



Các doanh nghiệp AI dẫn đầu của Trung Quốc và các cơ quan chính phủ Trung Quốc

Nơi chính sách, doanh nghiệp, dữ liệu và phần cứng ăn khớp trong nền kinh tế trí tuệ

Luật sư Kim Kyung-jin

AI Library - kimkj.com - 2026

Mục lục

Phần I. Chiến lược Quốc gia và Hệ thống Quản trị nhà nước

1. Chương 1. Từ AI Thế hệ Tiếp theo đến AI+: Chiến lược Hướng tới Kế hoạch Năm năm thứ 15

Phần I. Chiến lược quốc gia và hệ thống quản trị

2. Chương 2. Vai trò của Quốc vụ viện, NDRC, CAC, MIIT và SAC
3. Chương 3. Các quy tắc kiểm soát dữ liệu được xây dựng giữa phát triển và an ninh
4. Chương 4. Cục Dữ liệu Quốc gia, chính sách bộ dữ liệu và mối liên kết giữa mô hình và dữ liệu

Phần II. Hệ sinh thái của các công ty dẫn đầu

5. Chương 5. DeepSeek và những con hổ AI: hai vòng lặp do mã nguồn mở tạo ra

Phần II. Hệ sinh thái của các doanh nghiệp dẫn đầu

6. Chương 6. Alibaba, Tencent, Baidu, ByteDance và iFLYTEK trong cuộc đua toàn bộ chuỗi công nghệ
7. Chương 7. SenseTime và mô hình công ty niêm yết dành cho các doanh nghiệp thị giác máy tính và trí tuệ không gian
8. Chương 8. Huawei Ascend, Quý Lớn và Đồng Số Tây Toán trong nỗ lực tự chủ phần cứng

Phần III. Sự dung hợp giữa chính sách và công nghiệp, cùng cuộc cạnh tranh quyền lực

9. Chương 9. AI+ trong nền kinh tế thực: Chính sách bước vào sản xuất, giáo dục, tiêu dùng và robot hình người

Phần III. Sự hòa nhập giữa chính sách và công nghiệp, cùng cuộc cạnh tranh quyền lực

10. Chương 10. Cạnh tranh AI Mỹ-Trung, kiểm soát xuất khẩu và sự đảo chiều trong kiểm soát tiếp cận ở nước ngoài

Phần III. Sự hòa trộn giữa chính sách và công nghiệp, cùng cuộc cạnh tranh quyền lực

11. Chương 11. Những giới hạn đã được xác nhận, những rủi ro chưa được kiểm chứng, và Trung Quốc sẽ đưa trí tuệ quân sự cùng AGI đi xa đến đâu
12. Chương 12. Những lựa chọn của Hàn Quốc: Vẽ lại khoảng cách giữa chính phủ, doanh nghiệp và dữ liệu

Chương 1. Từ AI Thế hệ Tiếp theo đến AI+: Chiến lược Hướng tới Kế hoạch Năm năm thứ 15

Tháng 7 năm 2017, một tài liệu xuất hiện trên trang web của Hội đồng Nhà nước Trung Quốc. Tiêu đề của nó là "Kế hoạch Phát triển Trí tuệ Nhân tạo Thế hệ Mới" (). Nó xuất hiện bốn tháng sau khi từ "trí tuệ nhân tạo" lần đầu tiên xuất hiện trong Báo cáo Công tác Chính phủ vào tháng Ba năm đó. Một báo cáo về xu hướng chính sách của Viện Chính sách Kinh tế Quốc tế Hàn Quốc mô tả nó là tài liệu có hệ thống đầu tiên của Trung Quốc trong lĩnh vực trí tuệ nhân tạo. Trước đó, các bộ riêng lẻ đã phát hành các kế hoạch riêng của họ, nhưng đây là tài liệu đầu tiên nâng trí tuệ nhân tạo lên vị trí hàng đầu trong chiến lược quốc gia và đặt các mục tiêu cho đến năm 2030 trên một bản thiết kế thống nhất.

Khoảng cách bốn tháng này đáng lưu ý. Khi trí tuệ nhân tạo được đề cập trong một dòng của Báo cáo Công tác Chính phủ, một kế hoạch quốc gia nhanh chóng theo sau. Điều đó báo hiệu rằng trí tuệ nhân tạo đã chuyển từ một chương trình cấp bộ sang một vấn đề được Hội đồng Nhà nước xử lý trực tiếp.

Khi bạn mở tài liệu, một đặc điểm bất thường nổi bật. Các mục tiêu không phải là những khẩu hiệu. Chúng được cố định bằng các con số và năm. Kế hoạch đặt ra ba mốc thời gian, 2020, 2025 và 2030, và gán một quy mô ngành công nghiệp tính bằng nhân dân tệ cho mỗi mốc. Theo báo cáo của Viện Chính sách Kinh tế Quốc tế Hàn Quốc và Samjong KPMG, mục tiêu vào năm 2020 là phát triển ngành trí tuệ nhân tạo cốt lõi lên 150 tỷ nhân dân tệ và các ngành liên quan lên 1 nghìn tỷ nhân dân tệ.

