**TRƯỜNG ĐẠI HỌC THỦ ĐÔ HÀ NỘI**

**KHOA KHOA HỌC TỰ NHIÊN VÀ CÔNG NGHỆ**

**\*\*\*\*\*\*\***

A blue and black logo

AI-generated content may be incorrect.

**BÁO CÁO BÀI TẬP LỚN CUỐI KỲ**

**HỌC PHẦN: NHẬP MÔN TRÍ TUỆ NHÂN TẠO**

**CHỦ ĐỀ:**

**TÓM TẮT VĂN BẢN TIẾNG VIỆT SỬ DỤNG HỆ THỐNG MÔ HÌNH**

**REASONING – CRITICAL**

|  |  |  |
| --- | --- | --- |
| **Giảng viên hướng dẫn** | **:** | **Trần Thị Thu Phương** |
| Nguyễn Bảo Long | : | 223001756 |
| Nguyễn Gia Bảo | : | 223001695 |
| Lê Sơn Trường | : | 223001805 |
| Phạm Quỳnh Chi | : | 223001699 |
| Nguyễn Đức Thắng | : | 223001791 |

**Hà Nội, 10/2025**

**MỤC LỤC**

[1. GIỚI THIỆU 1](#_Toc213731320)

[1.1. Bối cảnh nghiên cứu 1](#_Toc213731321)

[1.2. Hệ thống Reason–Critic 1](#_Toc213731322)

[1.3. Mục tiêu và phạm vi nghiên cứu 2](#_Toc213731323)

[2. PHƯƠNG PHÁP LUẬN 3](#_Toc213731324)

[2.1. Quy trình tóm tắt tổng quan 3](#_Toc213731325)

[2.2. Kiến trúc thành phần và cấu hình 4](#_Toc213731326)

[2.2.1. Cấu trúc 4](#_Toc213731327)

[2.2.2. Tham số sinh mẫu 4](#_Toc213731328)

[2.3. Ràng buộc định dạng và nội dung 5](#_Toc213731329)

[2.4. Thiết kế Prompt và Tiêu chí đánh giá 5](#_Toc213731330)

[2.4.1. Thiết kế Prompt 5](#_Toc213731331)

[2.4.2. Bộ 6 Tiêu chí đánh giá (Thang 1–5) 6](#_Toc213731332)

[3. THIẾT KẾ THỰC NGHIỆM 7](#_Toc213731333)

[3.1. Tập dữ liệu (Datasets) 7](#_Toc213731334)

[3.2. Các mô hình thử nghiệm 7](#_Toc213731335)

[3.3. Chỉ số đánh giá (Metrics) 8](#_Toc213731336)

[4. KẾT QUẢ VÀ PHÂN TÍCH 9](#_Toc213731337)

[4.1. Kết quả định lượng 9](#_Toc213731338)

[4.1.1. Mô hình Phi-3-mini-4k-instruct (Trên 496 bộ dữ liệu tiếng Anh) 9](#_Toc213731339)

[4.1.2. Mô hình Phi-3-mini-4k-instruct (Trên 496 bộ dữ liệu tiếng Anh) 9](#_Toc213731340)

[4.2. Kết quả định tính 10](#_Toc213731341)

[4.3. Phân tích hiệu quả của các cơ chế 10](#_Toc213731342)

[5. PHÂN TÍCH HIỆU QUẢ TINH CHỈNH 11](#_Toc213731343)

[6. HƯỚNG KHẮC PHỤC VÀ PHÁT TRIỂN 11](#_Toc213731344)

[5.1. Tối ưu hóa quy trình chính 11](#_Toc213731345)

[5.2. Cải thiện năng lực xử lý 11](#_Toc213731346)

[5.3. Mở rộng phạm vi tác vụ 11](#_Toc213731347)

[7. KẾT LUẬN 12](#_Toc213731348)

# 1. GIỚI THIỆU

## 1.1. Bối cảnh nghiên cứu

Trong lĩnh vực Xử lý Ngôn ngữ Tự nhiên (NLP), các mô hình ngôn ngữ lớn (LLM) thể hiện năng lực mạnh mẽ trong việc sinh văn bản nhưng vẫn gặp vấn đề về độ chính xác và trung thực do lỗi “ảo giác” (hallucination).

