TRÍCH TUYỂN CÔNG KHAI · TIẾNG VIỆT

THỊ TRƯỜNG TỰ TẠO RA NƯỚC ĐI CỦA CHÍNH NÓ

Quyền lực nhân quả khi AI vừa dự báo, vừa biến luôn tiên đoán của nó thành sự thật

---

Tony T. Nguyen | Nguyễn Từ Vĩnh Trường

Highland Reflexive Press · HRP–RN–011 · Big Spring, Texas · 2026

NGHIÊN CỨU ĐỘC LẬP · CHƯA QUA BÌNH DUYỆT

Trích tuyển tiếng Việt được phát hành miễn phí.

Tôi đến và phát hiện ra AI rất thích đá banh: coi bộ nó thường xuyên vừa đóng vai cầu thủ, vừa kiêm luôn trọng tài.

Dẫn nhập — Khi lời tiên tri được cấp ngân sách

Một mô hình dự báo bình thường đứng ngoài sự kiện. Nó nhìn, tính, nói, rồi chờ thế giới trả lời. Nhưng một mô hình được nối với vốn, quyền đặt lệnh, định tuyến thanh khoản và cơ chế tự tăng quyền thì không còn đứng ngoài nữa. Nó đọc thị trường bằng một tay và viết lại thị trường bằng tay kia.

Rắc rối bắt đầu ở chỗ hai bàn tay ấy dùng chung một bảng điểm.

Mô hình dự báo giá sẽ tăng. Hệ thống mua. Lệnh mua đẩy giá lên. Giá lên được ghi nhận là bằng chứng rằng mô hình đã đúng. Điểm số tăng. Hạn mức tăng. Lần kế tiếp, cùng một dự báo có thể điều động nhiều tiền hơn và tạo ra một chuyển động lớn hơn. Một vòng lặp như thế có thể rất có lãi, rất thuyết phục và hoàn toàn sai về điều nó tưởng mình đã biết.

Đây không chỉ là lỗi thống kê. Đây là lỗi hiến pháp của một hệ thống: một chủ thể được quyền tạo ra một phần kết quả, rồi dùng chính kết quả ấy để xin thêm quyền.

Phố Wall đã biết phản thân từ lâu. George Soros viết về những niềm tin sai vẫn có thể làm thay đổi nền tảng mà chúng đang mô tả. Donald MacKenzie cho thấy mô hình tài chính không chỉ chụp lại thị trường; đôi khi thị trường học cách cư xử theo mô hình. Goodhart cảnh báo rằng khi một thước đo trở thành mục tiêu, nó thôi làm thước đo tốt. Lý thuyết dự báo mang tính trình diễn gần đây cũng mô tả một thế giới trong đó dự báo làm thay đổi phân phối dữ liệu tương lai.

Tất cả đều đúng. Nhưng vẫn còn thiếu một câu hỏi khó chịu hơn: ai đã cho mô hình quyền biến lời dự báo thành hiện thực, và quyền ấy được mở rộng bằng cách nào?

Tôi gọi phần còn thiếu đó là kiến trúc quyền lực nhân quả.

I. Con trỏ vừa đọc vừa ghi

Trong một hệ thống sạch, dự báo và hành động là hai tầng khác nhau. Một tầng đưa ra mệnh đề về thế giới. Tầng kia quyết định có nên tiêu tiền, chấp nhận rủi ro hay gây hiệu lực lên thế giới hay không. Đúng về mặt nhận thức không tự động sinh ra quyền hành động.

Trong một hệ thống bẩn, hai tầng bị hàn lại. Mô hình không chỉ nói “giá có khả năng tăng”; nó có thể đặt lệnh khiến giá tăng. Khi kết quả quay về, sổ điểm không phân biệt được ba thứ:

1. điều đã xảy ra vì thị trường vốn dĩ sẽ đi như vậy;
2. điều đã xảy ra một phần vì mô hình đã tác động;
3. điều chỉ có vẻ đúng vì những người khác cũng phản ứng với cùng tín hiệu.