Đến năm 2025, các mục tiêu là 400 tỷ nhân dân tệ cho ngành cốt lõi và 5 nghìn tỷ nhân dân tệ cho các ngành liên quan. Đến năm 2030, chúng là 1 nghìn tỷ nhân dân tệ cho ngành cốt lõi và 10 nghìn tỷ nhân dân tệ cho các ngành liên quan. Đây là ngôn ngữ của một kế hoạch công nghiệp hơn là của một kế hoạch công nghệ. Việc viết các mục tiêu bằng nhân dân tệ cũng có nghĩa rằng hiệu suất sẽ được đo lường không phải bằng số lượng bài báo hoặc bằng sáng chế, mà bằng kích thước thị trường.

Mỗi giai đoạn trong ba giai đoạn mang một trọng lượng khác nhau. Giai đoạn đầu nói bằng ngôn ngữ đặt nền móng. Mục tiêu là bắt kịp trình độ công nghệ của các nước tiên tiến, đồng thời xây dựng tiêu chuẩn kỹ thuật và hệ thống dịch vụ. Giai đoạn thứ hai vượt ra khỏi phạm vi ứng dụng và đặt mục tiêu đạt kết quả trong lý thuyết nền tảng. Giai đoạn này kêu gọi chuyển đổi thông minh trong các lĩnh vực như sản xuất, y tế, đô thị, nông nghiệp và quốc phòng, đồng thời xây dựng các quy phạm pháp luật và hệ thống lý thuyết cho trí tuệ nhân tạo. Ngôn ngữ của giai đoạn thứ ba cho thấy tham vọng lớn hơn nhiều. Văn kiện tuyên bố Trung Quốc sẽ dẫn đầu thế giới về lý thuyết, công nghệ và ứng dụng, đồng thời sẽ "hoàn thiện" các quy phạm pháp luật và hệ thống chính sách của mình.

Việc chọn năm 2030 làm mốc mục tiêu có ý nghĩa quan trọng. Văn kiện nhìn xa hơn nhiệm kỳ của bất kỳ một chính phủ nào và hướng tới một chân trời cách đó hơn một thập kỷ. Điều cũng đáng chú ý là pháp luật và quy phạm xuất hiện trong mọi giai đoạn. Đặt tiến bộ công nghệ và tiến bộ quy phạm cạnh nhau trong cùng một bảng có nghĩa là quản lý không được xem như phần phụ kiện gắn thêm về sau. Ngay từ đầu, nó đã được xem là một trục phải được thiết kế cùng với phát triển.

Các mục thuộc lý thuyết nền tảng và công nghệ cốt lõi cũng rất cụ thể. Trong báo cáo của Korea Information Society Development Institute, những khái niệm như trí tuệ dựa trên dữ liệu lớn, điện toán nhận thức xuyên phương tiện, và trí tuệ tăng cường lai ghép, nơi con người và máy móc cùng phán đoán, được liệt kê là các lý thuyết nền tảng. Động cơ điện toán tri thức, trí tuệ bầy đàn và hệ thống tự chủ không người lái được xác định là công nghệ cốt lõi. Những tên gọi này sẽ xuất hiện trở lại về sau. Kế hoạch hành động ba năm do Ministry of Industry and Information Technology (MIIT) công bố vào tháng 12 năm 2017 đã chuyển các lý thuyết này thành những sản phẩm hữu hình. Danh sách bao gồm xe kết nối thông minh, robot dịch vụ, drone, hình ảnh y tế bằng AI, hệ thống tương tác và dịch giọng nói, cùng thiết bị nhà thông minh.

Kế hoạch cũng bao gồm các mục trực tiếp xây dựng cảm biến thông minh, chip mạng nơ-ron và nền tảng mã nguồn mở nhằm giảm sự phụ thuộc vào công nghệ nước ngoài. Ngôn ngữ về tự chủ công nghệ đã có vị trí của nó trong văn kiện. Cấu trúc này, trong đó State Council vạch hướng đi còn MIIT chuyển hướng đó thành danh sách sản phẩm và hạ tầng, đã trở thành mẫu hình cho sự phân công lao động được lặp lại trong các kế hoạch sau này.

