PHÁT TRIỂN KỸ NĂNG SỬ DỤNG AI CÓ TRÁCH NHIỆM VÀ ĐẠO ĐỨC TRONG HỌC TẬP VÀ NGHIÊN CỨU HỌC THUẬT

PHẦN 1: NGHIÊN CỨU & PHÂN TÍCH CHÍNH SÁCH SỬ DỤNG AI CỦA TRƯỜNG ĐẠI HỌC

Hiện nay, việc tích hợp Trí tuệ nhân tạo (AI) vào giáo dục đại học vừa là cơ hội lớn mang tính cách mạng, vừa là một thách thức kiểm định tính liêm chính học thuật. Trong phạm vi báo cáo này, tác giả tiến hành nghiên cứu sâu chính sách về AI của Đại học Quốc gia và khung hướng dẫn thực hành của Đại học RMIT Việt Nam để làm hệ quy chiếu phân tích.

1.1 Tóm tắt các điểm chính trong chính sách của các trường

Các trường đại học lớn tại Việt Nam đã chuyển dịch từ tư duy "cấm đoán hoàn toàn" sang cấu trúc "quản lý rủi ro và định hướng sử dụng". Cụ thể:

• Phân định rõ ràng ranh giới hành vi: Sinh viên được quyền sử dụng AI để tìm kiếm ý tưởng ban đầu (brainstorming), tối ưu hóa cấu trúc bài viết, hiệu đính ngữ pháp (proofreading). Tuy nhiên, tuyệt đối cấm việc sử dụng AI để tạo ra sản phẩm cuối cùng rồi nộp dưới danh nghĩa cá nhân (AI-authored text submission) mà không có sự chỉnh sửa hay trích dẫn.

• Nghĩa vụ minh bạch thông tin: Bất kỳ sản phẩm học thuật nào có sự can thiệp của Generative AI (như ChatGPT, Claude, Gemini) đều phải được khai báo cụ thể tại phần mục lục tài liệu tham khảo hoặc đính kèm phụ lục mô tả các bước 'Prompts' đã thực hiện.

1.2 So sánh mở rộng với xu hướng quốc tế

Khi so sánh với Khung quy định của các đại học top đầu thế giới như Oxford hay Harvard, chính sách của các trường tại Việt Nam có tính tương đồng cao ở mô hình chia ba phân vùng (Three-Zone Framework):

1. Vùng trắng (White Zone - Tự do): Sử dụng AI tra cứu khái niệm, tìm kiếm từ khóa, định dịch tài liệu thô.

2. Vùng xám (Grey Zone - Có điều kiện): Sử dụng AI để lập dàn ý bài viết, tối ưu hóa code lập trình. Cần khai báo chi tiết và có sự phê duyệt ngầm định trong đề bài bài tập.

3. Vùng đỏ (Red Zone - Nghiêm cấm): Sử dụng AI tạo lập toàn bộ bài văn, giải hộ đề thi trực tuyến, tạo số liệu giả khoa học.

> **★ LƯU Ý ĐẠO ĐỨC**
> *Sự khác biệt cốt lõi nằm ở mức độ kiểm soát kỹ thuật. Các trường quốc tế ứng dụng chặt chẽ hệ thống AI-Detector (Turnitin AI, GPTZero) kết hợp vấn đáp trực tiếp, trong khi đa phần đại học Việt Nam hiện vẫn dựa trên tinh thần tự giác tôn trọng đạo đức học thuật của người học.*

PHẦN 2: NHẬT KÝ THỰC HIỆN NHIỆM VỤ HỌC TẬP CÙNG AI

Nhiệm vụ lựa chọn: Nghiên cứu, xây dựng đề cương và viết bài luận học thuật chủ đề: "Tác động của biến đổi khí hậu đến nông nghiệp vùng Đồng bằng sông Cửu Long và giải pháp thích ứng bền vững".

2.1 Nhật ký các câu lệnh (Prompts) và Đầu ra của AI (AI Output)

• Câu lệnh 1 (Định hình dàn ý): "Đóng vai là một chuyên gia môi trường học thuật. Hãy thiết lập một cấu trúc dàn ý chi tiết cho bài luận nghiên cứu về tác động của biến đổi khí hậu đến nông nghiệp Đồng bằng sông Cửu Long, bao gồm các phần chính, phụ, tập trung vào thực trạng hạn mặn và giải pháp chuyển đổi sinh kế."

