use of Technology (UTAUT). The results indicate that, on average, participants produced more creative output when collaborating with a dissenting AI than a conforming AI. However, the difference did not reach statistical significance, rendering the results inconclusive. The UTAUT analysis revealed a slight variation in adoption tendencies between the two AI conditions. These variations were minor, however, and did not provide significant evidence that adoption likelihood differed for the two conditions. These findings suggest that, in co-creative contexts, dissenting AI systems are equally adoptable as conforming AI systems. However, further research is necessary to determine whether dissenting behavior in AI meaningfully enhances creative performance.

Introduction

In the field of computational creativity, the goal is to design programs that can enhance human creativity without necessarily being creative themselves (Jordanous 2014). Presently it has become increasingly common to engage in co-creative ideation with AI systems (Cox, Djernæs, and van Berkel 2025). Especially when working with large language models, it becomes possible to have 'brainstorming sessions.'

A key issue with AI systems is that they contain biases (Schwartz et al. 2022). Some of these biases, such as groupthink and confirmation biases, can cause AIs to be overly conforming, which can hinder ideation processes when working with AIs. This paper poses the research question: "In the context of co-creative story-writing, does interaction with a dissenting co-creative AI lead to more creative outcomes compared to a conforming AI, and how does interacting with a dissenting AI impact its adoptability?"

This study hypothesizes that interaction with a dissenting co-creative AI in story-writing leads to more creative outcomes than interaction with a conforming AI. This hypothesis was developed by first examining human-human interactions (HHI), and then analyzing how these concepts translate to human-computer interaction (HCI). Existing literature indicates that in HHI, dissent stimulates creative ideation

Related Work

Divergent Thinking Versus Groupthink

Divergent thinking is widely recognized as a key component of creative ideation, often serving as a reliable indicator of novel ideas, though it does not equate to creativity directly (Runco 2010). A common strategy to attempt to reach creative ideas is brainstorming, however this form of group-based ideation has been shown to suppress divergent thoughts rather than enhance them, which is detrimental for the creative process (Runco 2010). Runco found a negative linear relationship between group size and idea quality, suggesting that individual ideation is often more effective. This reduction in quality is attributed to groupthink—the tendency to conform to group norms and avoid proposing unconventional ideas (Janis 1972). Janis characterizes groupthink as a barrier to optimal decision-making, particularly in cohesive groups under pressure or guided by authoritative leaders. To counter this, Nemeth (1995) proposes dissent as a mechanism for stimulating divergent thought. Her studies show that minority disagreement encourages group members to explore multiple perspectives, breaking the conformity that stifles creativity.

Social Behavior Towards Computer Systems

Co-creative AI systems typically display conforming behavior, often reinforcing rather than challenging user input (Schwartz et al. 2022). If users respond to excessive conformity in co-creative systems in the same socially driven ways as they do to other humans, we should expect a detrimental effect on idea quality (Runco 2010). Reeves and Nass (1996)

argue that humans have a tendency to project social behavior on computer systems. In their experimental work, they demonstrate that users often apply the same social norms to computers as to people. They found that participants rated a computer's performance more favorably when giving feedback on the same machine they had interacted with, compared to when evaluating it from a separate terminal, or on pen and paper. The effect was statistically significant, with users reporting the computer as more friendly and competent. These findings suggest that people apply at least some of the same social filters—such as politeness and conformity—in their interactions with computers.

As Reeves and Nass conclude, "When in doubt, treat it as human." If humans default to social behavior even when interacting with computer systems, co-creative AI may inherit the same group-level dynamics—such as groupthink—that hinder creativity in human-computer ideation.

Evaluation of Creativity and Adoption

A persistent challenge in creativity research is the ambiguity of the term creativity itself. Although most individuals possess an intuitive sense of what constitutes a creative product, this perception is inherently subjective (Amabile 1982). Such subjectivity complicates the collection of empirical data and necessitates structured evaluation methods. For this study, the Consensual Assessment Technique (CAT) (Amabile 1982) was adopted as a practical and widely accepted approach for assessing creative output.