Một phân tích ở nước ngoài, được tóm lược trong nghiên cứu của University of Virginia, đã nắm bắt bản chất của kế hoạch hành động này trong một đoạn văn. Đó là một thứ ngữ pháp đặt hợp tác quốc tế và chia sẻ nguồn lực bên cạnh an ninh quốc gia, đồng thời nói về mở cửa và kiểm soát. Một câu kêu gọi mở cửa, câu kế tiếp lại kêu gọi duy trì trật tự. Cách hai giọng nói này cùng tồn tại trong một văn kiện trở thành đặc điểm xuyên suốt toàn bộ chính sách AI của Trung Quốc. Thói quen viết phát triển và an ninh song song với nhau đã hình thành trong văn kiện năm 2017. Nghiên cứu này cũng xem tính hai mặt ấy là thứ kéo dài sang hoạt động quản trị thực tế. Những câu nói về mở cửa trở thành nền tảng cho việc lan tỏa công nghiệp, còn những câu nói về trật tự trở thành nền tảng cho kiểm soát nội dung và thủ tục đăng ký, tạo ra một cấu trúc trong đó hai mặt này hỗ trợ lẫn nhau.

Văn kiện này cũng có những lời hướng tới xã hội. Nó liệt kê máy móc nông nghiệp thông minh và cảm biến nông nghiệp, giáo dục được điều chỉnh theo người học, robot hỗ trợ chăm sóc lâm sàng và phẫu thuật, chăm sóc người cao tuổi, cùng quản lý sức khỏe thông qua thiết bị đeo. Văn kiện cũng có một đoạn nói về việc khu vực quân sự và dân sự cùng chia sẻ công nghệ trí tuệ nhân tạo. Văn kiện kêu gọi xây dựng một cấu trúc trong đó doanh nghiệp, giới học thuật,

viện nghiên cứu và các tổ chức quân sự thường xuyên trao đổi và hợp tác. Phần này cần được đọc cẩn thận. Định hướng được ghi trong văn kiện và tiến trình thực tế của thông minh hóa quân sự thuộc hai cấp độ khác nhau.

Ngôn ngữ của một kế hoạch không đồng nghĩa với năng lực đã hoàn tất. Văn bản này chỉ xử lý các tuyên bố liên quan đến quân sự trong phạm vi những gì đã được xác nhận và những gì vẫn còn là rủi ro chưa được xác nhận. Ngôn ngữ về dung hợp quân sự-dân sự trong văn kiện năm 2017 nên được đọc như một tuyên bố về hướng đi dự kiến, chứ không phải như một kết quả. Các nguồn chỉ xác nhận rằng văn kiện đã nói sẽ xây dựng một cấu trúc hợp tác như vậy. Chúng không xác nhận cấu trúc ấy đã đi vào bất kỳ hệ thống vũ khí cụ thể nào đến mức nào.

Ngôn ngữ về đạo đức và hệ thống pháp lý cũng không bị bỏ sót. Kế hoạch kêu gọi nghiên cứu các vấn đề pháp lý, đạo đức và xã hội của trí tuệ nhân tạo, cũng như các chuẩn mực đạo đức và quy tắc ứng xử dành cho nhà phát triển và nhà thiết kế. Nó còn có một câu về việc xây dựng khuôn khổ đạo đức cho hợp tác trong những môi trường nơi con người và robot làm việc cùng nhau. Cách diễn đạt này dẫn tới giá trị được mô tả trong một văn kiện năm 2025 là "trí tuệ vì điều thiện" (). Cụm từ này, có nghĩa là dẫn dắt công nghệ theo hướng tốt, phản ánh một thái độ đặt quy định không ở phía đối lập với phát triển, mà như một điều kiện để phát triển.

Tuy vậy, các nhà quan sát bên ngoài lại đọc tư thế này theo cách khác. Một phân tích về quản trị của University of Virginia chỉ ra rằng cùng một ngôn ngữ về chuẩn mực cũng có thể trở thành nền tảng cho quản lý nội dung và kiểm soát biểu đạt. Một câu nói về việc hướng mọi thứ theo chiều hướng tốt có thể trở thành ngôn ngữ thúc đẩy hoặc ngôn ngữ kiểm soát, tùy vào việc nó định nghĩa thế nào là tốt. Sự căng thẳng này trở lại ở Chapter 3, nơi kiểm duyệt và đàn áp xuất hiện như những lập luận chống lại hệ thống.

Đó là ngôn ngữ của năm 2017. Nếu di sản của văn kiện này có thể rút lại trong một dòng, thì đó là: trí tuệ nhân tạo đã trở thành đối tượng để nhà nước thiết kế trực tiếp. Ngữ pháp đặt ra thời hạn, quy mô công nghiệp, danh mục công nghệ, ví dụ sản phẩm, phát triển và an ninh đều hình thành từ thời điểm ấy. Tuy vậy, văn kiện năm 2017 đặt trọng tâm vào hạ tầng và lý thuyết. Đó là giai đoạn xây dựng hệ thống và sắp xếp khung khổ. Việc đưa bản thiết kế ấy xuống nhà máy, cửa hàng và lớp học được để lại cho văn kiện tiếp theo.