• Phản hồi của AI (Tóm tắt): AI trả về cấu trúc gồm 4 phần tiêu chuẩn: (1) Mở đầu; (2) Thực trạng hạn mặn và cơ chế tác động nông nghiệp; (3) Các giải pháp thích ứng; (4) Kết luận. Dàn ý logic nhưng mang tính lý thuyết chung chung, chưa đi vào chi tiết địa bàn trọng điểm.

• Câu lệnh 2 (Yêu cầu dẫn chứng tài liệu): "Cung cấp cho tôi 3 nguồn tài liệu học thuật thực tế, kèm tên tác giả và năm xuất bản tại Việt Nam để minh chứng cho số liệu diện tích đất trồng lúa bị ảnh hưởng bởi hạn mặn tại ĐBSCL giai đoạn 2020-2024."

• Phản hồi của AI (Tóm tắt): AI cung cấp 3 tài liệu, trong đó có một nguồn ghi rõ: 'Nguyễn Văn A (2022). Tạp chí Khoa học Nông nghiệp Việt Nam - Nghiên cứu sự suy giảm diện tích lúa vùng nhiễm mặn'.

2.2 Đánh giá phản biện và Quá trình Chỉnh sửa lỗi sai của AI

Khi tiến hành kiểm tra chéo (cross-check) thông tin, tác giả phát hiện ra lỗi nghiêm trọng nhất của Generative AI: Sự ảo tưởng thông tin (Hallucination). Tài liệu của tác giả 'Nguyễn Văn A (2022)' hoàn toàn không tồn tại trên Google Scholar hay hệ thống cơ sở dữ liệu của Tạp chí Khoa học Nông nghiệp Việt Nam. AI đã tự động 'bịa ghép' ra một tên gọi có vẻ học thuật dựa trên phân tích xác suất từ ngữ.

Quy trình xử lý, can thiệp chỉnh sửa của người học:

1. Loại bỏ hoàn toàn nguồn dẫn chứng ảo do AI sinh ra.

2. Tìm kiếm thủ công: Tác giả tự truy cập vào trang số liệu chính thống của Tổng cục Thống kê Việt Nam và báo cáo cập nhật của Bộ Nông nghiệp & Phát triển Nông thôn năm 2023 để lấy con số chính xác: hơn 80.000 ha cây ăn trái và lúa bị thiệt hại do hạn mặn mùa khô 2019-2020.

3. Viết lại nội dung (Paraphrasing): Sử dụng giọng văn mang tính cá nhân, loại bỏ các từ ngữ sáo rỗng thường thấy của AI (ví dụ: 'Như chúng ta đã biết', 'Tóm lại là', 'Một khía cạnh quan trọng khác là'). Thêm các góc nhìn biện giải mang tính cá nhân.

2.3 Trích dẫn minh bạch việc sử dụng AI

Để tuân thủ tính minh bạch, việc sử dụng AI trong bài luận này được trích dẫn theo định dạng chuẩn APA 7th như sau:

• Khai báo trong văn bản: "Mô hình cấu trúc thô của bài viết được tham khảo thông qua hỗ trợ định hình từ công cụ trợ lý ngôn ngữ lớn ChatGPT (OpenAI, 2026), sau đó được người viết điều chỉnh và bổ sung dữ liệu thực địa."

• Danh mục tài liệu tham khảo: OpenAI. (2026). ChatGPT (Phiên bản GPT-4o) [Mô hình ngôn ngữ lớn]. https://chatgpt.com

PHẦN 3: PHAN TÍCH SÂU SẮC CÁC VẤN ĐỀ ĐẠO ĐỨC TRONG HỌC THUẬT

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

Ranh giới đạo đức nằm ở khâu "Tư duy cốt lõi". Hỗ trợ hợp lý là khi AI đảm nhiệm vai trò công cụ cơ giới: sửa lỗi chính tả, dịch thuật ngữ, tìm kiếm từ khóa, gợi ý cấu trúc bài viết để người học vượt qua hội chứng sợ trang giấy trắng (Writer's block).

