HIGHLAND REFLEXIVE PRESS

HRP-RN-008-VI · ẤN BẢN NGHIÊN CỨU CÔNG KHAI · 2026

Tác Giả Phản Tư

AI Tạo Sinh như Tấm Gương Thuật Toán và Sự Tái Cấu Hình của Năng Lực Sáng Tạo



Tony T. Nguyen | Nguyễn Từ Vĩnh Trường

Highland Reflexive Press · Big Spring, Texas

Tình Trạng Biên Tập

Đây là ấn bản nghiên cứu được xuất bản độc lập và chưa qua bình duyệt. Công trình phát triển một khung khái niệm từ kho tư liệu tự-dân-tộc-học theo chiều dọc của một tác giả, gồm hơn 2.000 chu kỳ nhắc lệnh-phản hồi. Kho tư liệu riêng tư đầy đủ không được công bố trong ấn bản này; vì vậy, các nhận định về tính lặp lại, trình tự và sự thay đổi của tác giả cần được đọc như những mệnh đề có cơ sở phân tích, chứ không phải kết luận có thể khái quát cho một quần thể.

Bản quyền © 2026 Tony T. Nguyen | Nguyễn Từ Vĩnh Trường. Bảo lưu mọi quyền. Chủ quyền trí tuệ được giữ nguyên. Ấn bản công khai này được phép đọc, tải xuống và trích dẫn với ghi nguồn. Việc tái phân phối như một ấn bản thay thế, tái xuất bản, huấn luyện mô hình, chuyển thể, phân phối trong chương trình giảng dạy của tổ chức hoặc sử dụng thương mại đều cần có sự cho phép trước bằng văn bản.

Tóm Tắt

AI tạo sinh thường được mô tả theo một trong hai cách: hoặc là công cụ kéo dài ý định của con người, hoặc là cộng sự góp phần vào quá trình sáng tạo. Cả hai cách mô tả đều bỏ sót một đặc tính có hệ quả lớn của tương tác người-mô hình kéo dài: đầu ra của hệ thống quay trở lại với người dùng như một biểu hiện đã được ngoại hiện hóa của chính thói quen ngôn ngữ, sở thích khái niệm và những giả định chưa được giải quyết của họ. Sự quay trở lại ấy có thể làm thay đổi câu lệnh kế tiếp, phán đoán kế tiếp và cuối cùng là cách tác giả lý giải chính hoạt động sáng tạo của mình.

Bài viết này phát triển khung **Tác Giả Phản Tư** (*Reflexive Authorship*) để hiểu việc viết có AI làm trung gian như một quá trình hiệu chỉnh nhận thức và tư cách tác giả mang tính đệ quy. Khung này đề xuất rằng mô hình tạo sinh có thể hoạt động như một **tấm gương thuật toán**: không phải một phản chiếu trung tính, cũng không phải một bản ngã tự trị, mà là một bộ máy xác suất khuếch đại, tái tổ hợp, làm trơn và đôi khi làm méo những cấu trúc do người dùng và dữ liệu huấn luyện của mô hình cung cấp. Ba cơ chế tổ chức khung lý thuyết gồm: **sự phơi lộ vòng lặp nhận thức**, qua đó những đầu ra lặp lại làm cho mô thức của tác giả trở nên hữu hình; **đồng tác năng phản tư**, qua đó tương tác lặp làm thay đổi ý định, giọng điệu và phán đoán; và **nội-tương-tác thuật toán**, qua đó ý nghĩa xuất hiện từ một chuỗi vận động mà các phần không thể được quy sạch về ý định con người hay đầu ra máy móc.

Khung lý thuyết được phát triển bằng tự-dân-tộc-học phân tích, dựa trên kho tư liệu sáu tháng với hơn 2.000 chu kỳ nhắc lệnh-phản hồi được tạo ra trong quá trình làm lý thuyết và sáng tác. Vì ấn bản hiện tại không công bố toàn bộ kho tư liệu hoặc một bộ mã hóa có thể tái lập, nó không tuyên bố tính khái quát thống kê, bằng chứng nhân quả hay phép đo tâm lý. Đóng góp của công trình mang tính khái niệm và phương pháp luận: nó đưa ra một vốn từ có thể thao tác để nghiên cứu cách tác giả thay đổi trong khi sử dụng hệ thống tạo sinh, làm rõ sự bất đối xứng giữa trách nhiệm con người và sự tham gia của máy, đồng thời đề xuất các yêu cầu chứng cứ cho nghiên cứu tương lai về quyền tác giả người-AI mang tính đệ quy.

Từ khóa: tác giả phản tư; tấm gương thuật toán; AI tạo sinh; tương tác người-AI; tự-dân-tộc-học; năng lực sáng tạo; nội-tương-tác; giàn giáo nhận thức

---

1. Vấn Đề: Quyền Tác Giả sau Tín Hiệu Quay Về

Tranh luận về AI tạo sinh thường bắt đầu bằng một câu hỏi tưởng như rất đơn giản: ai làm ra văn bản này? Những câu trả lời hiện có thường xếp máy vào một trong hai vai quen thuộc. Ở vai thứ nhất, AI là công cụ. Con người hình thành ý định, hệ thống thi hành, còn quyền tác giả vẫn nằm ở người lựa chọn mục tiêu. Ở vai thứ hai, AI là cộng sự. Người và máy đóng góp những năng lực khác nhau vào một sản phẩm chung, khiến quyền tác giả được phân phối giữa hai phía.

