From Proactive Assistants to Agentic Debate : Human-in-the-Loop Multi-Agent AI for Co-Creative Design

Oussama BADDI* and Hafid GARHOUM*

**ENSA Berrechid, Université Hassan 1er, Morocco*
Email : {oussama.baddi, mr.hafid.garhoum}@gmail.com

Résumé

Most contemporary AI design tools operate as single-agent systems that react to user prompts through linear execution pipelines. While effective for automation, such systems fail to capture the inherently deliberative, conflict-driven, and iterative nature of human creative practice. This paper introduces an agentic multi-agent architecture for human-AI co-creative design, in which creativity emerges through structured debate among autonomous AI agents, governed by human-in-the-loop validation.

The proposed system orchestrates a team of specialized design agents—artist, critic, UX researcher, and brand strategist—engaged in multi-round argumentative exchanges moderated by an orchestrator agent. Rather than suppressing disagreement, the system externalizes design tensions as explicit debates, enabling traceable reasoning, improved explainability, and richer creative outcomes. Human users retain final authority, validating or redirecting agent consensus at each debate cycle.

Through conceptual evaluation and comparative analysis with monolithic AI systems, we argue that agentic debate offers a robust paradigm for co-creative systems, transforming AI from a passive assistant into a collective reasoning partner.

Index Terms

Human-AI collaboration, multi-agent systems, co-creative AI, agentic debate, human-in-the-loop, design intelligence.

I. Introduction

Artificial intelligence has rapidly permeated creative domains such as design, art, and ideation. Despite impressive generative capabilities, most AI tools remain fundamentally limited by a single-agent, prompt-response interaction model. In these systems, creativity is treated as an execution problem rather than a deliberative process.

Human creative practice, by contrast, is inherently dialogical : ideas are proposed, critiqued, refined, and negotiated across multiple perspectives. Design teams rely on disagreement, trade-offs, and debate to surface constraints and explore alternative solutions. Monolithic AI systems obscure this process by collapsing reasoning into latent model states.

Recent research in human-AI co-creation emphasizes the importance of treating AI as a creative partner rather than an automation tool [1], [4]. However, existing systems rarely operationalize this partnership at the architectural level.

This paper proposes a shift from adaptive single-agent AI to *agentic collective intelligence*, where creativity emerges through structured debate among specialized AI agents, under explicit human governance.

Our contributions are :

— A novel agentic debate architecture for co-creative design
— A structured multi-round debate mechanism enabling explainable creative reasoning
— A human-in-the-loop governance model ensuring bounded autonomy

II. Related Work

A. Human-AI Co-Creation

Prior work highlights AI as a complementary creative force capable of expanding design spaces [1]. Studies show increased idea diversity and user satisfaction when AI participates in creative tasks [3].

B. Paradoxes of Human-AI Collaboration

Designing co-creative systems introduces tensions such as ambiguity vs. precision and control vs. serendipity [2]. Most systems resolve these tensions implicitly through automation, limiting transparency.

C. Multi-Agent Creative Systems

Recent advances in LLM-based multi-agent systems demonstrate improved reasoning through role specialization and interaction [5], [6]. However, their application to human-governed creative debate remains underexplored.

III. Agentic Multi-Agent Debate Architecture

A. System Overview

The proposed architecture conceptualizes AI creativity as an emergent property of debate among autonomous agents. Figure 1 illustrates the system.