Cảnh tiếp theo là Government Work Report tháng 3 năm 2024. Đây là nơi cụm từ 'Artificial Intelligence Plus (AI+)' lần đầu xuất hiện. Đó là một dòng ngắn, nhưng hướng đi đã thay đổi. Nếu năm 2017 đặt câu hỏi Trung Quốc sẽ xây dựng trí tuệ nhân tạo như thế nào, thì cụm từ này trong năm 2024 hỏi trí tuệ nhân tạo nên được gắn vào đâu. Dấu cộng gắn vào công nghiệp. Ngữ pháp ở đây là thêm nó vào sản xuất, thêm nó vào giáo dục, và thêm nó vào tiêu dùng. Trọng tâm dịch chuyển từ việc dùng công nghệ sau khi đã hoàn tất sang việc đặt nó ngay tại các địa điểm công nghiệp và đồng thời cải thiện nó ở đó.

Hướng đi này thành hình trong ý kiến tháng 8 năm 2025 của State Council, mang tên 'Opinions on Deepening the Implementation of the Artificial Intelligence Plus (AI+) Action.' Theo các chỉ số được trích trong một báo cáo của Korea Information Society Development Institute, văn kiện này đặt mục tiêu nâng tỷ lệ thâm nhập của các hệ thống thông minh lên 70 percent vào năm 2027 và 90 percent vào năm 2030. Nếu văn kiện năm 2017 nêu quy mô công nghiệp bằng nhân dân tệ, thì văn kiện này diễn đạt lại mục tiêu bằng ngôn ngữ thâm nhập, thông qua tỷ lệ áp dụng. Câu hỏi chuyển từ việc Trung Quốc sẽ tạo ra một thị trường lớn đến đâu sang việc AI sẽ lan sâu đến mức nào.

Văn kiện nêu sáu lĩnh vực: khoa học và công nghệ, công nghiệp, tiêu dùng, đời sống người dân, quản trị, và hợp tác toàn cầu. Văn kiện nói Trung Quốc sẽ thúc đẩy việc tích hợp trí tuệ nhân tạo trong từng lĩnh vực. Đích đến cuối cùng là 'Trung Quốc số.' Việc đặt mức độ thâm nhập làm mục tiêu cũng đồng nghĩa với việc thay đổi chủ thể đo lường hiệu quả. Thước đo không còn là một mô hình lớn duy nhất, mà là AI được cắm sâu đến đâu trong nhiều ngành công nghiệp.

Một báo cáo của U.S.-China Economic and Security Review Commission (USCC), thuộc U.S. Congress, phân tích sự chuyển dịch này từ bên ngoài. Báo cáo xác định cốt lõi trong quá trình phổ biến AI của Trung Quốc là một 'vòng phản hồi của nền kinh tế thực.' So với cách lan tỏa mô hình ở phương Tây, vốn thường còn nằm trong một chu trình số xoay quanh lưu lượng web và phần mềm, Trung Quốc đang đưa trí tuệ nhân tạo trực tiếp vào nhà máy, logistics, và các hiện trường robot, dựa trên một trong những nền sản xuất hàng đầu thế giới, Internet of Things, và mạng 5G.

Một mô hình được đặt trong một hiện trường vật lý liên tục tạo ra dữ liệu vận hành, loại dữ liệu không thể có được từ dữ liệu quét trên web hay dữ liệu tổng hợp. Dữ liệu đó sau đó làm thay đổi mô hình ngay tại hiện trường, và mô hình đã được chỉnh sửa lại tạo ra thêm dữ liệu. Trọng tâm của phân tích này là chu trình ấy vận hành trong thế giới vật lý. Khái niệm này là một trong những cấu trúc cốt lõi của cuốn sách, và về sau nó xuất hiện lại bên cạnh vòng lặp mã nguồn mở như một trong hai chu trình.

Một ví dụ trong báo cáo của USCC làm cho chu trình này trở nên cụ thể. Một nhà máy thông minh ở Guangdong Province đã gắn camera độ nét cao 5G và một hệ thống kiểm tra bằng AI để giám sát chất lượng sản phẩm. Báo cáo nói việc này đã giảm tỷ lệ sửa chữa thiết bị 20 percent và tiết kiệm 1 million yuan mỗi năm, khoảng \$140,000. Việc tỷ lệ sửa chữa và khoản tiết kiệm chi phí, chứ không phải những tên mô hình hào nhoáng, được dùng làm thước đo hiệu quả cho thấy tính chất của cách tiếp cận này.

Nói cách khác, trí tuệ nhân tạo có vận hành tốt hay không không được đo bằng điểm chuẩn benchmark, mà bằng chi phí của nhà máy. Điểm này cần được đọc kỹ. Khoản tiết kiệm trong một ví dụ là số liệu của riêng nhà máy đó, không phải mức trung bình của toàn ngành công

nghiệp quốc gia. Việc một trường hợp thành công xuất hiện trong văn kiện chính sách và việc tuyên bố rằng kết quả tương tự đã được tái lập ở các nhà máy khác là hai chuyện khác nhau.