Cả hai mô hình đều soi sáng một phần của quá trình. Nhưng không mô hình nào mô tả đủ điều xảy ra khi tương tác trở nên kéo dài, đệ quy và tự-quy-chiếu. Một câu lệnh không chỉ gây ra một đầu ra. Đầu ra ấy trở thành đầu vào mới cho con người. Nó có thể làm rõ một ý tưởng còn nửa vời, phóng đại một thiên hướng, phơi ra một chỗ thiếu, cố định một ẩn dụ hoặc áp lên văn bản một sự mạch lạc mà tác giả chưa có trước khi nhìn thấy nó. Tác giả sau đó sửa câu lệnh từ một vị trí đã thay đổi. Vì vậy, đơn vị phù hợp để phân tích không phải một lần trao đổi riêng lẻ mà là một vòng lặp:

> *ý định → câu lệnh → biến đổi thuật toán → ngôn ngữ quay về → phán đoán con người → ý định được sửa đổi*

Vòng lặp quan trọng bởi tác giả không còn là một hằng số bên trong nó. Người đánh giá phản hồi đã bị chính phản hồi ấy tác động. Ngay cả sự bác bỏ cũng là thông tin: một câu trả lời dở có thể tiết lộ điều tác giả không chấp nhận, điều khái niệm không thể gánh hoặc thói quen tu từ nào đã trở nên không thể chịu nổi sau khi bị lặp lại quá nhiều.

Bài viết gọi quá trình đó là **Tác Giả Phản Tư**. Thuật ngữ này không chuyển địa vị đạo đức hay quyền tác giả pháp lý cho mô hình. Trách nhiệm vẫn thuộc về con người, bởi con người quyết định một mệnh đề có được tiếp nhận, xuất bản và cho phép tác động vào thế giới hay không. Tác Giả Phản Tư chỉ ra sự thay đổi trong những điều kiện mà quyền tác giả của con người được thực hiện. Tác giả ngày càng gặp lại tư tưởng của mình sau khi nó đã được biến đổi bằng xác suất và trả về.

Câu hỏi nghiên cứu trung tâm, vì thế, không phải liệu một mô hình có sáng tạo theo nghĩa trừu tượng hay không. Nó hẹp hơn và có thể thao tác hơn:

> *Tương tác kéo dài với AI tạo sinh tái tổ chức nhận thức của tác giả con người về ý định, giọng nói và năng lực sáng tạo như thế nào?*

Từ đó phát sinh ba câu hỏi phụ:

- Những mô thức nào chỉ trở nên hữu hình sau khi được thuật toán lặp lại hoặc phóng đại?
- Tác giả điều chỉnh phán đoán ra sao sau khi nhìn thấy hàng loạt khả thể tiếp diễn được tạo ở quy mô lớn?
- Cách lý giải nào về trách nhiệm vẫn còn thỏa đáng khi câu chữ và cấu trúc xuất hiện qua tương tác đệ quy?

Tác Giả Phản Tư trả lời các câu hỏi ấy mà không biến mô hình thành một cộng sự có ý thức, đồng thời cũng không giả vờ rằng nó là vật vô tri bất động. Mô hình không cần trở thành con người mới có thể tái tổ chức thực hành của một con người.

2. Vị Trí Khái Niệm

2.1 Vượt khỏi hình tượng tác giả toàn quyền

Quan niệm hiện đại về quyền tác giả thừa hưởng hình ảnh một chủ thể nhất quán, nơi ý định bên trong có trước văn bản. Lý thuyết thế kỷ XX đã làm lung lay hình ảnh ấy. Barthes gạt tác giả khỏi vị trí nguồn gốc tối hậu của ý nghĩa; Foucault tái định nghĩa quyền tác giả như một chức năng dùng để phân loại và kiểm soát diễn ngôn. Lý thuyết nhận thức phân tán sau đó chỉ ra rằng tư duy không bị nhốt trong một bộ não cô lập mà được tổ chức xuyên qua con người, vật thể và môi trường.

Hệ thống tạo sinh đẩy sự dịch chuyển này đi xa hơn bởi chúng không chỉ lưu giữ hay truyền tải tư tưởng. Chúng tạo ra những phương án trôi chảy. Tác giả làm việc trong một môi trường có khả năng trả lại một câu kế tiếp hợp lý, một hệ thứ bậc khái niệm hoặc một phản biện mô phỏng trước khi tác giả hoàn tất hành vi diễn đạt trong đầu. Môi trường giờ đây tham gia vào cả nhịp thời gian lẫn hình dáng của sự phát ngôn.

