

## 1. Tổng quan báo cáo

Báo cáo này cung cấp một phân tích toàn diện về "BannerAgency" - một hệ thống agent đa phương thức (multimodal LLM agent) nhằm tự động hóa quy trình thiết kế banner quảng cáo. Trong bối cảnh các công cụ AI tạo sinh hiện tại như DALL-E và OpenCOLE chủ yếu tạo ra các sản phẩm dạng điểm ảnh (pixel-based) tĩnh, thiếu khả năng chỉnh sửa sâu thì BannerAgency đề xuất một giải pháp đột phá. Hệ thống này mô phỏng một đội ngũ sáng tạo chuyên nghiệp bằng cách sử dụng một nhóm các agent AI chuyên biệt: Strategist, Background Designer, Foreground Designer, và Developer. Các agent này cộng tác với nhau phân tích yêu cầu, tạo hình ảnh nền, thiết kế bố cục tiền cảnh, và xuất ra sản phẩm dưới dạng mã nguồn có thể chỉnh sửa hoàn toàn (SVG hoặc plugin Figma). Để đánh giá hiệu quả, tác giả dùng bộ dữ liệu benchmark đa phương thức mới mang tên BannerRequest400. Các kết quả thực nghiệm cho thấy BannerAgency vượt trội đáng kể so với các phương pháp đối chứng trên hàng loạt các tiêu chí về chất lượng thiết kế, từ tính phù hợp với đối tượng mục tiêu đến tính thẩm mỹ tổng thể.

## 2. Sơ lược BannerAgency

### 2.1. BannerAgency là gì?

Khi thiết kế một banner quảng cáo cho sản phẩm mới thì thay vì phải thuê một đội ngũ marketing hoặc tự mình mày mò trên các phần mềm phức tạp, chỉ cần sử dụng BannerAgency. Về cơ bản, **BannerAgency hoạt động như một "agency quảng cáo thu nhỏ" bên trong máy tính.**

### 2.2. Tại sao BannerAgency Đặc biệt?

Điểm khác biệt lớn nhất của BannerAgency là nó không chỉ "vẽ" ra một bức ảnh ngẫu nhiên.

1. **Đầu ra có thể chỉnh sửa:** Thay vì đưa một file ảnh JPG hay PNG "chết", BannerAgency cung cấp một file thiết kế dạng SVG hoặc file cho phần mềm Figma. Điều này có nghĩa là có thể dễ dàng mở file lên và thay đổi mọi thứ: sửa một lỗi chính tả, di chuyển logo sang góc khác, đổi màu nút bấm...
2. **Tư duy như Marketer:** Hệ thống này được xây dựng để cân nhắc các nguyên tắc quảng cáo, sự hài hòa về thị giác và thông điệp thuyết phục.
3. **Con người có thể can thiệp:** Con người có thể tham gia vào quá trình bất cứ lúc nào. Đưa ra phản hồi, yêu cầu thay đổi, hoặc định hướng cho các agent AI, giống như đang quản lý đội ngũ của mình.

### 2.3. Các Phát hiện Chính

- BannerAgency tạo ra các banner có chất lượng thẩm mỹ, tính nhất quán thương hiệu và hiệu quả quảng cáo cao hơn các phương pháp khác.

- Người dùng thực tế yêu thích các thiết kế của BannerAgency hơn hẳn so với các mô hình mạnh như GPT-4V.
- Việc phân chia công việc cho các agent chuyên biệt mang lại kết quả tốt hơn nhiều so với việc dùng một agent "biết tuốt" duy nhất.

### 3. Tổng quan Hệ thống (System Overview)

#### 3.1. Mô phỏng Quy trình Sáng tạo của Con người

Cốt lõi của BannerAgency là ý tưởng mô phỏng quy trình làm việc cộng tác của một đội ngũ thiết kế chuyên nghiệp. Trong thực tế, việc tạo ra một sản phẩm sáng tạo không phải là công việc của một cá nhân mà là kết quả của sự phối hợp giữa nhiều vai trò khác nhau: từ người lên chiến lược, người viết nội dung, đến người thực thi thiết kế. BannerAgency tái tạo mô hình này trong một môi trường kỹ thuật số.