Văn bản hướng dẫn đưa chu trình này xuống cấp nhà máy là 'Kế hoạch triển khai Kế hoạch hành động chuyên đề Trí tuệ nhân tạo + Sản xuất.' Các mục tiêu do Korea Institute for Advancement of Technology tóm lược dày đặc con số. Đến năm 2027, Trung Quốc đặt mục tiêu nhúng sâu ba đến năm mô hình ngôn ngữ lớn vào sản xuất, phát hành 1.000 tác nhân AI công nghiệp, xây dựng 100 bộ dữ liệu công nghiệp chất lượng cao, và phát triển 500 kịch bản ứng dụng tiêu biểu. Trung Quốc cũng muốn chỉ định hai đến ba doanh nghiệp dẫn đầu thế giới và 1.000 doanh nghiệp mô hình chuẩn có thể làm mốc tham chiếu benchmark, đồng thời hướng dẫn doanh nghiệp lập vị trí 'Chief Data Officer (CDO).' Ý định xem dữ liệu công nghiệp như một tài sản chiến lược xuất hiện như một ô trong sơ đồ tổ chức.

Kế hoạch này chia công việc thành nhiều nhiệm vụ: xây dựng nền tảng đổi mới, nâng cấp thông minh, đổi mới sản phẩm, nuôi dưỡng các chủ thể công nghiệp, mở rộng hệ sinh thái, thiết lập an toàn, và hợp tác quốc tế. Nó buộc việc xây dựng mô hình và công việc an toàn vào cùng một văn kiện. Cuộc thảo luận rằng bên nào bảo đảm được ứng dụng công nghiệp trước sẽ quyết định thứ tự kiểm soát các điểm dữ liệu cũng đi theo cùng hướng. Bên chiếm được vị trí cao trong ứng dụng sẽ nắm dữ liệu vận hành, và bên nắm dữ liệu sẽ đưa mô hình tiếp theo tiến lên sớm hơn.

Cách văn bản hướng dẫn này đi vào thực địa thể hiện trong ngành luyện thép. Trong một trường hợp về 'dự đoán thông minh nhiệt độ gang lỏng trong lò cao' được đưa vào các tài liệu liên quan đến National Data Administration, bên trong lò cao trước đây được xem như một hộp đen không thể quan sát. Công nhân dựa vào kinh nghiệm để điều chỉnh miệng gió và tỷ lệ làm giàu oxy. Nay, với một cấu trúc kết hợp tính toán cục bộ và tính toán biên, đồng thời dùng máy chủ GPU, một mô hình dự đoán chuỗi thời gian dự báo nhiệt độ gang lỏng trước hai giờ và tự điều chỉnh các giá trị vận hành. Chất lượng tốt hơn và mức tiêu thụ năng lượng thấp hơn theo sau.

Mô hình tiếp quản phần việc từng do kinh nghiệm con người đảm nhận, và dữ liệu vận hành tích lũy trong quá trình đó lại trở thành vật liệu mới cho mô hình. Đây là vòng phản hồi của nền kinh tế thực đang quay ngay trước lò cao. Cũng cần để ý rằng trường hợp này xuất hiện trong các tài liệu phía Cục Dữ liệu Quốc gia. Một cải tiến quy trình của một doanh nghiệp không chỉ dừng lại ở thành quả riêng của doanh nghiệp đó. Nó được sắp xếp thành một kịch bản công nghiệp trong tài liệu của cơ quan quốc gia phụ trách chính sách dữ liệu.

Các agent công nghiệp và trí tuệ hiện thân là phương tiện đẩy chu trình này vào thế giới vật lý. Nếu một mô hình dựa trên phần mềm muốn xử lý máy móc thật, nó cần có tay và chân. Agent và robot đảm nhận vai trò đó. Trong trường hợp được Samjong KPMG dẫn ra, Xiaomi đã đưa robot hình người vào thay công nhân ở công đoạn lắp đai ốc tại một nhà máy xe điện, cho

chúng làm việc tự động trong ba giờ liên tiếp, và ghi nhận tỷ lệ thành công 90,2% khi xử lý đồng thời cả hai bên. Con số này là giá trị chứng minh cho một công đoạn. Cách đọc chính xác hơn không phải là bằng chứng rằng robot đã thay thế hoàn toàn con người, mà là một ghi nhận rằng việc lặp lại tự động đã trở nên khả thi đối với một nhiệm vụ cụ thể.

Một báo cáo của Mirae Asset Securities xem robot hình người là phần cứng chiến lược tiếp theo sau máy tính, điện thoại thông minh và xe điện. Báo cáo này cũng ghi nhận rằng Trung Quốc chiếm một tỷ trọng đáng kể trong phân bố toàn cầu của các công ty robot hình người. Dù vậy, phần này cần được đọc mà không phóng đại. Phân bố về số lượng công ty và việc sản xuất hàng loạt cùng lợi nhuận thực tế thuộc hai tầng khác nhau, và chuyển động trên sân khấu trình diễn không phải là cùng một vấn đề với vận hành liên tục trên dây chuyền nhà máy. Chương 9 sẽ quay lại robot hình người và tách các giá trị chứng minh đã được xác nhận khỏi những dự báo chưa được xác nhận.