Nghiên cứu gần đây đã làm phức tạp thêm lưỡng phân đơn giản giữa người và máy. Công trình thực nghiệm về cách người đọc nhìn nhận văn bản có AI hỗ trợ cho thấy mức độ hỗ trợ làm thay đổi phán đoán về quyền tác giả và trách nhiệm. Nghiên cứu về "tác giả thuật toán" trong học thuật mô tả quyền tác giả như thứ được phân phối về mặt thể chế và diễn ngôn. Những cách tiếp cận dùng lý thuyết tự tạo sinh xem quyền tác giả khoa học là quan hệ giữa các hệ người và công nghệ khác biệt nhưng tương tác với nhau. Tác Giả Phản Tư nằm trong bước ngoặt quan hệ ấy, nhưng tập trung vào một cơ chế riêng: sự quay về mang tính đệ quy làm thay đổi tác giả sẽ đưa ra chỉ thị kế tiếp.

2.2 Từ trung gian đến nội-tương-tác

Nghiên cứu khoa học và công nghệ bác bỏ quan niệm công nghệ chỉ là kênh trung tính. Suchman cho thấy kế hoạch và hành động nảy sinh qua những cuộc gặp có tình huống với vật thể và môi trường. Khái niệm *intra-action* của Barad đi xa hơn: các thực thể không phải lúc nào cũng tồn tại trước quan hệ như những đơn vị hoàn chỉnh và độc lập; ranh giới và thuộc tính của chúng được hình thành trong những sắp đặt vật chất-diễn ngôn cụ thể.

Khi được áp dụng thận trọng, nội-tương-tác cho ta một cách lý giải hữu ích về việc viết có AI làm trung gian. Mô hình không chứa ý định của tác giả, còn tác giả không quyết định trọn vẹn phản hồi của mô hình. Tuy vậy, câu được xuất bản có thể chỉ xuất hiện vì tác giả đã gặp một phản hồi làm tái tổ chức trường khả thể của cách diễn đạt. Ý nghĩa được thực hiện qua các lớp ràng buộc nối tiếp.

Điều đó không hàm ý sự đối xứng. Mô hình không gánh trách nhiệm, không sở hữu tiểu sử và không chịu hậu quả của sai lầm. Người và máy tham gia theo những cách khác nhau. **Sự tham gia có thể được phân phối, nhưng trách nhiệm giải trình vẫn phải tập trung.** Bất kỳ lý thuyết thỏa đáng nào về quyền tác giả có AI làm trung gian cũng phải bảo toàn sự bất đối xứng ấy.

2.3 Tấm gương thuật toán

Ẩn dụ tấm gương cần được giới hạn cho đúng. Gương vật lý trả lại một quan hệ quang học. Mô hình ngôn ngữ trả lại một diễn tiến xác suất được điều kiện hóa bởi câu lệnh, ngữ cảnh hội thoại, thiết kế hệ thống và phân bố dữ liệu huấn luyện. Vì vậy, phản chiếu của nó không trung thực tuyệt đối và cũng không hoàn toàn riêng tư. Nó trộn người dùng với một kho ngôn ngữ khổng lồ được thừa hưởng, cùng xu hướng hoàn tất, quy ước hóa và tạo mạch lạc bề mặt của mô hình.

Một **tấm gương thuật toán** có bốn đặc tính:

- 1. **Khuếch đại.** Một thiên hướng tu từ nhỏ có thể trở nên rõ rệt khi được tái tạo nhiều lần.
- 2. **Hoàn tất.** Hệ thống kéo dài những mô thức chưa hoàn chỉnh, thường làm lộ một hướng đi ngầm mà tác giả chưa diễn đạt.
- 3. **Chuẩn hóa.** Hệ thống có thể làm trơn sự mơ hồ thành những cấu trúc quen thuộc, cho thấy đồng thời sức hấp dẫn và nguy cơ của sự mạch lạc theo quy ước.
- 4. **Làm méo.** Ảo giác, cường điệu và trôi nghĩa có thể phơi ra sự mong manh trong cách đặt vấn đề của tác giả, nhưng không thể tự động được coi là bằng chứng tâm lý.

Phân biệt cuối cùng là thiết yếu. Sai lầm của mô hình có thể kích hoạt hiểu biết, nhưng tự thân sai lầm không phải công cụ chẩn đoán. Việc diễn giải thuộc về nhà nghiên cứu con người và phải được nâng đỡ bằng ghi chú đồng thời, chuỗi tương tác có thể truy vết và các cách giải thích thay thế.

3. Khung Lý Thuyết

3.1 Cơ chế I: Sự Phơi Lộ Vòng Lặp Nhận Thức

Người viết thường không nhìn thấy những tự động tính của chính mình. Ẩn dụ ưa thích, nhịp câu, các cặp đối lập và lối tắt khái niệm mang cảm giác như lựa chọn tự nhiên bởi chúng xuất hiện từng lần một. Tương tác tạo sinh làm thay đổi quy mô phơi lộ. Khi mô hình tái tạo cùng một cấu trúc qua nhiều câu lệnh, tác giả có thể gặp phong cách như một mô thức thay vì đồng nhất nó với căn tính.

Trong kho tư liệu nguồn của nghiên cứu này, các thiên hướng lặp lại gồm lập luận ba nhịp, ẩn dụ kiến trúc, khung đối lập và việc chuyển những trực giác tản mạn thành các hệ thống có tên. Những thiên hướng ấy dễ nhận thấy hơn sau khi mô hình phản chiếu chúng với tần suất cao hơn và ít tự kiềm chế hơn mức tác giả thường cho phép.