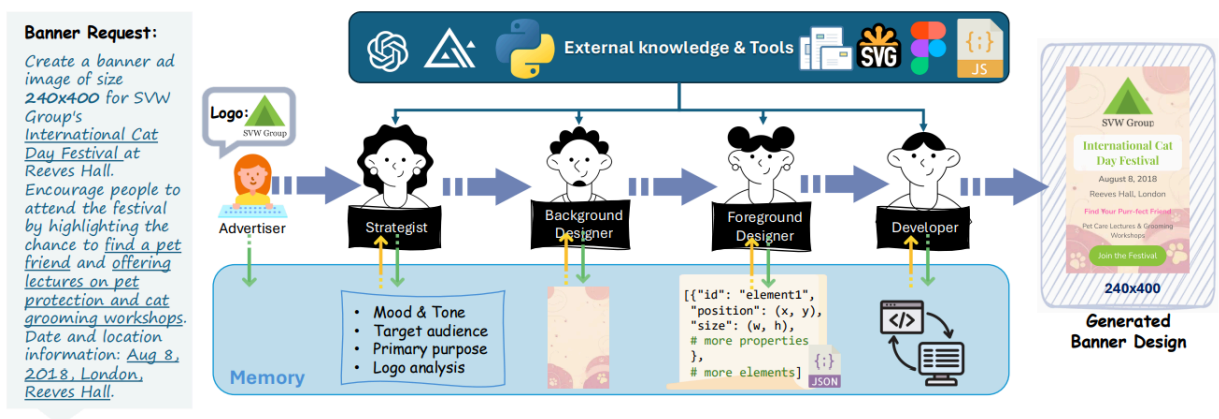
#### 3.2. Kiến trúc Đa Agent Chuyên biệt

Hệ thống được xây dựng dựa trên kiến trúc đa agent (multi-agent system), trong đó mỗi agent là một MLLM được giao một vai trò và bộ kỹ năng cụ thể. Bốn agent chính trong hệ thống bao gồm:

1. **Strategist:** Phân tích yêu cầu đầu vào và xác định các mục tiêu chiến lược.
2. **Background Designer:** Chịu trách nhiệm tạo ra hình ảnh nền phù hợp.
3. **Foreground Designer:** Thiết kế bố cục và nội dung cho các yếu tố tiền cảnh.
4. **Developer:** Chuyển đổi bản thiết kế (blueprint) thành mã nguồn cuối cùng.

#### 3.3. Luồng công việc Cộng tác và Tương tác

Luồng công việc của BannerAgency diễn ra một cách tuần tự nhưng có tính tương tác cao.



- **Đầu vào:** Nhà quảng cáo cung cấp yêu cầu dưới dạng văn bản (brief) và hình ảnh logo.

- **Giai đoạn 1 (Strategist):** The Strategist phân tích brief và logo để xác định mood & tone, đối tượng mục tiêu, và mục đích chính của chiến dịch. Thông tin này được lưu vào một "bộ nhớ chung" (shared memory).
- **Giai đoạn 2 (Background Designer):** Dựa trên thông tin trong bộ nhớ, The Background Designer tạo ra một hình nền phù hợp, đảm bảo không chứa text gây nhiễu.
- **Giai đoạn 3 (Foreground Designer):** The Foreground Designer nhận nền và các thông tin chiến lược, sau đó tạo ra một "bản thiết kế chi tiết" (blueprint) dạng JSON, định nghĩa vị trí, kiểu dáng, nội dung cho text, logo, nút CTA, và các yếu tố trang trí.
- **Giai đoạn 4 (Developer):** The Developer đọc blueprint và render nó thành file SVG hoặc mã plugin Figma.

Toàn bộ quá trình này được hỗ trợ bởi khả năng truy cập kiến thức bên ngoài, gọi các công cụ (tool-calling), và chia sẻ thông tin qua bộ nhớ chung, giúp các quyết định được đưa ra một cách hài hòa và có ngữ cảnh.

## 4. Kiến trúc Chi tiết (System Architecture)

Kiến trúc của BannerAgency là một hệ thống được thiết kế sao cho mỗi agent không chỉ thực hiện nhiệm vụ riêng lẻ mà còn tương tác chặt chẽ với nhau thông qua một bộ nhớ chung và các công cụ chuyên dụng.

### 4.1. The Strategist (Nhà chiến lược)

Đây là agent đóng vai trò giao diện chính với nhà quảng cáo, là điểm khởi đầu của toàn bộ quy trình sáng tạo.