Gian lận học thuật xảy ra khi sinh viên giao phó hoàn toàn trách nhiệm tư duy (Cognitive Outsourcing) cho máy tính. Việc copy nguyên văn đoạn văn do AI viết, sử dụng AI tạo mã code bài tập lập trình rồi nộp bài mà không hiểu bản chất, bản chất là hành vi mạo nhận tri thức và gian lận công sức sáng tạo.

3.2 Vấn đề quyền sở hữu trí tuệ và trích dẫn

Generative AI không tự sáng tạo ra tri thức mới, nó hoạt động dựa trên thuật toán xác suất tổng hợp dữ liệu từ hàng tỷ văn bản có sẵn trên Internet - vốn là tài sản trí tuệ của các tác giả, nhà nghiên cứu thực thụ. Do đó, sử dụng sản phẩm của AI mà không trích dẫn nguồn gốc không chỉ cấu thành hành vi đạo văn cấp độ hai (secondary plagiarism) mà còn xâm phạm gián tiếp đến quyền sở hữu trí tuệ của cộng đồng học thuật gốc.

3.3 Tác động dài hạn đến quá trình học tập và phát triển kỹ năng

Nếu quá phụ thuộc vào AI, người học sẽ mắc hội chứng "Lười biếng tư duy chủ động" (Cognitive Offloading). Các kỹ năng tối quan trọng trong môi trường đại học như: đọc hiểu tài liệu chuyên sâu, phân tích và phản biện dữ liệu, xây dựng lập luận cá nhân độc lập sẽ dần bị thui chột. Sinh viên có thể ra trường với những tấm bằng điểm số rất cao nhờ AI, nhưng rỗng tuếch về năng lực thực tế và hoàn toàn mất khả năng giải quyết các vấn đề phức tạp khi không có máy tính trợ giúp.

PHẦN 4: XÂY DỰNG BỘ NGUYÊN TẮC CÁ NHÂN VỀ SỬ DỤNG AI

Để định hình trách nhiệm bản thân, tác giả tự thiết lập và cam kết thực hiện nghiêm ngặt bộ 5 nguyên tắc ứng dụng AI cốt lõi dưới đây:

Tên Nguyên Tắc	Nội Dung Cam Kết & Hành Động Cụ Thể
1. Nguyên tắc Làm chủ Tư duy (Human-in-the-loop)	AI luôn luôn đóng vai trò là một trợ lý phản biện, một người gợi ý thô. Toàn bộ cấu trúc luận điểm chính, kết luận và bản sắc học thuật của bài viết bắt buộc phải do chính bản thân người học tư duy và viết ra.
2. Nguyên tắc Kiểm chứng Tuyệt đối (Zero-Trust Output)	Không bao giờ thừa nhận trực tiếp bất kỳ số liệu, tên tác giả hay bài nghiên cứu nào do AI cung cấp. Mọi thông tin đầu ra từ AI bắt buộc phải được đối chiếu lại bằng các công cụ hàn lâm chính thống (Google Scholar, Scopus, Thư viện trường).
3. Nguyên tắc Minh bạch Toàn diện (Full Disclosure)	Luôn công khai một cách trung thực và rõ ràng trong phần phụ lục hoặc ghi chú của tất cả các bài tập về mức độ can thiệp của AI, bao gồm các câu lệnh (prompts) chính đã sử dụng để giảng viên có thể đánh giá đúng tiến trình.
4. Nguyên tắc Bảo mật Thông tin (Data Privacy Compliance)	Tuyệt đối không tải lên hệ thống AI trực tuyến các tài liệu nội bộ, thông tin cá nhân của bạn học, thầy cô hoặc các bài nghiên cứu chưa xuất bản khi chưa được sự cho phép, nhằm tránh rò rỉ dữ liệu học thuật.

5. Nguyên tắc Tập trung Năng lực (Skill-Centric Purpose)	Chỉ sử dụng AI làm đòn bẩy để đẩy nhanh tốc độ thực thi, không sử dụng AI như một giải pháp thay thế cho việc đọc và học. Nếu AI làm hộ một bài tập, bản thân phải tự giải lại hoặc tự viết lại tương đương để đảm bảo kiến thức được chuyển hóa vào não bộ.