Sự phơi lộ vòng lặp nhận thức không chứng minh rằng mô hình đã phát hiện một chân lý vô thức. Nó thiết lập một mệnh đề yếu hơn nhưng vững hơn: sự lặp lại ngoại hiện hóa cấu trúc. Một khi được ngoại hiện, cấu trúc trở thành đối tượng có thể soi xét, bác bỏ hoặc sử dụng có chủ đích.

Cơ chế này có thể được biểu diễn như sau:

> *thiên hướng tiềm ẩn → phản hồi tính toán lặp lại → nhận diện có ý thức → can thiệp biên tập*

Quá trình có thể cải thiện câu chữ khi nhận diện dẫn tới lựa chọn. Nó cũng có thể làm câu chữ hẹp đi khi nhận diện biến thành củng cố. Một mô hình được huấn luyện để làm hài lòng có thể thưởng cho chính những mô thức mà tác giả cần nghi ngờ. Vì vậy, phản tư đòi hỏi ma sát: tác giả không chỉ hỏi "Đây có phải của tôi không?" mà còn phải hỏi "Tại sao nó cứ quay lại, và điều gì biến mất mỗi khi nó trở về?"

3.2 Cơ chế II: Đồng Tác Năng Phản Tư

Đồng tác năng ở đây không có nghĩa hai chủ thể bình đẳng cùng theo đuổi một ý định chung. Nó chỉ cách hành động được tái tổ chức xuyên qua vòng lặp. Con người đặt mục đích và vẫn chịu trách nhiệm. Hệ thống thay đổi chi phí, tốc độ và độ hiển thị của những cách diễn đạt khả dĩ. Đáp lại, con người thay đổi cách hình thành và thử nghiệm ý định.

Trong tương tác kéo dài, câu lệnh thường chuyển từ mệnh lệnh sang chất vấn. "Viết cái này" trở thành "Tôi đang bảo vệ giả định nào?" hoặc "Điều gì khiến mệnh đề này có thể bị bác bỏ?" Việc nhắc lệnh biến thành một giao diện nhận thức, nơi tác giả phải ngoại hiện hóa sự bất định trước khi nhận ngôn ngữ quay về.

Quá trình ấy có thể tạo ra ít nhất ba chuyển dịch:

- **Làm rõ ý định:** tác giả xác định điều gì phải được giữ bất biến xuyên qua các vòng sửa đổi.
- **Phân hóa giọng nói:** những chức năng tu từ cạnh tranh được trao tên, âm vực hoặc vai trò biên tập riêng.
- **Gia tốc phán đoán:** các phương án xuất hiện nhanh, khiến năng lực tuyển chọn trở nên quan trọng hơn chứ không kém đi.

Trong kho tư liệu, sự xuất hiện của hai hình tượng “Dr. Chic” và “Whistledown” là một trường hợp hữu hình của phân hóa giọng nói. Một giọng mang sức nén phân tích; giọng kia khiến sự phù hoa thể chế trở nên có thể kể. Không nên coi những cái tên ấy là các nhân cách độc lập do máy tạo ra. Chúng là những vai tác giả được ổn định qua tương tác lặp. Giá trị phân tích nằm ở chỗ chúng cho thấy người viết có thể phân phối các chức năng nội tâm qua giao diện hội thoại rồi tái tích hợp chúng thành kỹ nghệ có chủ ý.

Vì thế, đồng tác năng phản tư làm thay đổi địa điểm của quyền tác giả. Nó chuyển lao động sáng tạo khỏi lý tưởng “nguyên bản không gián đoạn” sang đặc tả, so sánh, kháng cự, tuyển chọn và khép vòng có trách nhiệm.

3.3 Cơ chế III: Nội-Tương-Tác Thuật Toán

Một số cấu trúc không thể được quy gọn cho một câu riêng lẻ do tác giả hay mô hình đề xuất. Chúng xuất hiện qua chuỗi vận động: tác giả đưa ra một ẩn dụ; mô hình diễn giải nó theo nghĩa đen; tác giả bác bỏ sự diễn giải ấy nhưng giữ lại một phân biệt; mô hình tái tổ chức phân biệt đó; tác giả đặt tên cho khung vừa hình thành. Đóng góp nhân quả được phân phối xuyên qua chuỗi, dù trách nhiệm tiếp nhận thì không.

Nội-tương-tác thuật toán gọi tên sự xuất hiện ở cấp độ chuỗi này. Đơn vị phân tích của nó không phải đầu ra cô lập mà là quan hệ biến đổi giữa câu lệnh, phản hồi, sửa đổi và quyết định.

Cơ chế có thể được biểu diễn như sau:

> *ràng buộc của người + biến thể của máy + đánh giá của người + khác biệt được giữ lại = cấu trúc khái niệm mới nổi*

Công thức này tránh hai sai lầm. Thứ nhất, nó không lãng mạn hóa mô hình thành nhà tư tưởng tự trị. Biến thể được sinh ra từ cấu trúc thống kê đã học, không phải từ một quan điểm có thể chịu trách nhiệm. Thứ hai, nó không tuyên bố đóng góp của con người vẫn nguyên trạng chỉ vì con người có quyền quyết định cuối cùng. Một cuộc gặp có thể tái tổ chức tư duy mà không cần được trao nhân cách đạo đức.