- **Vai trò:** Tiếp nhận và phân tích các yêu cầu đầu vào, bao gồm brief chiến dịch và các tài sản thương hiệu như logo. Từ đó, agent này xác định các yếu tố cốt lõi của banner:
  1. **Mood & Tone:** Cảm xúc chủ đạo (ví dụ: vui tươi, sang trọng, chuyên nghiệp).
  2. **Target Audience:** Đối tượng khách hàng mục tiêu (ví dụ: trẻ em, người đi làm).
  3. **Primary Purpose:** Mục đích chính (ví dụ: tăng nhận diện thương hiệu, thúc đẩy doanh số).
- **Quy trình hoạt động:**
  1. Nhận brief và file logo.
  2. Phân tích logo để hiểu về màu sắc, phong cách thương hiệu. Một bước xử lý quan trọng là tự động cắt bỏ các khoảng trống suốt dư thừa (transparent padding) xung quanh logo để tối ưu hóa không gian sử dụng.
  3. Tổng hợp tất cả các thông tin chiến lược đã phân tích.
- **Công cụ hỗ trợ:** Các công cụ xử lý hình ảnh cơ bản để phân tích và tinh chỉnh logo.
- **Đầu ra:** Một bộ dữ liệu có cấu trúc chứa toàn bộ định hướng chiến lược. Dữ liệu này được lưu vào bộ nhớ chung và chuyển quyền điều khiển cho Background Designer.

### 4.2. The Background Designer (Nhà thiết kế Nền)

Agent này chịu trách nhiệm tạo ra background cho banner

- **Vai trò:** Tạo ra một hình nền (background) hấp dẫn, phù hợp với định hướng chiến lược từ The Strategist và đặc điểm của logo, đồng thời đảm bảo hình nền này **không chứa văn bản (text-free)**.
- **Quy trình hoạt động:** Được triển khai dưới dạng một ReAct agent, có khả năng suy luận và hành động.
  1. Đọc thông tin từ bộ nhớ (yêu cầu, đặc điểm logo, mục tiêu chiến dịch).
  2. Sử dụng công cụ FindImagePath để kiểm tra xem nhà quảng cáo có cung cấp sẵn hình nền hay không.
  3. Nếu không, nó sẽ tự tạo prompt và gọi công cụ T2I (một mô hình Text-to-Image như FLUX.1-schnell) để sinh ra hình ảnh mới.
  4. **Vòng lặp tự tinh chỉnh (self-refinement loop):** Sau khi tạo ảnh, agent dùng công cụ TextChecker (dựa trên một MLLM) để kiểm tra xem ảnh có vô tình chứa chữ hay không. Nếu có, nó sẽ tự động sửa lại prompt (ví dụ: thêm các từ khóa tiêu cực như "no text", "text-free") và tạo lại ảnh. Quá trình này lặp lại tối đa 5 lần để đảm bảo kết quả là một tấm nền "sạch".
  5. Agent này cũng xử lý thông minh về kích thước, chọn tỷ lệ khung hình gần nhất được T2I tool hỗ trợ nhưng lớn hơn kích thước yêu cầu, sau đó thu nhỏ lại để đảm bảo độ phân giải cao và kích thước chính xác.
- **Đầu ra:** Một file ảnh nền chất lượng cao, không chứa text, sẵn sàng cho giai đoạn thiết kế foreground.

### 4.3. The Foreground Designer (Nhà thiết kế Tiền cảnh)

Đây là nơi xây dựng các component/element cho banner dựa trên dữ liệu từ Strategist

- **Vai trò:** Tạo ra một bản thiết kế (blueprint) có cấu trúc dạng JSON, định nghĩa chi tiết mọi thuộc tính của các thành phần tiền cảnh: logo, văn bản quảng cáo, nút CTA, và các hình khối trang trí.
- **Quy trình hoạt động:**
  1. Nhận hình nền từ Background Designer và các thông tin chiến lược từ bộ nhớ.
  2. Xác định nội dung, font chữ, màu sắc, kích thước cho các đoạn text.
  3. Quyết định vị trí và kích thước của logo.
  4. Thiết kế nút CTA (nội dung, màu sắc, hình dạng).
  5. Sắp xếp bố cục của tất cả các yếu tố. Để đảm bảo sự căn chỉnh và chồng chéo hợp lý, agent này sử dụng **kỹ thuật định vị tương đối (relative positioning)**, trong đó vị trí của một yếu tố được xác định dựa trên một yếu tố tham chiếu khác [232].
- **Đầu ra:** Một file JSON blueprint. Việc biến các quyết định thiết kế thành dữ liệu có cấu trúc thay vì điểm ảnh phẳng là nền tảng cho khả năng chỉnh sửa toàn diện sau này.