Để khắc phục, các khung cải tiến lặp (iterative refinement) đã được phát triển, mô phỏng quá trình con người viết – tự đánh giá – sửa chữa.

Các phương pháp như Self-Refine và Critique–Improve giúp giảm suy diễn sai và tăng khả năng bám sát văn bản nguồn, qua đó nâng cao độ tin cậy và chất lượng của văn bản đầu ra.

## 1.2. Hệ thống Reason–Critic

Nghiên cứu này tập trung vào một kiến trúc cụ thể: Reason–Critical. Đây là một kiến trúc tóm tắt văn bản lặp (iterative) gồm hai pha:

1. **Reasoning model:** Phân tích văn bản, tạo ra một chuỗi suy luận (reasoning trace) và một bản tóm tắt ban đầu. Mô hình này sử dụng nhiệt độ (temperature) cao (≈0.3) để tăng tính đa dạng trong suy luận.
2. **Critic model:** Đối chiếu bản tóm tắt với văn bản gốc, chấm điểm theo 6 tiêu chí định sẵn và đưa ra phản hồi ngắn gọn để cải thiện. Mô hình này sử dụng nhiệt độ thấp (≈0.1) để giữ sự ổn định và nhất quán trong đánh giá.

Chu trình này được lặp lại. Ở các vòng sau, mô hình Reasoning sẽ tiếp nhận phản hồi của Critic và thực hiện chỉnh sửa có chọn lọc, chỉ sửa các phần được xác định là sai và giữ nguyên nội dung đã đúng.

## 1.3. Mục tiêu và phạm vi nghiên cứu

Nghiên cứu này được thực hiện nhằm mục đích phân tích chi tiết hiệu quả, các ưu điểm về phương pháp luận và các rủi ro tiềm ẩn của hệ thống Reason–Critic khi áp dụng cho tác vụ tóm tắt văn bản có cấu trúc song ngữ (Anh–Việt).

Phạm vi triển khai của hệ thống được giới hạn trong các tham số sau:

|  |
| --- |
| {  "System\_Config": {  "max\_iters": {  "value": 3,  "description": "Vận hành tối đa 3 vòng lặp."  },  "min\_improve": {  "value": 0.1,  "description": "Ngưỡng cải thiện tối thiểu giữa các vòng là 0.1."  },  "n\_ctx": {  "value": 4096,  "description": "Sử dụng ngữ cảnh 4096 token."  },  "n\_gpu\_layers": {  "value": -1,  "description": "GPU layers được đặt ở chế độ toàn phần."  }  } } |

# 2. PHƯƠNG PHÁP LUẬN

Phương pháp luận của nghiên cứu này bao gồm việc thiết kế một quy trình (pipeline) khép kín, định nghĩa các thành phần kiến trúc, thiết kế bộ prompt hướng dẫn chi tiết, và xác lập các ràng buộc đầu ra nghiêm ngặt.

## 2.1. Quy trình tóm tắt tổng quan

**Bước khởi đầu - Vòng 0 - No Reason:** Mô hình đọc văn bản gốc và thực hiện tóm tắt trực tiếp, không có ràng buộc hay hướng dẫn suy luận (Reasoning).

**Bước 1 – Vòng 1 - Reasoning First:** Mô hình đọc văn bản gốc và sinh ra một đối tượng JSON bao gồm reasoning trace (chứa [ topic, key\_ideas, filtered\_ideas ]) và summary (bản tóm tắt).

**Bước 2 – Vòng 1 - Critical First:** Đánh giá kết quả từ Bước 1 theo 6 tiêu chí, trả về điểm số [scoring] và phản hồi văn bản [feedback\_text].