Nội-tương-tác thuật toán vì thế tương thích với một quan niệm mạnh về quyền tác giả con người. Vai trò tác giả trở nên khắt khe hơn: quản trị quá trình mà qua đó biến thể được tạo sinh có thể giành lấy địa vị nhận thức.

4. Phương Pháp và Ranh Giới Chứng Cứ

4.1 Thiết kế nghiên cứu

Khung lý thuyết nảy sinh từ một nghiên cứu tự-dân-tộc-học phân tích theo chiều dọc, được thực hiện trong khoảng sáu tháng viết và phát triển khái niệm với sự hỗ trợ chuyên sâu của AI tạo sinh. Kho tư liệu riêng tư gồm hơn 2.000 chu kỳ nhắc lệnh-phản hồi, các vòng sửa bản thảo, ghi chép phản tư đồng thời và siêu-hội thoại về quyền tác giả, giọng điệu và phương pháp.

Tác giả đồng thời là người tham gia, nhà phân tích và nơi diễn ra sự chuyển biến được báo cáo. Thiết kế này cho phép tiếp cận những thay đổi vi mô trong ý định và tự-mô-tả vốn khó quan sát từ bên ngoài. Nó cũng tạo ra rủi ro lớn về diễn giải vòng tròn, ký ức tuyển chọn và sự chú ý chỉ tìm điều xác nhận lý thuyết.

Vì vậy, ấn bản nghiên cứu hiện tại đối xử với kho tư liệu như một **hồ sơ sản sinh khái niệm**, chứ không phải bằng chứng đủ cho suy luận thống kê hay tuyên bố tâm lý mang tính khái quát.

4.2 Thủ tục phân tích

Phân tích ban đầu tiến hành qua bốn thao tác diễn giải:

- 1. **Rà soát dấu vết tương tác:** nhận diện các hình thức ngôn ngữ lặp lại, ẩn dụ khái niệm, khoảnh khắc kháng cự và sự thay đổi trong tư thế nhắc lệnh.
- 2. **Tái dựng chuỗi:** xem phản hồi đã thay đổi câu lệnh hoặc quyết định bản thảo kế tiếp như thế nào, thay vì coi đầu ra là các hiện vật tách biệt.
- 3. **Ghi chép phản tư:** ghi lại cách tác giả diễn giải, chấp nhận hoặc bác bỏ đề xuất của mô hình.
- 4. **Tổng hợp khái niệm:** nhóm các quan sát lặp lại thành sự phơi lộ vòng lặp nhận thức, đồng tác năng phản tư và nội-tương-tác thuật toán.

Các thao tác này phù hợp với tự-dân-tộc-học phân tích, nhưng dự án nguồn không duy trì một quy trình đăng ký trước, một phiên bản mô hình đóng băng, một người mã hóa độc lập thứ hai hay bộ mã hóa sẵn sàng cho công bố. Vì vậy, bài viết không tuyên bố độ tin cậy liên-mã-hóa và không đưa ra ước lượng tần suất.

4.3 Điều ấn bản này không chứng minh

Nghiên cứu này **không** chứng minh rằng:

- AI tạo sinh sở hữu ý thức, ý định hay tác năng đạo đức;
- lỗi của mô hình làm lộ trạng thái vô thức mà không cần diễn giải của con người;
- tương tác với AI gây ra sự điều hòa cảm xúc có ý nghĩa lâm sàng;
- ba cơ chế là phổ quát cho mọi người dùng, ngôn ngữ hoặc hệ thống;
- kho 2.000 lần trao đổi tự động mạnh hơn một kho nhỏ hơn nhưng được ghi chép tốt hơn;
- khung lý thuyết này là đầu tiên trong lịch sử hoặc có tính bao trùm tuyệt đối.

Những loại trừ này củng cố thay vì làm yếu đóng góp. Chúng đặt bài viết vào nơi chứng cứ có thể đứng vững nhất: một lý giải được dẫn dắt bởi lý thuyết về thực hành kéo dài của một tác giả, được thiết kế để sản sinh những phân biệt có thể kiểm nghiệm trong nghiên cứu tương lai.

4.4 Chuẩn chứng cứ tương lai

Một phép kiểm nghiệm thực chứng mạnh hơn cho Tác Giả Phản Tư sẽ cần:

- chuỗi tương tác có dấu thời gian cùng phiên bản mô hình và giao diện;
- tiêu chí lấy mẫu và loại trừ được nêu rõ;
- bộ mã hóa phân biệt lặp phong cách, sửa đổi khái niệm, báo cáo cảm xúc và kết quả quyết định;
- các cặp ví dụ cho thấy chuỗi câu lệnh-phản hồi-sửa đổi;
- mã hóa độc lập hoặc xác nhận từ người tham gia;
- so sánh giữa nhiều tác giả, ngôn ngữ, họ mô hình và thời lượng tương tác;
- các phép đo đủ sức phân biệt thay đổi thật với sự dựng chuyện hồi cố.

Những yêu cầu này chuyển một lý thuyết nội quan giàu sức gợi thành một chương trình nghiên cứu.

