TRƯỜNG ĐẠI HỌC KINH TẾ - ĐẠI HỌC QUỐC GIA HÀ NỘI

Học phần: Nhập môn Công nghệ số và Ứng dụng Trí tuệ nhân tạo

BÀI 6 — SỬ DỤNG AI CÓ TRÁCH NHIỆM TRONG HỌC TẬP VÀ NGHIÊN CỨU

Bộ nguyên tắc cá nhân

Họ và tên: Bùi Lê Đình Yên Ái

Ngành học: Kinh tế phát triển

Lớp học phần: VNU1001_E252045

Bài tập: Bài tập 4 — Mục 6.4

I. MỤC TIÊU BÀI TẬP

Bài tập yêu cầu trình bày bộ nguyên tắc cá nhân về việc sử dụng AI có trách nhiệm dựa trên các nghiên cứu đã thực hiện. Điểm mấu chốt của yêu cầu nằm ở hai chữ "cá nhân" và cụm "dựa trên các nghiên cứu đã thực hiện": đây không phải bài chép lại một bộ quy tắc có sẵn, mà phải được rút ra từ chính trải nghiệm của người viết.

Vì vậy, mỗi nguyên tắc trong báo cáo này đều gắn với một tình huống có thật đã xảy ra trong quá trình thực hiện Bài 1 đến Bài 5 của học phần — những chỗ người viết đã suýt sai, hoặc đã sai và phải sửa lại.

II. XUẤT PHÁT ĐIỂM: MỘT MÂU THUẪN TỪ DỮ LIỆU THỰC TẾ

Khảo sát của Nhóm 1 trên 28 sinh viên (thực hiện tại Bài 4 — nghiên cứu "Hành vi sử dụng AI trong tiêu dùng của sinh viên") cho ra hai con số đặt cạnh nhau rất đáng suy nghĩ:

- 92,9% sinh viên có sử dụng AI, trong đó 57,1% sử dụng hàng ngày.
- 50% người tham gia khảo sát đồng thời bày tỏ lo ngại về việc AI cung cấp thông tin sai lệch hoặc có tính thiên vị.

Gần như ai cũng dùng, và một nửa trong số đó biết công cụ mình đang dùng có thể sai — nhưng vẫn dùng. Khoảng cách giữa nhận thức và hành vi đó chính là nơi các vấn đề về liêm chính học thuật nảy sinh. Bộ nguyên tắc dưới đây là câu trả lời cho khoảng cách ấy: nếu đã biết công cụ có thể sai, thì phải có quy trình để cái sai không đi vào bài nộp.

III. BỘ NGUYÊN TẮC CÁ NHÂN: 6 ĐIỀU

Nguyên tắc 1. AI có thể giúp tôi hiểu, nhưng không được học thay tôi

Tôi dùng AI để làm rõ một khái niệm khó, chứ không dùng để có sẵn một đáp án mà mình không hiểu. Phép thử tôi tự đặt ra: nếu bị hỏi lại mà không giải thích được bằng lời của mình, thì tôi chưa được phép đưa nội dung đó vào bài.

> *Minh chứng từ Bài 3: prompt nâng cao V3 đạt 9.5/10 khi tóm tắt mô hình tăng trưởng Solow. Nhưng chính vì kết quả trông rất "chuyên gia" nên nó nguy hiểm — tôi rất dễ*

chép lại phần phân tích trạng thái dừng (steady state) mà không thật sự hiểu phương trình tích lũy vốn. Do đó tôi đặt quy tắc: đọc kết quả của AI xong phải quay lại đối chiếu giáo trình, rồi tự viết lại bằng ngôn ngữ của mình.

Nguyên tắc 2. Mọi số liệu do AI đưa ra đều bị coi là chưa đúng cho tới khi được kiểm chứng

Đây là nguyên tắc cứng rắn nhất và cũng cần thiết nhất đối với một sinh viên ngành Kinh tế phát triển — ngành mà một con số sai có thể kéo theo cả một kết luận chính sách sai. AI có xu hướng tạo ra số liệu nghe rất thuyết phục nhưng không có thật (hiện tượng hallucination). Quy tắc: số nào không truy được về nguồn gốc thì số đó không được đưa vào bài.