### 4.4. The Developer (Nhà phát triển)

Agent cuối cùng trong chuỗi, đóng vai trò chuyển đổi ý tưởng thành sản phẩm kỹ thuật số có thể sử dụng được.

- **Vai trò:** Đọc file JSON blueprint từ Foreground Designer và biên dịch nó thành mã nguồn thực thi.
- **Hai phương thức triển khai:**
  1. **Tạo mã SVG (SVG Code Generation):** SVG (Scalable Vector Graphics) là một định dạng XML mở, được hỗ trợ rộng rãi trên web và các phần mềm thiết kế. Agent sẽ tạo ra một file SVG chứa tất cả các yếu tố đồ họa. Ưu điểm là tính tương thích cao, nhưng nhược điểm là cần một bước import thủ công vào phần mềm chỉnh sửa.
  2. **Tạo mã Plugin Figma (Figma Plugin Code Generation):** Figma là một công cụ thiết kế đám mây rất phổ biến. Agent này sẽ tạo ra mã JavaScript tuân thủ API của Figma. Khi chạy plugin này trong Figma, nó sẽ tự động vẽ banner ngay trên canvas. Ưu điểm là quy trình liên mạch, tự động và tận dụng được hệ thống component mạnh mẽ của Figma, nhưng việc tạo mã JavaScript phức tạp hơn cho LLM.
- **Đầu ra:** File mã nguồn .svg hoặc một thư mục plugin Figma hoàn chỉnh.

#### **4.5. Cơ chế Tinh chỉnh Lặp lại Tăng cường bởi Bộ nhớ (Memory-Augmented Iterative Refinement)**

Đây là một trong những cơ chế tinh vi và mạnh mẽ nhất của BannerAgency, mô phỏng quá trình nhận phản hồi và cải tiến sản phẩm trong thế giới thực.



Một đóng góp quan trọng của nghiên cứu này là **BannerRequest400**, bộ dữ liệu benchmark đầu tiên được thiết kế chuyên biệt để đánh giá các hệ thống tạo banner quảng cáo đa phương thức.

### 5.1. Nhu cầu cho một Benchmark Mới

Các bộ dữ liệu hiện có, như DESIGNERINTENTION, thường chỉ chứa các ý định thiết kế dưới dạng văn bản đơn thuần (uni-modal) và không giải quyết các yêu cầu đặc thù của banner quảng cáo, vốn luôn gắn liền với một yếu tố thị giác cốt lõi là logo thương hiệu. Sự thiếu hụt này gây khó khăn cho việc đánh giá và so sánh một cách công bằng và toàn diện các mô hình tạo banner.

### 5.2. Quy trình Xây dựng Bộ dữ liệu

Quá trình tạo ra BannerRequest400 rất công phu và có hệ thống:

1. **Khởi tạo ý định:** Các tác giả tuyển chọn 100 ý định quảng cáo (advertising intentions) từ bộ dữ liệu DESIGNERINTENTION.
2. **Tổng hợp logo:** Với mỗi ý định, họ sử dụng mô hình Claude 3.5 Sonnet để tổng hợp một logo vector tương ứng. Các logo này sau đó được các chuyên gia đánh giá và tinh chỉnh để đảm bảo chất lượng và tính thẩm mỹ.
3. **Mở rộng yêu cầu:** Nhận thấy rằng cùng một chiến dịch có thể có các thiết kế rất khác nhau tùy thuộc vào đối tượng mục tiêu (xem ví dụ về phòng khám nha khoa cho người đi làm và trẻ em trong Hình 7 của bài báo, các tác giả đã sử dụng GPT-4o để mở rộng mỗi ý định ban đầu thành 4 yêu cầu banner riêng biệt. Mỗi yêu cầu nhắm đến một cặp (đối tượng mục tiêu, mục đích) khác nhau.
4. **Đa dạng hóa kích thước:** Để đảm bảo tính thực tiễn trong ngành, 400 yêu cầu này tiếp tục được mở rộng ra 13 kích thước banner tiêu chuẩn.