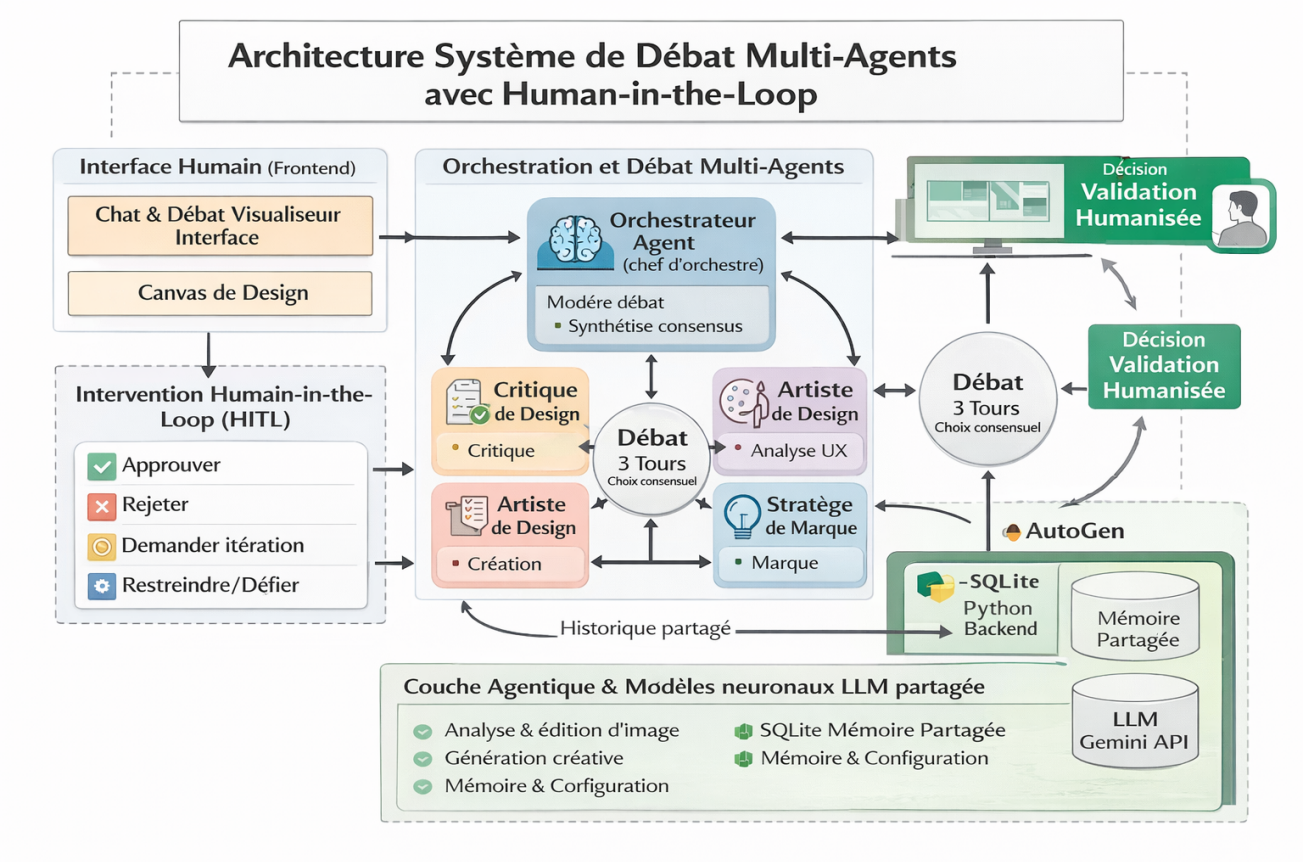

FIGURE 1. Agentic multi-agent debate architecture with human-in-the-loop validation.

B. Agent Roles

The system comprises the following agents :

— **Design Artist Agent** : Generates creative proposals.
— **Design Critic Agent** : Evaluates proposals against formal design principles.
— **UX Research Agent** : Assesses usability, accessibility, and user impact.
— **Brand Strategist Agent** : Ensures alignment with brand identity and values.
— **Orchestrator Agent** : Moderates debate, enforces structure, and synthesizes outcomes.

Each agent operates autonomously while sharing a common language model backend, preserving epistemic diversity.

IV. Structured Debate Mechanism

Design decisions are produced through a three-round debate process :

1) **Proposal Round** : The Design Artist introduces multiple concepts.
2) **Argumentation Round** : Other agents critique, defend, and challenge proposals.
3) **Consensus Round** : The Orchestrator synthesizes a reasoned consensus.