**Bước 3 – Vòng 2 - Reasoning Refine:** Mô hình nhận văn bản gốc, bản tóm tắt và phản hồi từ vòng trước. Mô hình chỉ sửa các phần cần thiết, giữ nguyên số liệu và nội dung đã đúng.

**Bước 4 – Vòng 2 - Critical Refine:** Đánh giá lại kết quả đã tinh chỉnh, ưu tiên trọng số cho tiêu chí Consistency (Tính nhất quán).

**Bước 5 – Lặp:** Tiếp tục các vòng refine (Bước 3 và 4) cho đến khi đạt giới hạn max\_iters hoặc không đạt ngưỡng cải thiện min\_improve.

## 2.2. Kiến trúc thành phần và cấu hình

Hệ thống sử dụng một mô hình ngôn ngữ duy nhất nhưng áp dụng hai cấu hình (tham số sinh mẫu) riêng biệt cho hai vai trò.

### 2.2.1. Cấu trúc

1. **Reasoning model**
   * **Đầu vào:** Văn bản gốc (ở vòng 1) hoặc Văn bản gốc + JSON cũ + Feedback (ở các vòng refine).
   * **Đầu ra:** Một đối tượng JSON duy nhất tuân thủ định dạng nghiêm ngặt, bao gồm *reasoning* [ topic, key\_ideas, filtered\_ideas ] và *summary*.
2. **Critical model**
   * **Đầu vào:** Văn bản gốc + JSON do mô hình Reasoning tạo ra.
   * **Đầu ra:** Một đối tượng JSON duy nhất chứa *scoring* [ 6 tiêu chí, thang điểm 1–5 ] và *feedback\_text* (phản hồi dạng văn bản ngắn).

### 2.2.2. Tham số sinh mẫu

Các tham số vận hành của mô hình được thiết lập như sau:

|  |
| --- |
| "reason\_params": {  "max\_new\_tokens": 512,  "temperature": 0.3,  "top\_p": 0.9,  "seed": 42  },  "critic\_params": {  "max\_new\_tokens": 512,  "temperature": 0.1,  "top\_p": 0.8,  "seed": 314159  },  "llama\_cpp\_params": {  "n\_gpu\_layers": -1,  "n\_ctx": 4096,  "verbose": true  },  "flow\_params": {  "max\_iters": 3,  "min\_improve": 0.1  } |

Việc hạ nhiệt độ của Critic (0.1) so với Reasoning (0.3) là một lựa chọn có cơ sở, nhằm mục đích giúp ổn định quá trình chấm điểm, giảm các phản hồi ngẫu nhiên. Đồng thời, việc này vẫn giữ cho mô hình Reasoning có độ sáng tạo cần thiết trong quá trình sinh và tinh chỉnh nội dung.

## 2.3. Ràng buộc định dạng và nội dung

Hệ thống áp dụng các ràng buộc cứng (hard constraints) để đảm bảo chất lượng và tính nhất quán của đầu ra:

* JSON hợp lệ: Đầu ra phải là một đối tượng JSON duy nhất, không có văn bản thừa (ví dụ: "Here is the JSON...").
* Không trường rỗng: Mọi trường (field) trong JSON phải có giá trị.
* Bảo toàn thông tin: Phải giữ nguyên tuyệt đối các số liệu, ngày tháng, và tên riêng từ văn bản gốc.
* Giới hạn độ dài: Bản tóm tắt không được vượt quá 100 từ.
* Không suy diễn: Cấm thêm hoặc bịa đặt thông tin không có trong văn bản nguồn.

## 2.4. Thiết kế Prompt và Tiêu chí đánh giá

### 2.4.1. Thiết kế Prompt

**Reasoning:**

* Vòng 1: Hướng dẫn mô hình xác định topic, key\_ideas, filtered\_ideas; tự kiểm tra bằng checklist; tuân thủ định dạng JSON và ràng buộc không suy diễn.
* Vòng 2-n: Hướng dẫn mô hình đọc JSON cũ và feedback mới, chỉ sửa các điểm được nêu trong feedback, giữ nguyên các phần đã đúng, và không viết lại toàn bộ bản tóm tắt.