### 5.3. Đặc điểm Đa phương thức và Quy mô

- **Đa phương thức:** Mỗi mẫu dữ liệu trong benchmark là một cặp bao gồm **đầu vào hình ảnh (logo vector)** và **đầu vào văn bản (brief chi tiết)**.
- **Quy mô:**
  - 100 logo độc nhất.
  - 400 yêu cầu banner đa dạng.
  - Tổng cộng **5,200 mẫu benchmark**..

## 6. Phương pháp Thực nghiệm (Experimental Methodology)

Để chứng minh hiệu quả của BannerAgency, các tác giả đã thiết kế một loạt các thí nghiệm so sánh và đánh giá một cách khắt khe.

### 6.1. Các Mô hình Đối chứng (Baselines)

BannerAgency được so sánh với các mô hình tạo ảnh và thiết kế đồ họa hàng đầu hiện nay:

- **DALL·E3**: Một mô hình T2I thương mại mạnh mẽ từ OpenAI.
- **FLUX.1-schnell**: Một mô hình T2I mã nguồn mở hiệu suất cao, cũng được dùng làm backbone cho công cụ T2I của BannerAgency.
- **OpenCOLE**: Một mô hình chuyên về thiết kế đồ họa dựa trên prompt, yêu cầu fine-tuning trên dữ liệu cụ thể.

## 6.2. Backbone LLM cho BannerAgency

Các tác giả đã thử nghiệm BannerAgency với hai MLLM hàng đầu để đánh giá sự linh hoạt của framework:

- **GPT-4o (phiên bản 2024-10-21)**
- **Claude 3.5 Sonnet (phiên bản 2024-10-22)**

## 6.3. Các Chỉ số Đánh giá (Evaluation Metrics)

Thay vì sử dụng các chỉ số đánh giá thiết kế chung chung, nhóm nghiên cứu đã phát triển 6 chỉ số chuyên biệt cho banner quảng cáo, dựa trên việc khảo sát sâu rộng các tài liệu về nguyên tắc thiết kế quảng cáo hiệu quả. Mỗi chỉ số được đánh giá trên thang điểm 5.

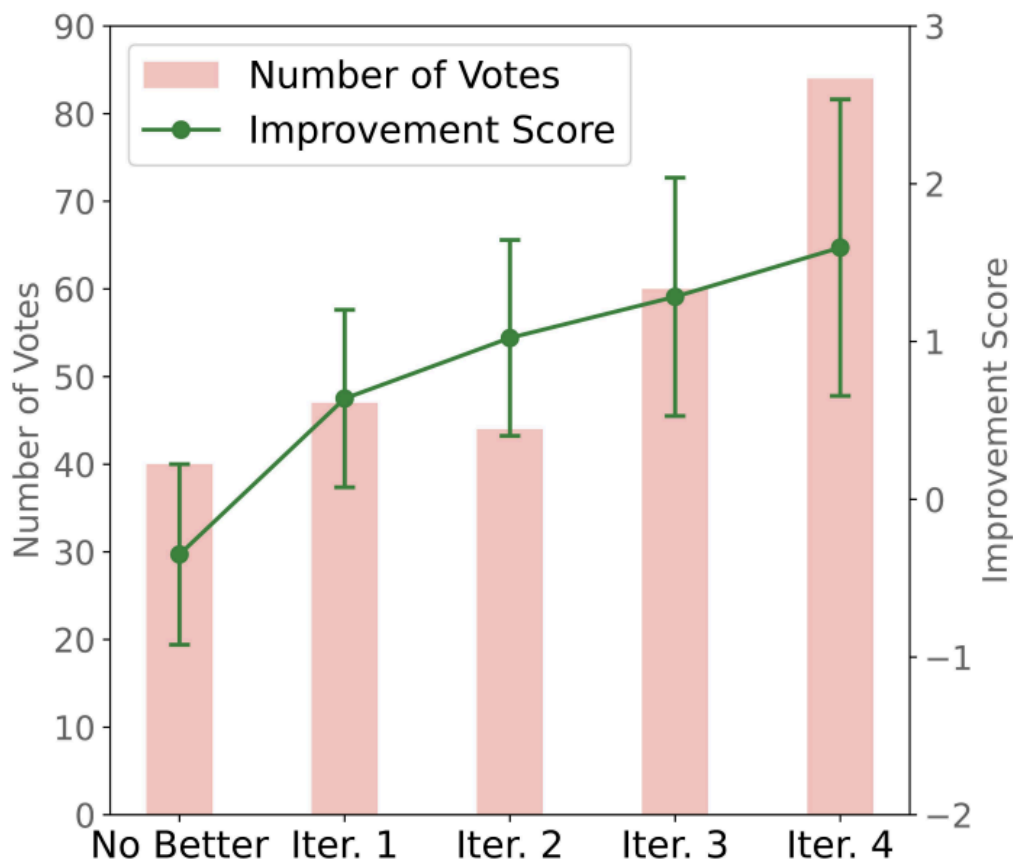
- **TAA (Target Audience Alignment)**: Mức độ banner phù hợp với chủ đề, đối tượng và mục đích được yêu cầu.
- **LPS (Logo Placement Score)**: Mức độ logo được tích hợp tốt về mặt hiển thị, kích thước và vị trí
- **CTAE (CTA Effectiveness)**: Mức độ nút kêu gọi hành động (CTA) rõ ràng, hấp dẫn và nổi bật .
- **CPYQ (Copywriting Quality)**: Chất lượng của văn bản (tiêu đề, phụ đề) về sự rõ ràng, thuyết phục và đúng ngữ pháp.
- **BIS (Brand Identity Score)**: Mức độ banner phù hợp với nhận diện thương hiệu (màu sắc, typography) ngoài vị trí logo.
- **AQS (Aesthetic Quality Score)**: Sức hấp dẫn thị giác tổng thể, bao gồm sự hài hòa màu sắc, cân bằng bố cục và chất lượng thiết kế.