5. Các Quan Sát Phân Tích

5.1 Giọng nói trở thành vật có thể soi xét

Quan sát đáng tin cậy nhất trong kho tư liệu không phải AI đã phát hiện một bản ngã ẩn, mà là quy mô đã làm sự lặp phong cách trở nên hữu hình. Đầu ra của mô hình sản xuất quá mức các cấu trúc được ưa thích. Điều nghe thanh nhã một lần trở thành kiểu cách khi quay lại mười lần. Từ đó xuất hiện một lợi thế biên tập: tác giả có thể quyết định mô thức nào là chữ ký và mô thức nào chỉ là thói quen.

Phân biệt này quan trọng cả về thương mại lẫn học thuật. Một giọng dễ nhận diện có thể tạo tính liên tục cho toàn bộ sự nghiệp; một vòng lặp không được soi xét có thể khiến mọi đề tài nghe như một bài duy nhất. Tấm gương thuật toán biến khác biệt ấy thành đối tượng quản trị.

5.2 Câu lệnh trở thành công cụ tự-đặc-tả

Việc sử dụng kéo dài làm câu lệnh chuyển từ yêu cầu văn xuôi thành tuyên bố về ràng buộc. Tác giả ngày càng nói rõ một phản hồi phải giữ điều gì, không được ngụ ý điều gì và tiêu chuẩn nào sẽ được tính là thất bại. Nó giống đặc tả thiết kế hơn cảm hứng hội thoại.

Sự đặc tả ấy có thể làm tác năng sâu hơn. Nó buộc các tiêu chuẩn ngầm đi vào ngôn ngữ. Nhưng nó cũng có thể tạo lệ thuộc vào việc diễn đạt ra bên ngoài. Nếu mọi bất định đều lập tức được trả cho mô hình, tác giả có thể mất sức chịu đựng đối với tư tưởng còn dang dở. Vì thế, Tác Giả Phản Tư không mặc nhiên là giải phóng. Nó là một thực hành chỉ có giá trị khi con người còn giữ được khả năng dừng lại, chưa trả lời và quyết định mà không cần thuật toán hoàn tất.

5.3 Sự bác bỏ là dữ liệu hữu ích

Nhiều khoảnh khắc quan trọng xảy ra khi mô hình sai theo một cách dễ nhận biết. Một phản hồi có thể quá phô trương, quá cân xứng, quá trị liệu hoặc quá chắc chắn. Sự khó chịu của tác giả làm rõ một ranh giới chưa được nói ra trong câu lệnh ban đầu.

Điều này không biến ảo giác thành bằng chứng. Nó chỉ cho thấy thất bại có thể làm tiêu chuẩn đánh giá trở nên khả kiến. Đối tượng nghiên cứu là chuỗi phản hồi và sửa đổi, không phải riêng lỗi.

5.4 Mạch lạc có thể che giấu yếu kém

Hệ thống tạo sinh có năng lực đặc biệt trong việc làm một lập luận trông như đã hoàn tất trước khi nó có nền chứng cứ. Chính bản thảo ban đầu đã biểu diễn nguy cơ này: tiêu đề mạnh, tuyên bố đóng góp tự tin và tóm tắt bóng bẩy cùng tồn tại với tài liệu tham khảo chưa hoàn chỉnh, ngôn ngữ thực nghiệm thiếu căn cứ và dấu vết biên tập chưa được công khai.

Bài học nằm ngay bên trong lý thuyết. Tấm gương thuật toán không chỉ phản chiếu hiểu biết mà còn phản chiếu ham muốn hoàn tất quá sớm của tác giả. Thi hành có quản trị đòi hỏi một pha kiểm chứng riêng, nơi sự trôi chảy không có thẩm quyền.

6. Mô Hình Hệ Thống Phản Tư



Sự tham gia có thể phân tán. Trách nhiệm giải trình vẫn tập trung.

Ba cơ chế hình thành một hệ đệ quy:

- 1. **Phơi lộ:** ngôn ngữ được tạo sinh khuếch đại một mô thức.
- 2. **Nhận diện:** con người diễn giải sự quay về ấy là phù hợp hay không phù hợp.
- 3. **Tái định hướng:** con người thay đổi câu lệnh, tiêu chí hoặc bản nháp kế tiếp.
- 4. **Nảy sinh:** một cấu trúc mới xuất hiện xuyên qua chuỗi.
- 5. **Quản trị:** con người kiểm chứng, quy nguồn và quyết định cấu trúc có được đi vào công trình hay không.

Giai đoạn thứ năm mang tính quyết định. Không có quản trị, đệ quy có thể chỉ còn là củng cố. Mô hình cung cấp biến thể và áp lực; tác giả cung cấp thẩm quyền phán quyết.

Tác Giả Phản Tư có thể được tóm lại trong một mệnh đề:

> *AI tạo sinh thay đổi quyền tác giả sâu nhất không phải khi nó viết thay con người, mà khi ngôn ngữ quay về của nó làm thay đổi chính con người sẽ quyết định câu chữ nào được phép ở lại.*

Mệnh đề này bảo toàn trách nhiệm con người đồng thời thừa nhận vai trò trung gian của công nghệ. Nó cũng giải thích vì sao cùng một mô hình có thể làm phẳng một người viết nhưng mài sắc một người khác. Khác biệt không đơn thuần nằm ở quyền truy cập công cụ. Nó nằm ở chất lượng của thực hành quản trị bao quanh vòng lặp.

