

BÁO CÁO:

SỬ DỤNG TRÍ TUỆ NHÂN TẠO (AI) CÓ TRÁCH NHIỆM VÀ ĐẠO ĐỨC TRONG HỌC TẬP VÀ NGHIÊN CỨU

Họ và tên: Đặng Thảo Vy

Mã sinh viên: 25020436

Môn học: Nhập môn Công nghệ số & Trí tuệ nhân tạo

I. NGHIÊN CỨU CHÍNH SÁCH VỀ SỬ DỤNG AI TRONG MÔI TRƯỜNG ĐẠI HỌC

1. Phân tích chính sách của Đại học RMIT Việt Nam (Trường tiêu biểu về chính sách AI)

Trong bối cảnh AI tạo sinh (GenAI) bùng nổ, RMIT là một trong những trường đi đầu trong việc ban hành khung chính sách rõ ràng. Thay vì cấm đoán, RMIT áp dụng phương pháp tiếp cận "Thích ứng và Tích hợp".

- Điểm chính của chính sách:** RMIT phân loại việc sử dụng AI thành 3 mức độ cho các bài kiểm tra/đánh giá: (1) Được phép sử dụng tự do (như một công cụ học tập); (2) Được phép sử dụng nhưng có giới hạn (ví dụ: chỉ dùng để lên ý tưởng, không dùng để viết nội dung); (3) Hoàn toàn không được phép sử dụng. Sinh viên *bắt buộc phải khai báo* (declaration) về việc sử dụng AI, ghi rõ tên công cụ, cách sử dụng và prompt đã dùng.
- Xử lý vi phạm:** Việc dùng AI tạo ra toàn bộ bài làm và nộp dưới tên mình bị coi là "Gian lận hợp đồng" (Contract Cheating) và vi phạm Tính toàn vẹn học thuật (Academic Integrity), có thể dẫn đến đình chỉ học.

2. So sánh với các trường Đại học công lập tại Việt Nam (Ví dụ: ĐHQG Hà Nội / ĐH Bách Khoa)

Hiện tại, nhiều trường đại học công lập lớn tại Việt Nam chưa có một bộ quy tắc độc lập và chi tiết dành riêng cho GenAI như RMIT. Đa số vẫn đang xếp việc lạm dụng AI vào chung nhóm "đạo văn" (Plagiarism) trong quy chế đào tạo hiện hành. Tuy nhiên, các trường như ĐH Bách Khoa Hà Nội (HUST) đã bắt đầu tích hợp các

công cụ kiểm tra AI (như Turnitin AI detection) và khuyến cáo giảng viên chủ động quy định mức độ dùng AI trong từng môn học.

- **Sự khác biệt:** Nếu RMIT thiên về "Hướng dẫn sử dụng minh bạch", thì phần lớn các trường khác đang ở giai đoạn "Phòng vệ và Dò tìm".

3. Nhận định cá nhân

Từ việc phân tích trên, tôi nhận thấy việc cấm AI hoàn toàn là không khả thi và đi ngược lại với xu thế phát triển năng lực số. Chính sách lý tưởng nhất là **khuyến khích sử dụng có kiểm soát và minh bạch hóa**. Các trường cần tập trung vào việc đào tạo sinh viên *cách* đặt câu hỏi (prompt engineering) và *cách* đánh giá thông tin thay vì chỉ chạy đua vũ trang với các phần mềm phát hiện AI.

II. THỰC HIỆN NHIỆM VỤ HỌC TẬP VỚI SỰ HỖ TRỢ CỦA AI

Nhiệm vụ lựa chọn: Tổng hợp tài liệu và lập đề cương cho bài tiểu luận nghiên cứu: *"Tác động của biến đổi khí hậu đến an ninh lương thực tại Đồng bằng sông Cửu Long (ĐBSCL)"*.

Công cụ sử dụng: ChatGPT (Model GPT-4o).