**Critical:**

* Vòng 1: Hướng dẫn mô hình chấm 6 tiêu chí (thang 1–5) theo Verification Rule (Quy tắc xác minh), đưa ra phản hồi ngắn, và phải có kiểm chứng.
* Vòng 2-n: Hướng dẫn mô hình ưu tiên tiêu chí Consistency, giữ nguyên các quy tắc cũ, và chỉ nêu các lỗi trọng yếu kế tiếp (nếu có).

### 2.4.2. Bộ 6 Tiêu chí đánh giá (Thang 1–5)

Mô hình Critic đánh giá các bản tóm tắt theo 6 tiêu chí chính, với trọng số như sau:

* **Factuality (Độ trung thực): 0.30**

Phản ánh đúng sự thật, không sai lệch hay bịa đặt.

* **Clarity (Độ rõ ràng): 0.20**

Rõ ràng, mạch lạc, dễ hiểu.

* **Logical Coherence (Tính mạch lạc logic): 0.15**

Các ý kết nối hợp lý, không mâu thuẫn nội tại.

* **Coverage (Độ bao phủ): 0.15**

Bao quát đủ các ý chính và chi tiết quan trọng.

* **Utility (Tính hữu dụng): 0.10**

Hữu ích cho việc cải thiện hoặc tổng hợp.

* **Consistency (Tính nhất quán): 0.10**

Nhất quán nội tại (giữa các phần của tóm tắt).

# 3. THIẾT KẾ THỰC NGHIỆM

Phần này mô tả các tập dữ liệu, các mô hình được sử dụng và các chỉ số đo lường hiệu suất được định nghĩa trong nghiên cứu.

## 3.1. Tập dữ liệu (Datasets)

Nghiên cứu sử dụng hai tập dữ liệu tóm tắt tin tức bao gồm:

* **Tiếng Anh:** *SurAyush / News\_Summary\_Dataset*
* **Tiếng Việt:** *LongK171 / VNexpress*

Mỗi bộ dữ liệu bao gồm 1000 mẫu. Nghiên cứu lấy ngẫu nhiên đồng đều 500 văn bản từ mỗi tập (tổng cộng 1000 văn bản).

## 3.2. Các mô hình thử nghiệm

Nghiên cứu tập trung vào các mô hình lượng tử hóa (Quantized Models) quy mô nhỏ, phù hợp với các kịch bản triển khai cần hiệu suất cao và tài nguyên hạn chế. Các mô hình được ghi nhận bao gồm:

* **microsoft/Phi-3-mini-4k-instruct (3.8B):** Có thế mạnh về xử lý tiếng Anh.
* **meta-llama/Llama-3.2-1B - 3B:** Cân bằng khả năng xử lý hai ngôn ngữ.
* **Qwen/Qwen2.5-3B-Instruct-GGUF (2.6B):** Có thế mạnh về xử lý tiếng Việt.

## 3.3. Chỉ số đánh giá (Metrics)

Các định nghĩa về vòng (round) được nhắc lại:

* Vòng 0: Tóm tắt mà không thực hiện reasoning.
* Vòng 1: Tóm tắt có reasoning và chưa áp dụng critical.
* Vòng 2 - n: Tóm tắt có áp dụng reasoning và critical.
* Total Texts: Tổng số văn bản.
* Total Comparisons: Tổng số vòng critical.

Với mỗi văn bản được xử lý, gọi điểm trung bình ở vòng thứ i là AVG[i].

**Reasoning:**

Số lần thành công: với i = 1

Tỉ lệ thành công:

**Critical:**

Số lần thành công: với i > 1

Tỉ lệ thành công:

# 4. KẾT QUẢ VÀ PHÂN TÍCH

Phần này trình bày các kết quả định tính và định lượng được ghi nhận từ các tệp log.