Đến đây, một khái niệm trong các văn kiện chính sách bắt đầu có trọng lượng: 'nền kinh tế thông minh' (). Các bình luận chính sách liên quan chỉ ra rằng tín hiệu đã trở nên rõ ràng về việc nền kinh tế thông minh là gì và sẽ được theo đuổi như thế nào. Thuật ngữ này chỉ một trạng thái trong đó trí tuệ nhân tạo không còn bị để lại như công cụ của một ngành riêng lẻ, mà dữ liệu, năng lực tính toán và mô hình trở thành tài nguyên cơ bản của toàn bộ nền kinh tế. Cuộc thảo luận về việc chiếm điểm cao trong các ứng dụng công nghiệp cũng đi theo cùng hướng. Bên nào giành được địa điểm ứng dụng trước thì nắm dữ liệu, và bên nào nắm dữ liệu thì cải thiện mô hình. Nếu ngôn ngữ của năm 2017 là ngôn ngữ của một hệ thống kỹ thuật, thì ngôn ngữ ở giai đoạn này là việc chiếm trước ứng dụng và dữ liệu.

Nơi AI bám vào trước tiên sẽ trở thành chiến trường cạnh tranh mang tính quyết định. Cần lưu ý rằng các thuật ngữ nền kinh tế thông minh, lực lượng sản xuất chất lượng mới, và AI+ có phần chồng lấn trong cách dùng. Ba khái niệm này nằm ở những tầng hơi khác nhau. AI+ chỉ hành động gắn AI vào một thứ gì đó. Nền kinh tế thông minh chỉ trạng thái kinh tế được tạo ra từ kết quả ấy. Lực lượng sản xuất chất lượng mới gần hơn với một chiếc ô bao trùm trạng thái đó trong câu chuyện tăng trưởng quốc gia. Nếu gom cả ba thuật ngữ thành một, sắc thái của các văn kiện sẽ bị mờ đi.

Dòng chảy này được gom lại thành một mục tiêu quốc gia trong 'Đề cương Kế hoạch 5 năm lần thứ 15,' được công bố vào tháng 3 năm 2026. Bao trùm các năm từ 2026 đến 2030, văn kiện này là lộ trình cấp cao nhất cho năm năm tiếp theo và là bản thiết kế cho hiện đại hóa kiểu Trung Quốc đến năm 2035. Phần mở đầu đặt ra tuyên bố rằng Trung Quốc sẽ rời khỏi mô hình tăng trưởng xoay quanh bất động sản và hạ tầng, rồi chuyển trực sang 'lực lượng sản xuất chất lượng mới' () và nền kinh tế thông minh. Nhìn vào những thay đổi trong các kế hoạch trước đây theo phần tổng hợp của Mirae Asset Securities, kế hoạch này đi sau con đường đã chạy qua 'Internet Plus' và 'Made in China 2025' của Kế hoạch lần thứ 13, rồi 'tuần hoàn kép' và

'Digital China' của Kế hoạch lần thứ 14.

Các từ khóa trong văn kiện này là lực lượng sản xuất chất lượng mới, hiện đại hóa kiểu Trung Quốc, tự lực về công nghệ tiên tiến, và Artificial Intelligence Plus. Nếu lần theo đường đi mà tên gọi các kế hoạch đã thay đổi, ta có thể thấy câu hỏi 'gắn nó vào đâu' đã được đặt lại một lần sau mỗi năm năm. Trung Quốc đã đi từ thời kỳ gắn nó vào internet, qua thời kỳ gắn nó vào sản xuất chế tạo, và nay bước vào một thời kỳ trong đó chính trí tuệ được đặt làm lớp nền của nền kinh tế.

Kế hoạch lần thứ 15 cố định các mục tiêu bằng cách chia chúng thành 20 chỉ tiêu định lượng. Trong số đó có một chỉ tiêu đo quy mô của nền kinh tế thông minh: tỷ trọng GDP do giá trị gia tăng từ các ngành cốt lõi của nền kinh tế số tạo ra. Mục tiêu là nâng con số này từ 10.5% vào năm 2024 lên 12.5% vào năm 2030. Văn kiện cũng đặt ra các chỉ tiêu duy trì mức tăng trưởng bình quân hằng năm của chi tiêu nghiên cứu và phát triển ở mức 7% trở lên, đồng thời tăng số bằng sáng chế phát minh có giá trị cao trên mỗi 10,000 người từ 16 lên 22 hoặc hơn. Thay vì đặt mục tiêu bằng cách tách riêng trí tuệ nhân tạo, văn kiện đưa quá trình thông minh hóa vào các chỉ tiêu tăng trưởng của toàn bộ nền kinh tế.