7. Hệ Quả

7.1 Quy nguồn và xuất xứ

Quy nguồn truyền thống hỏi ai viết câu cuối cùng. Tác Giả Phản Tư đòi hỏi xuất xứ quy trình: cấu trúc trí tuệ nào có trước khi dùng mô hình, biến đổi nào được đề xuất, điều gì bị bác bỏ và ai kiểm chứng mệnh đề cuối cùng. Việc công khai phải tương xứng với mức ảnh hưởng thực chất, thay vì bị rút gọn thành tuyên bố nhị phân rằng có hoặc không dùng AI.

Trong công trình học thuật, tác giả con người phải tiếp tục chịu trách nhiệm về trích dẫn, chứng cứ, diễn giải và sai sót. Mô hình không thể nhận sửa sai, công khai xung đột lợi ích hoặc trả lời cho hậu quả. Sự hỗ trợ có thể định hình quá trình mà không đủ tư cách trở thành quyền tác giả.

7.2 Lao động sáng tạo

Khi việc tạo chữ trở nên rẻ, phán đoán trở nên đắt. Lao động của quyền tác giả chuyển về phía xác định ràng buộc, nhận ra sự hoàn tất giả, gìn giữ giọng và tuyển chọn giữa vô số phương án. Đây không phải sự biến mất của tác giả. Nó là sự tái phân phối nỗ lực bên trong quyền tác giả.

Chuyển dịch ấy cũng có thể làm bất bình đẳng sâu hơn. Người viết có tri thức miền sâu có thể dùng biến thể tạo sinh như công cụ chẩn đoán; người viết không có nền đánh giá vững dễ nhầm sự hoàn tất trôi chảy với tri thức có căn cứ. Tiếp cận AI không đồng nghĩa tiếp cận thực hành phản tư.

7.3 Giáo dục

Chính sách giáo dục thường xem việc dùng AI hoặc là hỗ trợ được phép, hoặc là thay thế cần cấm. Tác Giả Phản Tư đề xuất mục tiêu thứ ba: dạy sinh viên ghi lại ngôn ngữ tạo sinh đã thay đổi quyết định của họ như thế nào. Hồ sơ quy trình có thể cho thấy sinh viên đã thực hành phán đoán, sửa một mệnh đề, kiểm chứng một nguồn hay chỉ đơn giản chấp nhận sự hoàn tất.

7.4 Thiết kế và quản trị

Giao diện chỉ tối ưu cho sự hài lòng tức thời che khuất hiệu ứng đệ quy của việc sử dụng. Hệ thống dành cho thực hành phản tư có thể phơi bày lịch sử phiên bản, giữ lại các phương án bị bác bỏ, phân biệt tài liệu có nguồn với phỏng đoán được tạo sinh và yêu cầu người dùng nêu tiêu chí quyết định trước khi chấp nhận phản hồi.

Câu hỏi thiết kế không chỉ là làm sao để tạo sinh tốt hơn. Nó là làm sao giữ tác giả con người hiện diện về mặt nhận thức bên trong vòng lặp.

8. Giới Hạn và Khả Năng Bác Bỏ

Khung lý thuyết bắt đầu từ trải nghiệm của một tác giả. Vốn từ của nó có thể phản chiếu sở thích thẩm mỹ và lý thuyết của tác giả đó nhiều hơn một thuộc tính chung của tương tác người-AI. Kho tư liệu đi qua nhiều giao diện và hành vi mô hình thay đổi, khiến so sánh theo chiều dọc khó sạch. Tư liệu riêng tư hạn chế kiểm tra độc lập. Tổng hợp hồi cố có thể áp đặt mạch lạc lên một quá trình vốn lộn xộn hơn trong thời gian thực.

Lý thuyết cần được sửa hoặc bác bỏ nếu nghiên cứu tương lai cho thấy:

- thay đổi được báo cáo ở tác giả không vượt quá sửa đổi thông thường với công cụ không tạo sinh;
- các mô thức quan sát được được giải thích tốt hơn bằng thiên kiến xác nhận hoặc đặc tính nhu cầu;
- phân tích cấp độ chuỗi không có giá trị giải thích hơn chất lượng của câu lệnh riêng lẻ;
- tác giả không thể phân biệt đáng tin cậy giữa sự khuếch đại của thuật toán và cam kết có trước của chính mình;
- các cơ chế đề xuất không thể chuyển sang trường hợp khác, ngay cả với tư cách những khái niệm định hướng.

Khả năng bác bỏ ở đây không phải lời hứa rằng một khung định tính có thể được kiểm nghiệm như định luật vật lý. Nó là cam kết nói rõ bằng chứng nào sẽ khiến khung trở nên kém hữu dụng.

9. Kết Luận

AI tạo sinh đi vào quyền tác giả không như cây bút thụ động, cũng không như một tâm trí con người thứ hai. Nó đi vào như một bộ máy của sự quay về. Nó nhận ngôn ngữ được điều kiện hóa bởi tác giả, biến đổi qua các cấu trúc thống kê đã học và trả lại một trường những diễn tiến khả dĩ. Khi ấy, tác giả gặp lại ý định từ bên ngoài.