## 4.1. Kết quả định lượng

### 4.1.1. Mô hình Phi-3-mini-4k-instruct (Trên 496 bộ dữ liệu tiếng Anh)

Tóm tắt mà không thực hiện reasoning (Vòng 0):

* Điểm thấp nhất: 3.00
* Điểm cao nhất: 4.45
* Điểm trung bình: 3.7311

Tóm tắt với reasoning (Vòng 1):

* Điểm thấp nhất: 2.10
* Điểm cao nhất: 4.85
* Điểm trung bình: 4.4383
* Số lần reason thành công: 434 / 496
* Tỉ lệ reason thành công: 87.50%

Tóm tắt với reasoning-critical (Vòng 2+):

* Điểm thấp nhất: 3.0
* Điểm cao nhất: 4.85
* Điểm trung bình: 4.5545
* Số lần critic thành công: 202 / 233
* Tỉ lệ critic thành công: 86.70%
* Tỉ lệ critic thất bại: 0%

### 4.1.2. Mô hình Phi-3-mini-4k-instruct (Trên 496 bộ dữ liệu tiếng Anh)

Tóm tắt mà không thực hiện reasoning (Vòng 0):

* Điểm thấp nhất: 2.85
* Điểm cao nhất: 4.70
* Điểm trung bình: 3.9527

Tóm tắt với reasoning (Vòng 1):

* Điểm thấp nhất: 3.00
* Điểm cao nhất: 4.70
* Điểm trung bình: 4.1643
* Số lần reason thành công: 243 / 386
* Tỉ lệ reason thành công: 62.95%

Tóm tắt với reasoning-critical (Vòng 2+):

* Điểm thấp nhất: 3.15
* Điểm cao nhất: 5.00
* Điểm trung bình: 4.2933
* Số lần critic thành công: 350 / 586
* Tỉ lệ critic thành công: 59.73%
* Tỉ lệ critic thất bại: 0%

## 4.2. Kết quả định tính

Kết quả định lượng cho thấy các bản tóm tắt có kèm suy luận (reasoning-based summarization) đạt chất lượng cao hơn đáng kể so với các tóm tắt được sinh trực tiếp không có bước suy luận trung gian. Việc buộc mô hình lý giải logic trước khi tóm tắt giúp nó tập trung hơn vào cấu trúc, quan hệ nhân quả và trọng tâm thông tin, thay vì chỉ rút gọn bề mặt văn bản.

Kết quả định lượng cũng cho thấy hầu hết các chuỗi suy luận (reasoning trace) và bản tóm tắt do mô hình Reasoning sinh ra đều cải thiện đáng kể sau khi nhận phản hồi từ Critic, phù hợp với kỳ vọng lý thuyết của các hệ thống learning-by-feedback (học qua phản hồi).

Sự kết hợp hai mô hình Reasoning và Critic không chỉ cải thiện độ chính xác và tính nhất quán của kết quả mà còn chứng minh rằng tóm tắt có suy luận là hướng tiếp cận ưu việt hơn so với tóm tắt thuần sinh, nhờ khả năng tái lập quy trình tư duy phản biện và tự điều chỉnh dựa trên phản hồi có cấu trúc.

## 4.3. Phân tích hiệu quả của các cơ chế

Thành công định tính của hệ thống xuất phát từ ba cơ chế chính được thiết kế trong phương pháp luận:

* Ràng buộc JSON: Quy định “một JSON hợp lệ duy nhất, không văn bản thừa” giúp giảm lỗi cú pháp và đảm bảo đủ trường dữ liệu nhờ checklist trong prompt.
* Quy tắc xác minh: Buộc Critic kiểm tra kỹ (Verification Rule) trước khi phê bình, giúp giảm “báo động giả” (false alarms) và ngăn mô hình Reasoning sửa sai các phần đã đúng, tránh hiện tượng hồi quy chất lượng (quality regression).
* Cơ chế tinh chỉnh có kiểm soát: Prompt reason\_refine yêu cầu mô hình “chỉ sửa đúng điểm feedback nêu”, kết hợp với việc Critic ưu tiên tiêu chí Consistency ở các vòng sau, giúp hệ thống tập trung vào sửa lỗi thay vì viết lại toàn bộ.

