A Multi-Agent Approach for Iterative Refinement in Visual Content Generation

Achala Nayak Adithya S Kolavi Nigel Teofilo Dias Srinidhi Somayaji P Ramamoorthy Srinath

INTRODUCTION

Foundational image generation models like Stable Diffusion produce high-quality images from text prompts but lack control beyond the initial input. This limitation is critical for industries such as advertising, where precise text alignment, layout customization, and brand consistency are essential. Manual creation of visual content (e.g., posters, banners) is time-consuming, repetitive, and often leads to inconsistent outputs, reducing user engagement and brand value. Our system addresses these challenges by enabling iterative refinement and human-in-the-loop editing, making it ideal for creating customizable, consistent visual content.

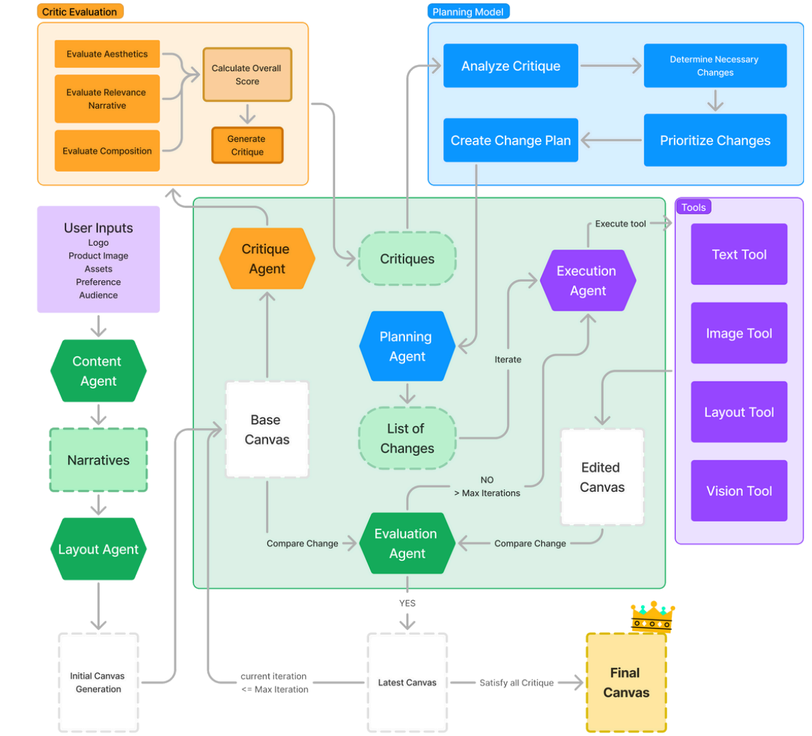
KEY CONTRIBUTIONS

- **Multi-Agent Architecture:** Integrates narrative generation and visual analysis through specialized agents, enabling seamless collaboration between AI and human designers.
- **Iterative Refinement Loop:** Combines LLMs and VLMs to progressively improve visual content, ensuring brand consistency and design coherence.
- **Efficient Implementation:** Runs on a single T4 GPU, demonstrating real-time interaction and reduced content creation time.

PROPOSED SOLUTION

Our multi-agent system generates an initial image from text and image-based prompts, followed by iterative refinement to address visual and semantic flaws. The process includes:

- **Initial Image Generation:** Uses LLMs for narrative creation, VLMs for analysis, and diffusion models for image synthesis.
- **Refinement Loop:** Employs Critic, Planning, Execution, and Evaluation agents to iteratively improve the design.
- **Human-in-the-Loop:** Allows users to edit any component via an interactive editor at each refinement stage.



SYSTEM ARCHITECTURE

- **Frontend**: Next.js and Fabric.js for a Figma/Canva-like editor with real-time interaction via REST APIs and SSE.
- **Backend**:
  - Control Server (FastAPI): Orchestrates agents and manages sessions.
  - Model Inference Server (T4 GPU): Hosts LLMs, VLMs, and diffusion models.
  - Database (serverless PostgreSQL): Stores canvas states and refinement history.
- **Agents**:
  - Critic: Identifies flaws.
  - Planning: Prioritizes changes.
  - Execution: Selects tools (e.g., Text, Image, Layout).
  - Evaluation: Validates changes via an "eval and apply" loop.
  - See Figure 1 (center) for the system architecture.

EXPERIMENTAL RESULTS

- Case Studies: Generated advertisements for "Mango Masti," "Crispo," and "ABC University," showing improved text alignment, layout, and brand consistency through iterative refinement.
- **Performance:**
  - **Generation time:** Minutes vs. days for human workflows.
  - **Quality:** Enhanced text-image alignment, brand integrity, and narrative coherence.
- **Benefits:** Faster iteration, better control, and professional outputs.

CONCLUSION

- **Summary**: Our system enhances image generation with iterative refinement and human-in-the-loop editing, ideal for advertising.
- **Future Work**: Improve aesthetic evaluation, reduce latency, and expand to other content types (e.g., videos).