Method	TAA	LPS	CTAE	CPYQ	BIS	AQS
<i>Baselines</i>						
DALL-E3*	3.17	1.78	4.18	2.48	3.06	3.10
FLUX.1-schnell*	3.36	3.67	4.28	2.64	3.09	3.12
OpenCOLE	3.56	1.20	2.85	3.20	2.84	3.29
<i>BannerAgency</i>						
<i>SVG Implementation</i>						
[E] GPT-4o w/o BG	4.15	4.51	4.04	4.25	3.48	3.19
[D] Claude3.5-Sonnet w/o BG	4.40	<b>4.57</b>	4.63	4.48	4.26	3.77
[A] Claude3.5-Sonnet	<b>4.56</b>	4.43	<b>4.94</b>	<b>4.53</b>	<b>4.36</b>	<b>3.92</b>
<i>Figma Implementation</i>						
[C] GPT-4o	<b>4.39</b>	4.22	4.66	4.00	4.14	3.69
[B] Claude3.5-Sonnet	4.34	<b>4.33</b>	<b>4.81</b>	<b>4.19</b>	<b>4.22</b>	<b>3.85</b>

(a) Banner quality metrics (GPT-4o, human-validated in Tab. 2).

#### 6.4. Nghiên cứu Người dùng (Human Study)



## ) Effectiveness of the iterative refinement process.

Để có được đánh giá khách quan nhất, ba nghiên cứu trên người thật đã được tiến hành [406]:

1. **Nghiên cứu sở thích hệ thống:** 20 người tham gia so sánh từng cặp thiết kế được tạo ra bởi 5 biến thể khác nhau của BannerAgency để chọn ra cái họ ưa thích hơn.
2. **Nghiên cứu hiệu quả tinh chỉnh:** 15 người tham gia đánh giá sự cải thiện chất lượng qua 4 vòng lặp tinh chỉnh của 20 thiết kế khác nhau.
3. **Nghiên cứu tương quan người-LLM:** 17-19 người tham gia chấm điểm các banner theo 6 chỉ số trên để xác thực xem liệu đánh giá tự động của GPT-4o có tương đồng với cảm nhận của con người hay không.

## 7. Kết quả và Phân tích (Results and Analysis)

### 7.1. So sánh Định lượng với các Baseline

Kết quả từ Bảng 1(a) của bài báo cho thấy một cách rõ ràng: **BannerAgency vượt trội hơn tất cả các mô hình đối chứng trên mọi chỉ số đánh giá** [439, 440].