Quan niệm của nền kinh tế thông minh rằng trí tuệ nhân tạo phải trở thành nền tảng của nền kinh tế, chứ không phải một ngành riêng biệt, được phản ánh trực tiếp trong cách thiết kế các chỉ tiêu này. Cũng đáng chú ý là các chỉ tiêu dân sinh xuất hiện trong cùng một bảng, gồm việc giữ tỷ lệ thất nghiệp khảo sát ở đô thị dưới 5,5% và nâng số năm đi học trung bình của dân số trong độ tuổi lao động từ 11,3 năm lên 11,7 năm. Việc các chỉ tiêu công nghệ, việc làm và giáo dục được quản lý song song trong cùng một văn kiện cho thấy nỗ lực điều phối tốc độ thông minh hóa và khả năng hấp thụ cú sốc của nó bằng cùng một bàn tay.

Bản đồ công nghiệp mà kế hoạch này vẽ ra rất rộng. Nó đưa ra bốn hướng công việc: làm cho các ngành truyền thống trở nên thông minh và xanh hơn, phát triển các ngành mới nổi như năng lượng mới, hàng không vũ trụ và kinh tế tầm thấp, đồng thời giành trước vị thế trong các ngành tương lai, gồm điện toán lượng tử, sản xuất sinh học, hydro, giao diện não-máy tính, robot hình người và 6G. Nâng đỡ phía dưới là hạ tầng thể hệ mới như mạng điện toán tích hợp và mạng truyền thông mới. Trí tuệ nhân tạo được đặt như một tầng chung chạy xuyên qua cả bốn hướng.

Nó không phải một ô ngành riêng, mà là nền tảng cắt ngang mọi ô. Mirae Asset Securities đọc cách sắp xếp này như một lựa chọn nhằm tìm lối đột phá bằng cách hoàn tất tự chủ công nghệ và hệ sinh thái công nghiệp trong nước, giữa bối cảnh Hoa Kỳ siết chặt các hạn chế đối với công nghệ tiên tiến. Sức ép quản lý từ bên ngoài càng chặt, ngôn ngữ tự chủ ở bên trong càng mạnh. Chương 10 sẽ quay lại mối quan hệ này khi bàn về kiểm soát xuất khẩu.

Lộ trình thâm nhập trí tuệ nhân tạo trong Kế hoạch 5 năm lần thứ 15 vận hành cùng với các ý kiến AI+ đã bàn ở trên. Đến năm 2027, phát triển tích hợp sẽ được thực hiện trong các lĩnh vực

như khoa học và công nghệ, công nghiệp, tiêu dùng, dân sinh, quản trị và hợp tác toàn cầu, trong khi tỷ lệ thâm nhập của các hệ thống thông minh sẽ tăng lên 70%. Sau đó con số này sẽ đạt 90% vào năm 2030, tiếp theo là việc hình thành một nền kinh tế và xã hội thông minh toàn diện vào năm 2035. Bên dưới đó, các mục tiêu cấp nhà máy của 'Artificial Intelligence + Manufacturing' tiếp tục được duy trì.

Các tầng được xếp thành một chuỗi liên mạch, từ tầm nhìn quốc gia đến định hướng công nghiệp, rồi từ định hướng công nghiệp đến lò cao và dây chuyền lắp ráp. Cấu trúc này được thiết kế để tỷ lệ áp dụng 90 phần trăm ở tầng trên và 1.000 agent ở tầng dưới cùng chỉ về một hướng. Cấu trúc nhiều tầng này cũng là lý do cuốn sách xem xét cơ quan nhà nước và doanh nghiệp cùng nhau. Nếu chỉ nhìn một tầng, ta không thể thấy vì sao những con số ấy xuất hiện.

Động cơ này cần sức mạnh tính toán để vận hành. Kế hoạch 5 năm lần thứ 15 kêu gọi xây dựng các cụm tính toán thông minh quy mô siêu lớn và đưa ra một hệ thống trong đó đám mây, biên và thiết bị cùng chuyển động. Khung vật lý là dự án 'Eastern Data, Western Computing', đưa dữ liệu từ miền đông sang miền tây, nơi điện năng và đất đai dồi dào hơn, để xử lý tính toán. Các trung tâm dữ liệu đang được xây dựng tại tám đầu mối, bao gồm vùng Jing-Jin-Ji, Đồng bằng sông Dương Tử, Guizhou và Inner Mongolia.

Cùng lúc đó, để ngăn đầu tư quá nóng, nhà nước đặt ra các quy định về hiệu suất điện năng cho trung tâm dữ liệu và chuyển đầu tư hạ tầng địa phương sang cơ chế phê duyệt. Lối tư duy từng xuất hiện trong các văn bản ghép phát triển với an ninh, vừa thúc đẩy tiến lên vừa siết chặt kiểm soát, lại xuất hiện trong chính sách hạ tầng. Khi nhà nước phân bổ mạng lưới tính toán và quản lý hiệu suất của nó, tính toán vừa là thiết bị của doanh nghiệp, vừa là tài nguyên công do nhà nước quản lý.