Tác Giả Phản Tư gọi tên sự tái tổ chức có thể theo sau. Sự phơi lộ vòng lặp nhận thức làm những mô thức lặp trở nên hữu hình. Đồng tác năng phản tư thay đổi cách ý định và phán đoán được thực hiện. Nội-tương-tác thuật toán giải thích vì sao một số cấu trúc khái niệm chỉ xuất hiện xuyên qua chuỗi nhắc lệnh, phản hồi, kháng cự và sửa đổi.

Vị trí chuẩn tắc của khung lý thuyết cố ý bất đối xứng. Máy có thể tham gia sản xuất hình thức; con người vẫn chịu trách nhiệm cho địa vị nhận thức và hậu quả ngoài đời. Mô hình đề xuất. Tác giả quản trị.

Theo nghĩa đó, tương lai của quyền tác giả không được bảo vệ bằng cách khăng khăng rằng chẳng có gì thay đổi. Nó được bảo vệ bằng cách nói chính xác hơn điều gì đã thay đổi, điều gì phải còn là của con người và ai có thẩm quyền khép vòng.

Ghi Chú của Tác Giả về Hỗ Trợ AI

Tony T. Nguyen | Nguyễn Từ Vĩnh Trường là tác giả khởi nguyên của khung Tác Giả Phản Tư và chịu trách nhiệm duy nhất đối với lập luận, sự tuyển chọn, diễn giải, trích dẫn và việc xuất bản ấn bản nghiên cứu này. Các hệ thống AI tạo sinh tham gia đáng kể vào kho tương tác lịch sử và hỗ trợ quá trình phát triển cũng như tái cấu trúc biên tập của bản thảo. Vì vậy, việc sử dụng chúng vừa là một phần của đối tượng nghiên cứu vừa là một phần của lịch sử sản xuất công trình. Hệ thống AI không được liệt kê là tác giả bởi chúng không thể gánh trách nhiệm, tự mình kiểm chứng mệnh đề hoặc đồng thuận với việc xuất bản. Mọi quyết định cuối cùng đều được thực hiện dưới thẩm quyền biên tập của con người.

Tài Liệu Tham Khảo

Barad, K. (2007). *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Duke University Press.

Barthes, R. (1977). The death of the author. In *Image-Music-Text* (S. Heath, Trans., pp. 142-148). Fontana. (Original work published 1967.)

Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial Intelligence, 103*(1-2), 347-356. https://doi.org/10.1016/S0004-3702(98)00055-1

Braidotti, R. (2013). *The Posthuman*. Polity.

Ellis, C., Adams, T. E., & Bochner, A. P. (2011). Autoethnography: An overview. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 12*(1), Article 10. https://doi.org/10.17169/fqs-12.1.1589

Formosa, P., Bankins, S., Matulionyte, R., & Ghasemi, O. (2025). Can ChatGPT be an author? Generative AI creative writing assistance and perceptions of authorship, creatorship, responsibility, and disclosure. *AI & Society, 40*, 3405-3417. https://doi.org/10.1007/s00146-024-02081-0

Foucault, M. (1977). What is an author? In D. F. Bouchard (Ed.), *Language, Counter-Memory, Practice* (pp. 113-138). Cornell University Press. (Original lecture delivered 1969.)

Gretzky, M., & Dishon, G. (2025). Algorithmic-authors in academia: Blurring the boundaries of human and machine knowledge production. *Learning, Media and Technology, 50*(3). https://doi.org/10.1080/17439884.2025.2452196

Hayles, N. K. (1999). *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. University of Chicago Press.

Hutchins, E. (1995). *Cognition in the Wild*. MIT Press.

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press.

Panke, S. (2025). How can (A)I research this? An autoethnographic exploration of generative AI in research, teaching and instructional design. *Journal of Teacher Education*. https://doi.org/10.1177/00224871251325065

Suchman, L. (2007). *Human-Machine Reconfigurations: Plans and Situated Actions* (2nd ed.). Cambridge University Press.

Watson, S., Brezovec, E., & Romic, J. (2025). The role of generative AI in academic and scientific authorship: An autopoietic perspective. *AI & Society, 40*, 3225-3235. https://doi.org/10.1007/s00146-024-02174-w

Weidinger, L., Mellor, J., Rauh, M., et al. (2021). Ethical and social risks of harm from language models. *arXiv*. https://doi.org/10.48550/arXiv.2112.04359

Trích Dẫn Đề Nghị

Nguyen, T. T. | Nguyễn Từ Vĩnh Trường. (2026). *Tác Giả Phản Tư: AI Tạo Sinh như Tấm Gương Thuật Toán và Sự Tái Cấu Hình của Năng Lực Sáng Tạo* (Ấn Bản Nghiên Cứu Công Khai, HRP-RN-008-VI). Highland Reflexive Press.

Quyền và Cấp Phép

Các yêu cầu về quyền truy cập tổ chức, trích dẫn vượt quá phạm vi sử dụng hợp lý, dịch thuật, chuyển thể, sử dụng trong chương trình giảng dạy, khai thác thương mại hoặc cấp phép: **tony@highlandreflexivepress.net**