CAT evaluates creativity through expert judgments within the relevant domain. It rests on two key assumptions: that domain experts are capable of recognizing creativity, and that creativity exists along a continuum—allowing for comparative assessments between products. Rather than enforcing strict criteria, the method leverages the consensus of informed evaluators to determine which products are relatively more creative.

Adoption and acceptance play an important role in the development of new information technologies (Venkatesh et al. 2003). The choice to measure the adoption likelihood of dissenting AI was made due to the antagonistic nature of a dissenting AI. If dissenting AI proves to be an effective tool for ideation, users' willingness to engage with the system decides its practical value. Should the system be perceived as excessively disagreeable or present other adoption-related barriers, its effectiveness may be rendered moot by low user uptake.

To measure adoption and acceptance, the unified theory of acceptance and use of technology (UTAUT), was used (Venkatesh et al. 2003). This model is a unification of eight competing acceptance models that describe user acceptance. The model uses determinants to make a statement about expected usage behavior. To measure whether dissenting AI and conforming AI are equally adoptable or not, Venkatesh et al. (2003) developed a standardized questionnaire that was used during this study.

The UTAUT model evaluates technology acceptance through eight determinants: *performance expectancy*, *effort expectancy*, *attitude toward using technology*, *social influence*, *self-efficacy*, *anxiety*, *behavioral intention to use the system*, and *facilitating conditions*. These constructs capture users' expectations, perceived ease of use, social and emotional influences, and enabling factors. Although UTAUT typically incorporates moderator variables such as age, gender, or experience to enhance predictive power, these were excluded due to the limited sample size and resulting lack of statistical power.

Methodology

To test the hypothesis as outlined in the introduction, an experiment was conducted with 22 participants who engaged in co-creative tasks. Specifically, participants completed a short story-writing exercise and received feedback from a large language model on their output. For this study, the LLaMA 3.1 model was chosen because it was practical to implement and, through initial testing, was found to be sufficiently capable of expressing both conforming and dissenting responses when appropriately prompted. While there is no formal benchmark for dissent or conformity expression in large language models, models like LLaMA 3.1 are trained on diverse texts and have demonstrated flexibility in generating outputs aligned with various stances. Prior work has shown that language models can take on different viewpoints when prompted appropriately (Bang et al. 2024).

The LLaMA 3.1 model is a pre-trained AI model and was used to simulate both conforming and dissenting feedback styles. The literature suggests that large language models naturally lean toward conformity (Schwartz et al. 2022), prompting dissenting behavior required careful prompt engineering. Prompts were designed to produce actionable feedback within 10 minutes, stay within 10 sentences, and conceal the AI's instructed attitude. The dissenting version was developed first using phrases like "devil's advocate" and then adapted into a conforming version by replacing attitude cues. Due to a lack of scientific standards for prompting dissenting AI, a trial-and-error approach was used to refine prompts while aiming to minimize bias, though some bias remains unavoidable.

The experiment was conducted in two phases. In the first phase, all participants were presented with the same short story—*Casabianca* by James Baldwin—with the ending deliberately omitted, and were instructed to generate three distinct story endings. These initial responses were subsequently evaluated by an AI system, which provided individualized feedback for each of the participant's story endings. In the second phase, participants were asked to produce three additional story endings for the same story. They were given the option to revise or build upon their earlier responses, incorporating the AI-generated feedback. Furthermore, they were instructed to apply the feedback in any way they liked, they were given the goal to: "be as creative as you can." Final outputs were evaluated by an expert panel based on their creativity.

The panel comprised individuals with professional or academic expertise in Literary Studies (MA), Linguistics (MA), and Performing Arts, including experience in performance, directing, and playwriting. While story writing was selected as the medium for this study, the underlying principles of the experimental design are applicable to other forms of creative

ideation (Runco 2010). At the end of the experiment, participants completed a survey with questions in accordance with the UTAUT model. All questions were answered on a scale from 'Strongly agree' (1) to 'Strongly disagree' (5).