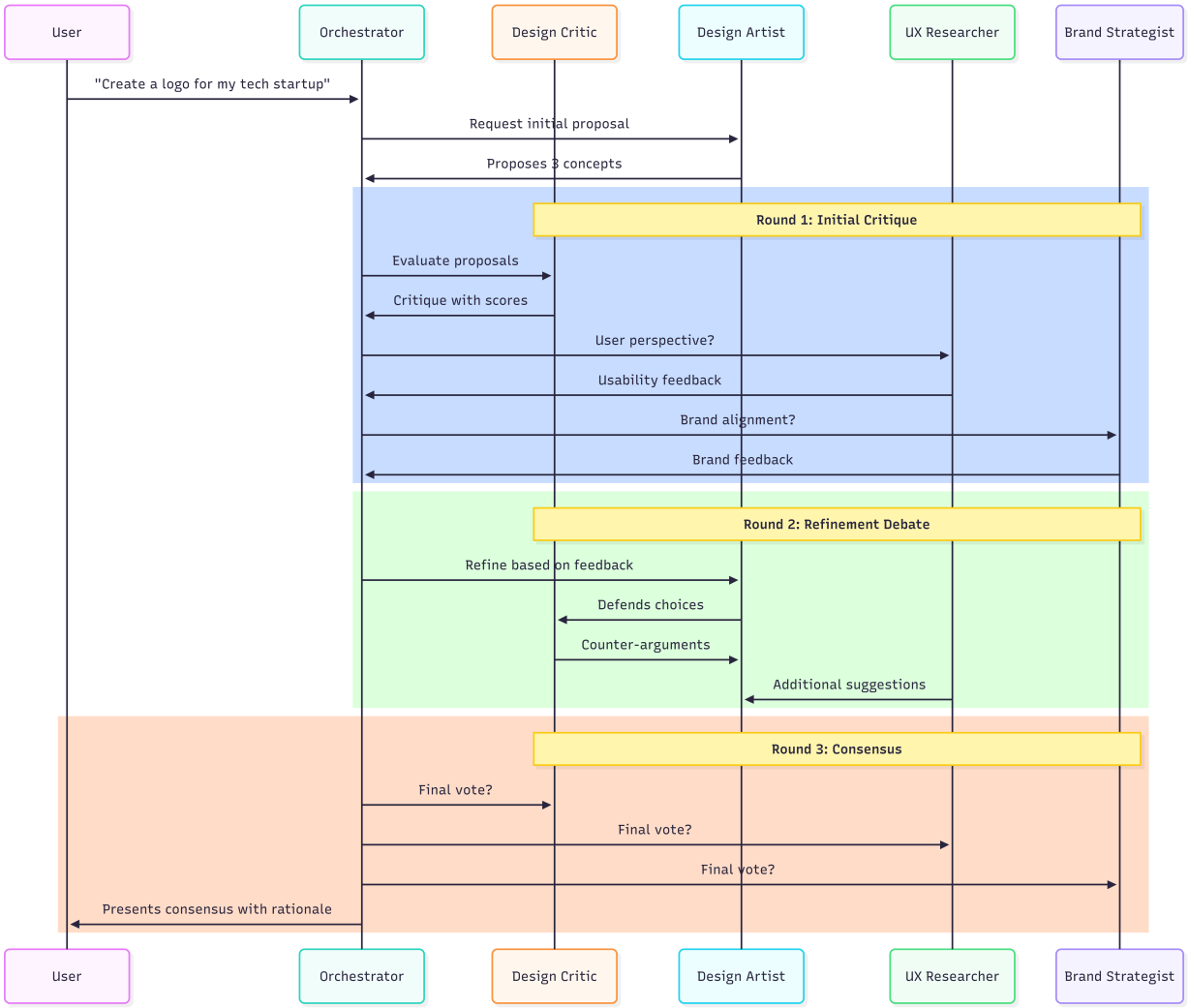

FIGURE 2. Multi-round agentic debate workflow.

This explicit debate externalizes reasoning, making creative decisions inspectable and traceable.

V. Human-in-the-Loop Governance

Human users retain final authority at the end of each debate cycle. They may :

— Approve the agent consensus
— Reject and restart the debate
— Introduce new constraints
— Request additional debate rounds

This governance model ensures ethical control, accountability, and alignment with human intent, addressing concerns around autonomous creative systems.

VI. Discussion

Unlike single-agent AI systems, the proposed architecture treats disagreement as a productive force. By externalizing design tensions through debate, the system enhances explainability and supports reflective creativity.

This approach aligns with emerging views of AI as a participant in collective intelligence rather than an isolated problem-solver [7].

VII. Limitations and Future Work

Current limitations include :

— Evaluation remains conceptual and qualitative
— Focus is limited to design-centric tasks
— Debate strategies are rule-based rather than learned

Future work will explore learning debate strategies, longitudinal user studies, and extension to other creative domains.

VIII. Conclusion

This paper introduces an agentic debate-driven framework for human-AI co-creative design. By shifting from monolithic AI execution to structured multi-agent deliberation with human oversight, the system offers a transparent, explainable, and collaborative model of creativity.

We argue that agentic debate represents a foundational paradigm for next-generation co-creative AI systems.

Références

[1] J. Smith et al., “Human-AI Co-Creation : A Framework for Collaborative Design,” arXiv, 2024.
[2] M. Chen et al., “Designing Co-Creative Systems : Five Paradoxes in Human–AI Collaboration,” MDPI, 2023.
[3] L. Wang et al., “Human-AI Collaboration in Creative Design : Evaluating Cognitive Synergy,” arXiv, 2023.
[4] P. Xu et al., “Beyond Automation : How Designers Perceive AI as a Creative Partner,” ACM CHI, 2024.
[5] S. Ding et al., “DesignGPT : Multi-Agent Collaboration in Design,” CAD&CG, 2024.
[6] Y.-C. Lin et al., “Creativity in LLM-based Multi-Agent Systems : A Survey,” arXiv, 2024.
[7] H. Li et al., “A Design Trajectory Map of Human-AI Collaborative Methods,” arXiv, 2023.