1. Ghi lại quá trình sử dụng (Prompt & Đầu ra)

PROMPT	ĐẦU RA AI (tóm tắt)
"Đóng vai một nhà nghiên cứu môi trường, hãy liệt kê 3 tác động lớn nhất của biến đổi khí hậu đối với nông nghiệp tại Đồng bằng sông Cửu Long hiện nay."	(1) xâm nhập mặn làm giảm diện tích đất trồng lúa. (2) Hạn hán kéo dài ảnh hưởng đến nguồn nước ngọt. (3) Sạt lở bờ sông do thay đổi dòng chảy.
"Dựa vào 3 ý trên, hãy lập một đề cương chi tiết cho bài tiểu luận dài 2000 chữ. Yêu cầu có phần giải pháp thích ứng. Vui lòng cung cấp một số số liệu hoặc dẫn chứng để tôi tự đi tìm hiểu thêm."	Gợi ý cấu trúc 4 phần: - Mở đầu - Thực trạng (với gợi ý số liệu về mức độ thiệt hại GDP) - Tác động cụ thể - Giải pháp (chuyển đổi cơ cấu cây trồng, đê kè).

2. Mô tả cách đánh giá, chỉnh sửa và tích hợp đầu ra của AI một cách sáng tạo

Đây là bước quan trọng nhất để đảm bảo tôi làm chủ kiến thức chứ không phụ thuộc AI:

- **Đánh giá (Evaluation):** Tôi nhận thấy đầu ra của AI về "Sạt lở bờ sông" chủ yếu do khai thác cát và thủy điện thượng nguồn, biến đổi khí hậu chỉ là yếu tố phụ. AI đã có sự nhầm lẫn (hallucination) hoặc gộp chung nguyên nhân.
- **Chỉnh sửa (Editing):** Tôi gạch bỏ phần "Sạt lở bờ sông" trong nguyên nhân trực tiếp của BĐKH, thay thế bằng yếu tố *"Sự gia tăng nhiệt độ cực đoan ảnh hưởng đến sinh lý cây lúa"*.
- **Tích hợp & Kiểm chứng (Integration & Fact-checking):** AI gợi ý *"xâm nhập mặn làm thiệt hại hàng ngàn tỷ đồng"*. Tôi không dùng số liệu chung chung này. Tôi đã vào website của Tổng cục Thống kê và Bộ NN&PTNT để tìm kiếm báo cáo năm 2023, lấy số liệu thực tế *"mùa khô 2019-2020 thiệt hại 39.000 ha lúa"* để đắp vào khung sườn AI đã gợi ý.
- **Tính sáng tạo:** Tôi chỉ dùng AI để tạo "bộ xương" (skeleton) cho bài viết. Toàn bộ "phần thịt" (lập luận, trích dẫn thực tế, số liệu) đều do tôi tự nghiên cứu và viết bằng văn phong của mình.

3. Trích dẫn việc sử dụng AI một cách minh bạch

Theo chuẩn APA 7th Edition (cập nhật về AI), tôi thêm phần khai báo vào cuối bài tiểu luận như sau:

"Tuyên bố về việc sử dụng công cụ hỗ trợ: Trong quá trình hình thành ý tưởng và xây dựng cấu trúc ban đầu cho bài viết này, tác giả đã sử dụng công cụ ChatGPT (phiên bản GPT-4o) do OpenAI phát triển vào ngày [Ngày tháng]. Sau khi nhận được gợi ý cấu trúc, tác giả đã tự rà soát, kiểm chứng tính chính xác của các luận điểm, tự thu thập dữ liệu từ các nguồn tài liệu khoa học và hoàn thiện văn bản cuối cùng. Tác giả chịu hoàn toàn trách nhiệm về nội dung của bài viết này."

III. PHÂN TÍCH CÁC VẤN ĐỀ ĐẠO ĐỨC LIÊN QUAN ĐẾN SỬ DỤNG AI

Việc tích hợp AI vào học thuật mở ra nhiều tiềm năng nhưng cũng đặt ra 3 thách thức đạo đức cốt lõi:

1. Ranh giới giữa hỗ trợ hợp lý và gian lận học thuật

- **Hỗ trợ hợp lý:** Sử dụng AI như một người hướng dẫn (tutor) hoặc trợ lý nghiên cứu: giải thích các khái niệm phức tạp, gợi ý cấu trúc bài viết, kiểm tra lỗi ngữ pháp, hoặc vượt qua "hội chứng sợ trang giấy trắng" (writer's block).
- **Gian lận học thuật:** Sao chép nguyên văn (copy-paste) đầu ra của AI và nhận đó là tác phẩm của mình; yêu cầu AI giải các bài tập tính toán/lập trình để nộp lấy điểm; hoặc dùng AI để tạo ra các số liệu, bài phỏng vấn không có thật.
- **Ranh giới:** Nằm ở khái niệm "**Sự đóng góp nhận thức**" (Cognitive load). Nếu sinh viên vẫn giữ vai trò quyết định, phản biện và tổng hợp kiến thức, đó là hỗ trợ. Nếu sinh viên khoán trắng toàn bộ quá trình tư duy cho máy móc, đó là gian lận.