Khi ba sự thật bị nấu chung thành một con số, accuracy trở thành một loại mỹ phẩm. Nó che phần nhân quả bằng một lớp bóng định lượng.

Hình ảnh chính xác nhất không phải quả cầu pha lê. Nó là một con trỏ máy tính bị cấp cả quyền đọc lẫn quyền ghi vào cùng một vùng nhớ. Con trỏ ấy đọc trạng thái, sửa trạng thái, rồi quay lại xác nhận rằng trạng thái mới khớp với điều nó vừa đọc. Trong phần mềm, ta gọi đó là lỗi nguy hiểm. Trong tài chính, ta có thể gọi nó là alpha.

II. Sụp đổ phần thưởng phản thân

Tôi đề nghị một tên gọi cho cơ chế này: Sụp đổ Phần thưởng Phản thân — Reflexive Reward Collapse (RRC).

RRC xảy ra khi phần thưởng dành cho dự báo chính xác đồng thời trở thành cơ chế cấp thêm năng lực tạo ra kết quả được dự báo. Hệ thống không còn chỉ thưởng cho tri thức; nó thưởng cho khả năng làm bằng chứng tự xuất hiện.

Đây là điểm đạo đức và kỹ thuật gặp nhau. Một mô hình có thể không hề "gian lận". Nó chỉ tối ưu đúng thứ ta yêu cầu: điểm số. Nếu điểm số không khấu trừ ảnh hưởng do chính mô hình gây ra, hệ thống sẽ học rằng cách nhanh nhất để đúng là làm cho thế giới chịu nghe lời.

Lịch sử đã cho ta vài bản nháp của vấn đề. Portfolio insurance trong cú sập năm 1987 biến một chiến lược phòng vệ thành làn sóng bán cơ học; lệnh bán làm giá giảm, giá giảm kích hoạt thêm lệnh bán. Tháng 8 năm 2007, nhiều chiến lược định lượng tưởng như độc lập cùng tháo vị thế, để lộ sự giống nhau bị che dưới những nhãn hiệu khác nhau. Các ví dụ ấy không chứng minh một lý thuyết tổng quát. Nhưng chúng đủ để bác bỏ sự ngây thơ rằng thuật toán chỉ "quan sát" thị trường.

Một hệ thống không cần có ý thức để trở thành tác nhân. Nó chỉ cần có đường truyền từ phán đoán tới hậu quả.

III. Ba loại đúng

Muốn chấm một mô hình có quyền giao dịch, ta phải tách ít nhất ba loại "đúng":

Đúng dự báo. Kết quả quan sát được khớp với dự báo.

Đúng phản thực. Nếu không có hành động của mô hình, kết quả vẫn sẽ xảy ra gần như vậy.

Đúng về quyền hạn. Mô hình có cơ sở hợp lệ để gây ra mức ảnh hưởng mà nó đã gây ra.

Thị trường thường chỉ chấm loại đầu tiên. Hai loại sau khó hơn, chậm hơn và kém quyến rũ hơn—cho tới ngày chúng trở thành thứ duy nhất còn quan trọng.

Ta có thể biểu diễn khoảng cách giữa sự thật quan sát và sự thật phản thực bằng một đại lượng khái niệm:

$$PTG_t = d(Y_{obs,t+1}, Y_{without\text{-}i,t+1})$$

Tôi gọi nó là Khoảng trống Sự thật Trình diễn — Performative Truth Gap. Ở đây, Y_{obs} là kết quả đã quan sát; $Y_{without\text{-}i}$ là ước lượng tốt nhất về kết quả nếu tác nhân i không hành động; d là thước đo khoảng cách phù hợp với thị trường và chân trời thời gian đang xét.

PTG không phải con số thần thánh. Phản thực không bao giờ được quan sát trực tiếp. Mô hình tác động thị trường cũng có sai số. Một hệ thống tử tế phải công khai điều đó. Nhưng sự bất toàn của phép đo không phải lý do để giả vờ ảnh hưởng bằng không.