Results

Three experts each produced a list with all stories written by the participants, ranked from most creative to least creative. Based on those lists, a 4th list was produced based on the average ranking of each story. A cropped representation of these lists are represented in Table 1. On average, each expert rated stories written with dissenting AI partners as more creative than those written with conforming AI partners. This can be seen in Table 1 where stories written with dissenting AI partners cluster at the higher rankings than those that were written with conforming AI partners. This observation is supported by the statistical test results in Table 2, which indicate that although only one expert graded the work from the dissenting group (group D) as significantly more creative than that of the conforming AI group (group C), there was general agreement among experts that group D's output was, on average, more creative. This trend is further reflected in the combined average rankings of all experts, which yielded a p-value of 0.111. While this value does not achieve statistical significance, it may suggest a potential relationship between interacting with dissenting AI and enhanced creative output.

After completing the grading process, the experts were asked to reflect on their evaluations through a questionnaire. The responses indicate that while technical quality and aesthetic appeal had some influence on their perception of creativity, this effect was not particularly strong. However, experts note that the variability in technical quality among the submissions made it challenging to fully disregard these factors in their assessments. On average, the experts characterized the overall range of creativity in the submissions as "somewhat varied." This degree of variation is beneficial, as it facilitates differentiation between submissions, ranking stories with highly similar levels of creativity would have been a more complex task (Amabile 1982).

To validate the CAT, an analysis was performed to grade the expert agreement. Figure 1 demonstrates the degree of agreement of all experts. Each grading of each story is compared to the average ratings of the experts combined. The degree of agreement is also expressed in Table 3 using the coefficient of concordance, W (Kendall 1945). A W of 0.741 is classified as *strong agreement* by the heuristic posed by Siegel and Castellan (1988). A caveat of these data is that each data point is compared to the average, in which the data point itself is contained, favoring the clustering of the data.

To provide context for this data, a correlation coefficient was produced for each pair of experts. These coefficients can be found in Table 3. According to the heuristic proposed by Schober, Boer, and Schwarte (2018), the level of agreement between expert 1 and 2 ($\tau = 0.580$) could be classified as *moderate agreement*. Both Expert 1 and Expert 2 appear to have comparatively lower agreement with Expert 3, a finding further supported by the Kendall's Tau coefficients in Table 3. The degree of agreement between

Table 1: Expert ranking

R	E1	E2	E3	Avg	R	E1	E2	E3	Avg
1	D	C	D	D	...				
2	D	C	D	D	56	C	C	D	C
3	C	D	D	D^1	57	D	D	D	C
4	C	C	D	C^1	58	C	D	D	D
5	D	D	C	D^1	59	C	C	D	D
6	D	D	D	D^2	61	D	C	D	C
7	D	D	C	C^2	62	C	C	C	C
8	C	D	C	C	63	C	C	C	C
9	D	C	C	D	64	C	D	C	C
10	C	C	D	D	65	C	D	D	D
...					66	C	C	D	D

Note. Gray cells represent a dissenting AI partner. White cells represent a conforming AI partner. Cells with a superscript represent a tie in a ranking with other cells with the same superscript.

Table 2: Mann-Whitney U test: Expert ratings

Expert	p-value
Expert 1	0.041
Expert 2	0.254
Expert 3	0.404
Average lower bound	0.103
Average upper bound	0.120
Average mean	0.111