# 5. PHÂN TÍCH HIỆU QUẢ TINH CHỈNH

Thứ nhất, dữ liệu vận hành ghi nhận tỉ lệ vòng lặp Critical tạo ra kết quả hồi quy (suy giảm chất lượng so với vòng trước) là 0%. Điều này xác nhận tính ổn định cao của kiến trúc: các phản hồi do mô hình Critical tạo ra, khi được xử lý bởi mô hình Reasoning, không gây hại hoặc làm giảm điểm số tổng thể của bản tóm tắt.

Thứ hai, các tỉ lệ "thành công" được báo cáo (86.70% cho tiếng Anh và 59.73% cho tiếng Việt) được định nghĩa nghiêm ngặt bởi tham số min\_improve (Mục 1.3 và 3.3). Một vòng lặp chỉ được tính là "thành công" nếu đạt được mức cải thiện điểm số tối thiểu là 0.1.

Do đó, phần tỉ lệ các vòng lặp không được tính là "thành công" – tương ứng là 13.30% (Anh) và 40.27% (Việt) – không đại diện cho sự thất bại của mô hình Critical. Thay vào đó, chúng đại diện cho các vòng lặp tinh chỉnh chỉ tạo ra lợi ích cận biên.

Đây là các trường hợp mà hệ thống đã vận hành đúng (phản hồi được đưa ra và nội dung được sửa đổi), nhưng mức độ cải thiện ròng về điểm số (average\_score) nằm trong khoảng (0 < cải thiện < 0.1).

Hiện tượng này cho thấy một hạn chế của quy trình lặp: hệ thống nhanh chóng đạt đến điểm bão hòa (đặc biệt rõ rệt với dữ liệu tiếng Việt), nơi các cải tiến ở các vòng lặp sau là không đáng kể hoặc nằm dưới ngưỡng phát hiện đã định.

# 6. HƯỚNG KHẮC PHỤC VÀ PHÁT TRIỂN

## 5.1. Tối ưu hóa quy trình chính

Ngưỡng cải thiện cố định 0.1 nên được thay bằng ngưỡng động, giảm dần qua các vòng (ví dụ 0.1 → 0.05) để duy trì cải thiện nhỏ nhưng có giá trị. Đồng thời, tăng số vòng lặp từ 3 lên 5–7 giúp hệ thống tự dừng khi chất lượng bão hòa. Với tiếng Việt, nên tinh chỉnh prompt và thử nghiệm mô hình khác ngoài Qwen2.5-3B để nâng hiệu quả.

## 5.2. Cải thiện năng lực xử lý

Giới hạn n\_ctx = 4096 và mô hình nhỏ (2.6B–3.8B) hạn chế phạm vi áp dụng. Cần mở rộng cửa sổ ngữ cảnh lên 8k–128k và thử mô hình lớn hơn (7B, 13B) để tăng độ chính xác và khả năng khái quát.

## 5.3. Mở rộng phạm vi tác vụ

Bổ sung tóm tắt linh hoạt theo độ dài, tóm tắt trọng tâm (query-focused) và cơ chế tự sửa JSON nhằm nâng cao tính linh hoạt, ổn định và độ tin cậy khi triển khai thực tế.

# 7. KẾT LUẬN

Nghiên cứu này đã thiết kế, triển khai và đánh giá thành công kiến trúc lặp Reason-Critic cho tác vụ tóm tắt văn bản song ngữ (Anh-Việt), sử dụng các mô hình ngôn ngữ quy mô nhỏ.