Minh chứng từ Bài 5: Google Gemini viết bài blog rất hay nhưng số liệu thống kê về rác thải nhựa còn chung chung. Tôi đã phải thay thế toàn bộ bằng số liệu chính thống của Bộ Tài nguyên và Môi trường, đồng thời áp dụng quy tắc kiểm chứng chéo độc lập (double-check) với nguồn từ cơ quan chính phủ hoặc tổ chức quốc tế uy tín trước khi xuất bản.

Nguyên tắc 3. Ghi nhận sự tham gia của AI một cách minh bạch

Giấu việc mình dùng AI là hình thức thiếu trung thực dễ mắc nhất, vì gần như không ai kiểm tra được. Nhưng liêm chính học thuật là thứ mình giữ cả khi không bị soi. Tôi ghi rõ công cụ nào đã tham gia vào công đoạn nào — ghi nhận cụ thể, không ghi chung chung.

Minh chứng từ Bài 5: tôi đã tích hợp dòng ghi chú "Sản phẩm có sự hỗ trợ kỹ thuật từ Generative AI trong giai đoạn phát triển ý tưởng và minh họa đồ họa" ở cuối bài viết và infographic — nêu rõ AI tham gia ở giai đoạn nào, thay vì một lời thú nhận mơ hồ.

Nguyên tắc 4. Phần đóng góp của tôi phải là phần quan trọng nhất

AI cung cấp các mảnh ghép; việc chọn mảnh nào, ghép ra sao và vì sao lại ghép như vậy phải là của tôi. Nếu bỏ phần đóng góp của tôi ra mà sản phẩm vẫn gần như nguyên vẹn, thì đó không còn là bài của tôi nữa.

Minh chứng từ Bài 5: đóng góp cá nhân chiếm hơn 50% trọng số sản phẩm (thiết kế lưới bố cục, lựa chọn hệ font chữ, sắp xếp phân cấp thị giác, viết lại toàn bộ tiêu đề), và tỷ lệ tổng thể của dự án là 40% AI — 60% con người. Con số đó là một ngưỡng tôi tự đặt ra để tự kiểm tra mình.

Nguyên tắc 5. Tôn trọng bản quyền và công sức của người khác

Các mô hình tạo ảnh được huấn luyện trên tác phẩm của hàng triệu nghệ sĩ thật, phần lớn không được xin phép hay đền bù thỏa đáng. Vấn đề này chưa có lời giải trọn vẹn, nhưng đó không phải lý do để người dùng phủi tay. Nguyên tắc: không sử dụng trực tiếp kết quả thô của AI cho mục đích thương mại, luôn chỉnh sửa sâu và tái cấu trúc để sản phẩm cuối mang dấu ấn nguyên bản của mình.

Minh chứng từ Bài 2: nguyên tắc này cũng áp dụng cho tài liệu học thuật. Mọi nguồn tôi sử dụng đều được trích dẫn đầy đủ theo chuẩn Harvard. Dùng ý tưởng của người khác mà không ghi nguồn là đạo văn, dù ý tưởng đó đến từ một cuốn sách hay từ một câu trả lời của AI.

Nguyên tắc 6. Bảo vệ dữ liệu cá nhân — của mình và của người khác

Mọi thứ gõ vào một chatbot đều rời khỏi máy tính của tôi và có thể được lưu lại. Tôi không đưa vào AI: thông tin cá nhân của người khác, dữ liệu khảo sát chưa được ẩn danh, hay tài liệu nội bộ chưa được phép công bố.

Minh chứng từ Bài 4: nhóm tôi thu về 28 phiếu khảo sát có chứa thông tin nhân khẩu học. Khi cần AI hỗ trợ diễn giải, tôi chỉ đưa vào các tỷ lệ phần trăm đã tổng hợp (ví dụ: "46,4% chi tiêu 2-5 triệu/tháng"), tuyệt đối không dán dữ liệu thô của từng người trả lời. Người tham gia khảo sát đã tin tưởng nhóm, nên việc giữ lời hứa ẩn danh là trách nhiệm tối thiểu.