2. Vấn đề về quyền sở hữu trí tuệ và trích dẫn

- **Vi phạm bản quyền ẩn:** Các mô hình ngôn ngữ lớn (LLMs) được huấn luyện trên lượng dữ liệu khổng lồ có bản quyền mà không có sự xin phép rõ ràng. Khi chúng ta sử dụng đầu ra của AI, chúng ta có nguy cơ vô tình sử dụng chất xám của các tác giả gốc mà không hề trích dẫn họ.
- **Hiện tượng "Ảo giác" (Hallucination) & Bịa đặt nguồn:** AI thường xuyên "bịa" ra các bài báo khoa học với tên tác giả và năm xuất bản trông rất thực tế nhưng không hề tồn tại. Nếu sinh viên mang các trích dẫn này vào bài, họ vi phạm lỗi nguy tạo dữ liệu – một vi phạm đạo đức nghiên cứu nghiêm trọng.

3. Tác động đến quá trình học tập và phát triển kỹ năng (Dài hạn)

- **Sự thui chột tư duy phản biện:** Khi quá phụ thuộc vào câu trả lời tức thời của AI, sinh viên sẽ mất đi khả năng tự tra cứu, chọn lọc và hoài nghi thông tin.
- **Mất đi "tiếng nói cá nhân":** Việc để AI viết hộ khiến bài luận của sinh viên trở nên rập khuôn, máy móc, mất đi văn phong cá nhân và khả năng biểu đạt cảm xúc, sự tinh tế trong ngôn ngữ.
- **Đạo đức với bản thân:** Mục đích cốt lõi của giáo dục không phải là tạo ra sản phẩm (bài luận), mà là **quá trình thay đổi nhận thức** thông qua việc vất vả tạo ra sản phẩm đó. Dùng AI làm hộ tức là sinh viên đang tự tước đoạt quyền được học và phát triển của chính mình.

IV. BỘ NGUYÊN TẮC CÁ NHÂN (6 NGUYÊN TẮC CỐT LÕI)

Dựa trên các phân tích đạo đức trên, tôi xây dựng bộ nguyên tắc cá nhân "6T" (The 6-T Principles) để định hướng bản thân khi sử dụng AI:

1. **Tính Làm chủ (Ownership):** Tôi luôn là người chịu trách nhiệm cuối cùng cho mọi nội dung mình nộp. AI là trợ lý, tôi là tác giả.
2. **Tính Minh bạch (Transparency):** Luôn chủ động khai báo rõ ràng công cụ AI nào đã được sử dụng, sử dụng ở khâu nào (lên ý tưởng, dịch thuật, hay sửa lỗi) trong các bài tập và báo cáo.
3. **Tính Kiểm chứng (Truth-checking):** Tuyệt đối không sử dụng thông tin, số liệu, hoặc trích dẫn do AI cung cấp nếu chưa được đối chiếu và xác minh bằng nguồn tài liệu sơ cấp hoặc học thuật uy tín.
4. **Tính Nguyên bản (Thought-first):** Phải tự suy nghĩ, lập dàn ý sơ bộ và định hình quan điểm cá nhân *trước khi* mở ChatGPT để tìm kiếm sự hỗ trợ, tránh bị tư duy của AI dẫn dắt.
5. **Tính Bảo mật (Trust/Privacy):** Không đưa các thông tin cá nhân, dữ liệu nghiên cứu chưa công bố, hoặc tài sản trí tuệ của người khác (bài giảng của thầy cô) vào prompt của AI để bảo vệ quyền riêng tư.
6. **Tính Rèn luyện (Training):** Chỉ dùng AI để làm những việc giải phóng thời gian cơ học, tuyệt đối không dùng AI để bỏ qua quá trình rèn luyện các kỹ năng cốt lõi (như kỹ năng tư duy phân tích, lập luận logic).

V. INFOGRAPHIC MINH HỌA