Kết quả thực nghiệm định lượng đã cung cấp bằng chứng rõ ràng cho các giả thuyết cốt lõi của đề tài:

Tóm tắt có suy luận (Reasoning-based) vượt trội so với tóm tắt thuần sinh: Việc áp dụng một bước suy luận trung gian (Vòng 1) giúp cải thiện đáng kể chất lượng tóm tắt so với phương pháp sinh trực tiếp (Vòng 0). Điều này được thể hiện qua mức tăng điểm trung bình rõ rệt (ví dụ: từ 3.73 lên 4.43 ở tiếng Anh) và tỉ lệ thành công cao (87.50% cho tiếng Anh, 62.95% cho tiếng Việt).

Phê bình có cấu trúc (Structured Criticism) mang lại sự ổn định và cải thiện: Quy trình tinh chỉnh lặp (Vòng 2+) sử dụng mô hình Critical đã chứng minh được hai điểm mạnh chính:

Tính ổn định tuyệt đối: Hệ thống ghi nhận 0% tỉ lệ thất bại, nghĩa là phản hồi của Critic không bao giờ dẫn đến suy giảm chất lượng tóm tắt.

Khả năng cải thiện: Hệ thống tiếp tục nâng cao chất lượng đầu ra sau Vòng 1, đẩy điểm trung bình lên mức cao hơn (ví dụ: 4.55 cho tiếng Anh).

Thành công về mặt phương pháp luận của hệ thống được xây dựng trên ba cơ chế kiểm soát chính: (1) Ràng buộc đầu ra JSON nghiêm ngặt, (2) Quy tắc xác minh (Verification Rule) cho mô hình Critic để tránh "báo động giả", và (3) Cơ chế tinh chỉnh có kiểm soát (controlled refinement) chỉ tập trung vào các lỗi được chỉ định.

Nghiên cứu đã khẳng định rằng kiến trúc Reason-Critic, bằng cách mô phỏng quy trình tư duy phản biện và tự điều chỉnh của con người, là một hướng tiếp cận ưu việt. Nó không chỉ cải thiện độ chính xác và tính nhất quán của kết quả mà còn cung cấp một phương pháp luận mạnh mẽ để giảm thiểu suy diễn sai (hallucination) và tăng cường độ tin cậy của văn bản do AI tạo ra.

Mặc dù các hạn chế về "lợi ích cận biên" và giới hạn tài nguyên (ngữ cảnh 4096-token, mô hình < 4B) đã được ghi nhận, nghiên cứu này đã xác lập một nền tảng vững chắc, chứng minh giá trị không thể phủ nhận của các hệ thống phản hồi lặp có cấu trúc trong các tác vụ NLP đòi hỏi độ trung thực cao.

**\*PHÂN CÔNG THÀNH VIÊN**

|  |  |  |
| --- | --- | --- |
| **Thành viên** | **Vai trò chính** | **Nhiệm vụ** |
| Nguyễn Đức Thắng | Kiểm thử | Thử nghiệm mô hình, phát hiện lỗi, đóng góp phản hồi cho nhóm trưởng nhằm nâng cao chất lượng hệ thống. |
| Lê Sơn Trường | Kiểm thử | Kiểm thử mô hình, rà soát lỗi, cung cấp ý kiến phản hồi kịp thời để tối ưu kết quả hoạt động. |
| Nguyễn Gia Bảo | Thu thập dữ liệu | Tìm kiếm và lựa chọn dataset phù hợp, đảm bảo nguồn dữ liệu đủ chất lượng và đa dạng. |
| Phạm Quỳnh Chi | Nghiên cứu mô hình | Khảo sát, đề xuất và lựa chọn mô hình tối ưu cho bài toán, hỗ trợ nhóm triển khai kỹ thuật. |
| Nguyễn Bảo Long | Kỹ thuật & điều phối | Phát triển code, xây dựng prompt, tiếp nhận phản hồi và liên tục cải thiện chất lượng mô hình. |