IV. THANG BỐN MỨC ĐỘ: KHI NÀO ĐƯỢC, KHI NÀO KHÔNG?

Sáu nguyên tắc trên vẫn còn trừu tượng khi đứng trước một bài tập cụ thể. Vì vậy tôi tự dựng một thang bốn mức độ để ra quyết định nhanh, sử dụng chính cách phân mức 1-4 mà học phần đã dạy:

Mức	Vai trò của AI	Ví dụ cụ thể	Đánh giá
1	Trợ lý kỹ thuật	Sửa lỗi chính tả, định dạng bảng biểu, gợi ý từ đồng nghĩa.	Luôn được phép. Không động đến nội dung học thuật.
2	Gia sư giải thích	Hỏi lại khái niệm "bẫy thanh khoản" cho tới khi hiểu; nhờ tóm tắt để nắm cấu trúc một chương.	Được phép, với điều kiện tự đối chiếu lại giáo trình.
3	Cộng sự sáng tạo	Brainstorm ý tưởng chiến dịch, dựng dàn ý, tạo ảnh minh họa (như đã làm ở Bài 5).	Được phép có điều kiện: phải ghi nhận minh bạch và đóng góp cá nhân trên 50%.
4	Người làm bài thay	Nộp nguyên văn bài AI viết; để AI trả lời thay trong bài kiểm tra; tạo số liệu khảo sát giả.	Không bao giờ. Đây là gian lận học thuật, bất kể có bị phát hiện hay không.

Ranh giới nằm giữa Mức 3 và Mức 4, và nó không nằm ở việc dùng AI nhiều hay ít, mà ở chỗ: tư duy trong bài là của ai? Ở Mức 3, AI làm nhanh những gì tôi đã biết mình muốn. Ở Mức 4, AI quyết định thay tôi.

V. BA RỦI RO ĐÃ NHẬN RA QUA TỪNG BÀI

1. Kết quả càng "mượt" càng dễ tin nhầm

Prompt V3 ở Bài 3 cho kết quả trình bày như một chuyên gia. Chính sự trau chuốt đó làm người đọc mất cảnh giác. Văn phong tự tin không phải là bằng chứng cho tính chính xác.

2. Phụ thuộc dần mà không nhận ra

Khảo sát ở Bài 4 cho thấy 57,1% người dùng thỉnh thoảng thay đổi quyết định mua sắm sau khi tham khảo ý kiến từ AI. Nếu AI định hình được cả quyết định tiêu dùng, nó cũng hoàn toàn có thể định hình cách chúng ta suy nghĩ trong học tập.

3. Thiên vị trong dữ liệu huấn luyện

AI học từ dữ liệu sẵn có trên internet, vốn nghiêng về tiếng Anh và các nền kinh tế phát triển. Đối với ngành Kinh tế phát triển, đây là một điểm mù cần cảnh giác thường trực khi tham khảo các phân tích do AI đưa ra về bối cảnh Việt Nam.

VI. CAM KẾT CỦA TÔI

Tôi sử dụng AI như một cộng sự, không phải một người thay thế. Tôi chịu trách nhiệm hoàn toàn về mọi nội dung mang tên mình — kể cả những phần có sự hỗ trợ của AI. Khi AI sai, đó là lỗi của tôi vì đã không kiểm chứng, chứ không phải lỗi của công cụ.

Bộ nguyên tắc này không sinh ra để làm khó bản thân. Nó xuất phát từ mục tiêu nghề nghiệp tôi đã đặt ra: phát triển theo lĩnh vực đúng chuyên ngành và trở thành nhân sự có chuyên môn cao. Chuyên môn không thể đến từ những gì mình chưa từng thật sự hiểu. Một tấm bằng có được nhờ AI làm hộ sẽ sụp đổ ngay ở buổi phỏng vấn đầu tiên — và đáng ra nó nên sụp đổ.