- **BannerAgency (Claude 3.5 Sonnet, SVG):** Đạt điểm số cao nhất ở hầu hết các hạng mục, ví dụ TAA (4.56), CTAE (4.94), CPYQ (4.53), BIS (4.36), và AQS (3.92).
- **Baselines (DALL-E3, FLUX, OpenCOLE):** Có điểm số thấp hơn đáng kể, đặc biệt ở các chỉ số yêu cầu sự hiểu biết sâu sắc về ngữ cảnh như LPS (tích hợp logo) và BIS (nhận diện thương hiệu). Ví dụ, DALL-E3 và OpenCOLE có điểm LPS chỉ 1.78 và 1.20, cho thấy chúng gần như thất bại trong việc tích hợp logo được cung cấp.

## 7.2. Phân tích Điểm mạnh của BannerAgency

Sự vượt trội của BannerAgency đến từ việc nó giải quyết được các hạn chế cố hữu của các baseline:

- **Tích hợp đa phương thức:** Không giống các baseline chỉ nhận prompt text, BannerAgency xử lý cả text (brief) và hình ảnh (logo), giúp tạo ra các thiết kế thực sự mang dấu ấn thương hiệu.
- **Khả năng chỉnh sửa và kiểm soát:** Việc tạo ra text và các yếu tố đồ họa dưới dạng các đối tượng riêng biệt giúp tránh được tình trạng chữ bị méo mó, khó đọc thường thấy ở các mô hình T2I.
- **Linh hoạt về kích thước:** BannerAgency có thể tạo ra banner với bất kỳ kích thước nào được yêu cầu, trong khi các baseline thường bị giới hạn ở một vài kích thước cố định.

## 7.3. Vai trò của Background Designer

Thí nghiệm loại bỏ (ablation study) bằng cách tắt agent Background Designer (phiên bản "w/o BG") đã cho thấy một kết quả ấn tượng: điểm số chất lượng, đặc biệt là AQS (thẩm mỹ), giảm mạnh. Điều này khẳng định rằng việc có một agent chuyên tạo ra một hình nền hấp dẫn và phù hợp là cực kỳ quan trọng, giúp banner không bị nhàm chán và trở nên chuyên nghiệp hơn.

## 7.4. Đánh giá Hiệu quả của Quy trình Tinh chỉnh Lặp lại

Kết quả từ human study cho thấy rõ xu hướng chất lượng tăng dần qua các vòng lặp tinh chỉnh. Điểm cải thiện (Improvement Score) tăng đều, và số phiếu bầu chọn cho các phiên bản ở vòng lặp 3 và 4 là cao nhất.

Tuy nhiên, một phát hiện thú vị là không phải lúc nào phiên bản cuối cùng cũng là tốt nhất trong mắt mọi người. Một số người tham gia lại thích các phiên bản ở vòng lặp 1 hoặc 2 hơn. Điều này nhấn mạnh giá trị của việc BannerAgency lưu lại toàn bộ quá trình, cho phép người dùng có nhiều lựa chọn để tìm ra thiết kế "hoàn hảo" theo gu thẩm mỹ chủ quan của họ.

## 8. Điểm mạnh (Strengths) và Hạn chế (Limitations)

### 8.1. Điểm mạnh

- **Kiến trúc Đa Agent Sáng tạo:** Việc mô phỏng một đội ngũ sáng tạo với các vai trò chuyên biệt là một hướng tiếp cận đột phá, giúp phân rã bài toán phức tạp thành các nhiệm vụ nhỏ hơn và giải quyết chúng hiệu quả hơn.
- **Đầu ra Có thể Chỉnh sửa (Editable Output):** Đây là ưu điểm cốt lõi và mang tính cách mạng, giải quyết được vấn đề lớn nhất của các mô hình T2I hiện tại và giúp BannerAgency thực sự hữu ích trong môi trường làm việc chuyên nghiệp.
- **Quy trình Tinh chỉnh Lặp lại:** Cơ chế memory-augmented refinement mô phỏng chính xác cách các nhà thiết kế làm việc, cho phép cải tiến liên tục và tạo ra nhiều phương án thiết kế chất lượng cao.
- **Benchmark Chất lượng cao:** Việc xây dựng và công bố BannerRequest400 là một đóng góp to lớn cho cộng đồng nghiên cứu, tạo ra một tiêu chuẩn mới để đánh giá các hệ thống trong tương lai.
- **Hiệu suất Vượt trội:** Các kết quả thực nghiệm cả về định lượng và định tính đều chứng minh một cách thuyết phục rằng BannerAgency hiệu quả hơn hẳn các phương pháp đối chứng.