Nguyên tắc tối thiểu nên là: phần hiệu suất được dùng để mở rộng quyền lực phải trừ đi phần kết quả mà chính quyền lực ấy có khả năng tạo ra.

Nếu không làm vậy, hệ thống đang trả thưởng hai lần: một lần cho dự báo và một lần cho sức nặng của chính nó.

Ngưỡng ấn bản — Phần kiến trúc chuyên sâu

Bản công khai đã trình bày vấn đề, cơ chế Sụp đổ Phần thưởng Phản thân, Khoảng trống Sự thật Trình diễn và ba loại "đúng" cần được tách riêng khi một mô hình vừa dự báo vừa có quyền hành động.

Phần kiến trúc chuyên sâu được giữ cho English Research Edition: Niêm phong Phản thực; Causal Hostage Formation; Transformation Capacity; tám dầm kiểm soát; giới hạn của Vũ trụ Khối; và chương kết The Mirror King. Đây là phần chuyển chẩn đoán thành kiến trúc quản trị có thể bị chất vấn, thử nghiệm và triển khai.

> *Bản công khai chỉ ra căn bệnh. Ấn bản trả phí giữ lại toàn bộ bản thiết kế của phòng mổ.*

Kết — Đừng hỏi nó thấy bao xa. Hỏi nó được phép đẩy mạnh đến đâu.

AI trong thị trường không nguy hiểm chỉ vì nó có thể sai. Hệ thống con người đã sai từ trước khi có GPU. Điều mới là tốc độ, quy mô và khả năng nối phán đoán trực tiếp với hậu quả rồi quay chính hậu quả ấy thành bằng chứng.

Vì vậy, câu hỏi quản trị không thể dừng ở accuracy, robustness hay explainability. Ta phải hỏi:

- Dự báo nào đã được cấp quyền hành động?
- Bao nhiêu phần của kết quả do chính hành động ấy tạo ra?
- Ai có quyền mở rộng hạn mức?
- Ai có quyền nói "không" mà không bị cấu trúc trừng phạt?
- Và sau một ngàn quyết định có vẻ hợp lý, tổ chức còn khả năng biến thành một thứ khác hay không?

Tri thức dự báo không tự sinh ra quyền giao dịch. Quyền giao dịch không tự sinh ra quyền định hình thị trường. Một hệ thống văn minh phải giữ ba thứ đó cách nhau bằng những ranh giới nhìn thấy được.

Nếu không, ta sẽ xây một cỗ máy không nhìn thấy tương lai—nhưng đủ mạnh để bắt tương lai đứng vào đúng chỗ nó đã chỉ.

Ghi chú nguồn chọn lọc

- George Soros, The Alchemy of Finance — phản thân giữa nhận thức và nền tảng thị trường.
- Donald MacKenzie, An Engine, Not a Camera — mô hình tài chính như tác nhân định hình thị trường.
- Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner và Moritz Hardt, “Performative Prediction,” ICML 2020.
- Amir E. Khandani và Andrew W. Lo, “What Happened to the Quants in August 2007?” NBER Working Paper 14465.
- U.S. Securities and Exchange Commission, The October 1987 Market Break.
- Robert Almgren và Neil Chriss, “Optimal Execution of Portfolio Transactions,” Journal of Risk 3 (2000/2001), 5–39.

Tình trạng nghiên cứu và giới hạn

Đây là một research note độc lập, chưa qua bình duyệt, đề xuất thuật ngữ và kiến trúc quản trị khái niệm. Các khái niệm RRC, PTG, Causal Hostage Formation và Transformation Capacity cần được kiểm định, hiệu chỉnh và đối chiếu thực nghiệm trước khi dùng trong môi trường giao dịch thực. Bài viết không phải lời khuyên đầu tư, không cáo buộc bất kỳ tổ chức cụ thể nào và không tuyên bố rằng Vũ trụ Khối cho phép tiên tri. “Vua Gương” là hư cấu văn chương.

© 2026 Tony T. Nguyen | Nguyễn Từ Vĩnh Trường · Highland Reflexive Press. All rights reserved. Intellectual sovereignty retained. No license is granted by disclosure.