Note. For all tests, the alternative hypothesis specifies that group C < group D. N = 22.
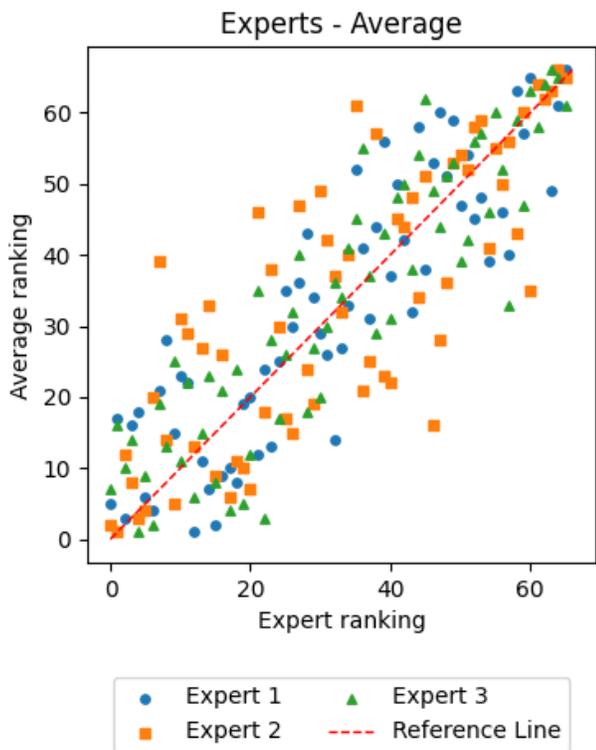
Figure 1: Rankings of each expert compared to the average rankings of all experts.

Experts 1 and 2 with Expert 3 ($\tau = 0.343$ and $\tau = 0.372$) could be categorized as *weak agreement*, following the same guidelines. This degree of agreement is slightly lower than expected, implying there might be some variation in how the experts ranked the stories. This is further discussed in the discussion section.

UTAUT

Table 4 presents the p-values of the Mann-Whitney U test, which examines whether responses differed significantly between participants in group D and group C ($\alpha \leq 0.050$). The results indicate significant differences for 'Effort Expectancy', while 'Anxiety' shows a notable trend (p = 0.079), albeit not reaching statistical significance. The mean responses for these determinants provide further insight. Participants in the dissenting AI group reported a mean response of 2.364 (SD = 0.848) for 'Effort Expectancy', compared to 1.900 (SD = 0.447) in the conforming AI group. This suggests that participants in the conforming AI group found the system somewhat easier to use, potentially indicating a higher willingness to adopt it.

'Anxiety' was rated higher in the dissenting AI group with a mean of 3.667 (SD = 1.155) than in the conforming AI group where the mean was 4.238 (SD = 1.221). This suggests that participants interacting with the dissenting AI reported somewhat more anxiety, which may neg-

Table 3: Kendall's Tau Correlations: Expert agreement

	Kendall's Tau
Expert 1 - Expert 2	0.580
Expert 1 - Expert 3	0.343
Expert 2 - Expert 3	0.372
Kendall's Tau: $0 \leq \tau < 0.40$ indicates weak agreement, $0.40 \leq \tau < 0.70$ indicates moderate agreement	
	Kendall's W
Average	0.741
Kendall's W: $0.70 \leq W \leq 1.00$ indicates strong agreement	

atively impact adoption in the dissenting AI group. For all other determinants, mean responses were similar across both groups, indicating minimal variation in perceived adoptability. Overall, this indicates that there is no significant difference in adoptability depending on the AI group a participant was in.

Table 4: Mann-Whitney U test: UTAUT

Category	p
Facilitating conditions	0.930
Performance expectancy	0.794
Social influence	0.783
Self-efficacy	0.350
Behavioral intention to use the system	0.297
Attitude towards technology	0.258
Anxiety	0.079
Effort expectancy	0.050

Note. For the Mann-Whitney U test, the alternative hypothesis specifies that group C $\neq$ group D. N = 22

Discussion

Dissenting AI exhibits the potential to influence co-creative ideation processes involving AI systems. Preliminary findings suggest a possible association between dissenting AI behavior and enhanced human creativity. Furthermore, concerns regarding user adoption do not appear to present significant barriers to the practical application of dissenting AI in ideation settings. While the observed effects did not reach statistical significance, the results offer promising indications that warrant further research. The lack of statistical significance could be due to several limitations in the study.