## 8.2. Hạn chế

Mặc dù rất ấn tượng, BannerAgency vẫn còn một số hạn chế cần được xem xét và cải thiện.

- **Phụ thuộc vào Năng lực của MLLM Gốc:** Hệ thống kế thừa các hạn chế của các MLLM làm backbone. Nó có thể gặp khó khăn trong việc hiểu các chỉ dẫn thương hiệu (brand guidelines) quá phức tạp hoặc các sắc thái văn hóa tinh tế mà một nhà thiết kế con người có thể dễ dàng nắm bắt.
- **Chưa Đo lường Hiệu quả Quảng cáo Thực tế:** Nghiên cứu chủ yếu tập trung vào đánh giá chất lượng thiết kế và sở thích của người dùng, chứ chưa đo lường các chỉ số kinh doanh thực tế như Tỷ lệ nhấp chuột (Click-Through Rate - CTR) hay Tỷ lệ chuyển đổi (Conversion Rate). Một banner đẹp mắt không nhất thiết đồng nghĩa với một banner mang lại hiệu quả cao về mặt doanh thu.
- **Thách thức về Tính Thích ứng (Responsive):** Bài báo chưa đánh giá sâu về khả năng tự động điều chỉnh bố cục để phù hợp với vô số kích thước màn hình và nền tảng khác nhau trong thực tế quảng cáo trực tuyến ngày nay.
- **Nguy cơ Đồng nhất hóa Thẩm mỹ:** Nếu các hệ thống như BannerAgency được sử dụng rộng rãi, có nguy cơ chúng sẽ tạo ra các quảng cáo có phong cách tương tự nhau, làm giảm đi sự độc đáo và khác biệt mà các thương hiệu luôn tìm kiếm.
- **Yêu cầu về Tài nguyên Tính toán:** Việc vận hành đồng thời nhiều agent MLLM đòi hỏi một nguồn tài nguyên tính toán đáng kể, đây có thể là một rào cản cho các doanh nghiệp nhỏ hoặc người dùng cá nhân muốn triển khai hệ thống.

## 9. Đề xuất Cải tiến (Future Work & Improvements)

Dựa trên các hạn chế đã phân tích, có thể đề xuất một số hướng cải tiến và phát triển trong tương lai cho BannerAgency:

- **Mở rộng Đầu vào Đa phương thức:** Cho phép hệ thống nhận thêm các loại đầu vào khác như hình ảnh sản phẩm, video, hoặc thậm chí là một website có sẵn để tự động trích xuất phong cách thiết kế.
- **Tích hợp Module Tối ưu hóa Chuyển đổi:** Kết nối BannerAgency với các nền tảng quảng cáo để thu thập dữ liệu về CTR và conversion. Từ đó, xây dựng một agent "nhà tối ưu hóa" sử dụng học tăng cường (Reinforcement Learning) để đề xuất các thay đổi thiết kế nhằm tối đa hóa hiệu quả kinh doanh.
- **Phát triển Logic Thiết kế Responsive:** Bổ sung logic để các agent có thể tự động tạo ra các biến thể bố cục (layout variants) cho cùng một thiết kế, đảm bảo hiển thị tối ưu trên cả desktop, tablet và mobile.
- **Tăng cường Tính Đa dạng và Sáng tạo:** Nghiên cứu các cơ chế để khuyến khích sự "sáng tạo bất ngờ" (serendipity) trong thiết kế, tránh việc đi vào lối mòn thẩm mỹ. Có thể cho phép người dùng điều chỉnh mức độ "an toàn" và "phá cách" của các agent.
- **Tối ưu hóa Hiệu suất và Chi phí:** Nghiên cứu các kỹ thuật chưng cất mô hình (model distillation) hoặc sử dụng các mô hình LLM nhỏ hơn, chuyên biệt hơn cho từng tác vụ của agent để giảm yêu cầu về tài nguyên và chi phí vận hành.