One notable limitation concerns the brevity of the AI interaction period. Multiple participants indicated that the interaction duration was insufficient for them to meaningfully engage with the dissenting or conforming AI. It is plausible

that extending the interaction—particularly through multi-turn dialogue—could yield more pronounced effects. Additionally, in the absence of established scientific standards for prompt engineering in this context, the design of AI prompts was based on iterative trial-and-error. This approach may have introduced researcher bias and, more critically, may have constrained the AI's capacity to effectively express dissent.

Several practical limitations may also have influenced the study's outcomes. These include a limited sample size, non-representative participant demographics consisting of mostly students and professionals with a background in creative disciplines, and the absence of double-blind experimental conditions. Finally, certain constraints were present in the application of the CAT. Due to resource limitations, the CAT could not be implemented to its full potential, which may partially account for the lower levels of inter-rater agreement observed.

Conclusion and future work

This study explored the following research question: In the context of co-creative story-writing, does interaction with a dissenting co-creative AI lead to more creative outcomes compared to a conforming AI, and how does interacting with a dissenting AI impact its adoptability?

The results indicate that differences in system adoption are primarily reflected in the constructs of anxiety and effort expectancy, both of which slightly favor the conforming AI system. However, these effects were not statistically strong enough to conclude a significant difference in overall adoptability between the two AI types. Notably, both AI systems were evaluated positively across most constructs of the UTAUT model, indicating a generally favorable disposition toward adoption.

Regarding creativity outcomes, no statistically significant evidence was found to support the hypothesis that interaction with dissenting AI produces more creative output than interaction with conforming AI. Nonetheless, the observed trend suggests that participants collaborating with the dissenting AI may generate more creative ideas. These findings highlight the need for future research. Future studies employing more robust experimental designs that would not be subjected to the aforementioned limitations, may validate this trend and provide empirical support for the integration of dissenting AI in co-creative ideation processes.

Author Contributions

Acknowledgements

References

Amabile, T. M. 1982. Social psychology of creativity: A consensual assessment technique. *Journal of personality and social psychology* 43(5):997.

Bang, Y.; Chen, D.; Lee, N.; and Fung, P. 2024. Measuring political bias in large language models: What is said and how it is said. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 11142–11159. Association for Computational Linguistics.

Cox, S. R.; Djernæs, H. B.; and van Berkel, N. 2025. Beyond productivity: Rethinking the impact of creativity support tools. In *Proceedings of the ACM Creativity and Cognition (C&C '25)*, 1–15. ACM.

Janis, I. 1972. Victims of groupthink; a psychological study of foreign-policy decisions and fiascoes. *Journal of American History* 60(3):857–858.

Jordanous, A. 2014. What is computational creativity?

Kendall, M. G. 1945. The treatment of ties in ranking problems. *Biometrika* 33(3):239–251.

Nemeth, C. J. 1995. Dissent as driving cognition, attitudes, and judgments. *Social Cognition* 13(3):273–291.

Reeves, B., and Nass, C. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People 10(10)*. Cambridge University Press.

Runco, M. A. 2010. *Divergent Thinking, Creativity, and Ideation*. Cambridge University Press.

Schober, P.; Boer, C.; and Schwarte, L. A. 2018. Correlation coefficients: appropriate use and interpretation. *Anesthesia & analgesia* 126(5):1763–1768.

Schwartz, R.; Vassilev, A.; Greene, K.; Perine, L.; Burt, A.; and Hall, P. 2022. *Towards a standard for identifying and managing bias in artificial intelligence*, volume 3. US Department of Commerce, National Institute of Standards and Technology.

Siegel, S., and Castellan, N. J. 1988. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, 2nd edition.

Venkatesh, V.; Morris, M. G.; Davis, G. B.; and Davis, F. D. 2003. User acceptance of information technology: Toward a unified view. *MIS quarterly* 425–478.