Hệ thống dữ liệu cũng vận hành song song. Dưới sự quản lý của National Data Administration, một nền tảng đăng ký các nguồn dữ liệu công, và Ministry of Finance đã đặt ra các chuẩn mực kế toán cho phép doanh nghiệp đưa dữ liệu công nghiệp vào báo cáo tài chính như tài sản vô hình hoặc tài sản tồn kho. Khi dữ liệu bắt đầu xuất hiện trên sổ sách như một tài sản, nó đã vượt ra khỏi khẩu hiệu chính sách và bước vào ngôn ngữ kế toán. Thói quen chính sách nói về mô hình và dữ liệu cùng nhau nay đã đi đến bảng cân đối kế toán của doanh nghiệp.

Chính quyền địa phương cũng đang đi theo cùng một hướng. Tô Châu đã khởi động một kế hoạch hành động cấp thành phố, lập ra một nhóm lớn các quỹ trí tuệ nhân tạo và chọn những mô hình trọng điểm để hỗ trợ. Sự ăn khớp giữa quy hoạch trung ương, quỹ địa phương và hạch toán doanh nghiệp chính là mối liên kết giữa chính sách, công ty và dữ liệu mà cuốn sách này mô tả.

Một đoạn nổi bật trong Kế hoạch 5 năm lần thứ 15 là cách kế hoạch này xử lý vấn đề việc làm. Đề cương kế hoạch nói rằng Trung Quốc sẽ xây dựng một hệ thống đánh giá việc làm đầy đủ

chất lượng cao, đồng thời ứng phó một cách tích hợp với tác động của các công nghệ mới như trí tuệ nhân tạo và những thay đổi của môi trường bên ngoài. Văn kiện này cũng nói Trung Quốc sẽ xác định các việc làm cộng tác mới, nơi con người và máy móc làm việc cùng nhau, và đưa trí tuệ nhân tạo vào những lĩnh vực gắn với đời sống hằng ngày của người dân, như nông nghiệp, y tế và chăm sóc người cao tuổi. Một văn kiện thúc đẩy công nghệ, ngay trong cùng văn kiện đó, cũng lo về những việc làm mà công nghệ có thể làm chao đảo. Điều đó cho thấy thái độ của kế hoạch. Cụm từ năm 2025 'trí tuệ vì điều thiện' và phần chỉ đạo về việc làm này cùng xuất phát từ một nơi. Nó nói về phát triển, trong khi ghi lại cả cái bóng của phát triển.

Dĩ nhiên, đưa cái bóng này vào các mục tiêu quốc gia không có nghĩa là cái bóng ấy đã biến mất. Câu nói về việc xây dựng một hệ thống đánh giá việc làm đầy đủ và những gánh nặng thực tế của lao động gắn nhãn dữ liệu hay khoảng cách vùng miền là những câu hỏi khác nhau. Một kế hoạch là văn kiện nêu mục tiêu, không phải văn kiện bảo đảm kết quả. Các nhà quan sát bên ngoài Trung Quốc xem phần này nghiêm túc hơn. Một phân tích về quản trị từ University of Virginia chỉ ra rằng thứ ngữ pháp nói về phát triển và an ninh cùng lúc cũng có thể trở thành cơ sở cho kiểm soát nội dung và trấn áp.

Mối lo nằm ở chỗ nhà nước quyết định điều gì sẽ bị lọc bỏ dưới lớp ngôn ngữ chuẩn mực về việc dẫn công nghệ hướng tới điều thiện. Cuốn sách này xử lý gánh nặng đó riêng ở Chương 3, như bằng chứng phản chứng trong phần bàn về kiểm duyệt và chi phí quản lý mà các công ty phải gánh. Phần chỉ đạo về việc làm trong Kế hoạch 5 năm lần thứ 15 và ngôn ngữ 'trí tuệ vì điều thiện' vẫn vừa là điểm xuất phát, vừa là đối tượng cần kiểm chứng của cuộc thảo luận ấy.

Nếu đặt văn kiện năm 2017 bên cạnh kế hoạch năm 2026 và nhìn vào cách ngôn ngữ thay đổi trong chín năm, mạch vận động sẽ hiện rõ. Ban đầu, đó là ngôn ngữ của việc tạo ra công nghệ. Lý thuyết cơ bản, công nghệ cốt lõi và các con số tính bằng nhân dân tệ về quy mô công nghiệp được đặt lên trước. Sau đó, nó chuyển sang ngôn ngữ gắn AI vào các lĩnh vực khác. Ngữ pháp AI+, được thêm vào sản xuất và giáo dục, dần đứng vững cùng với các tỷ lệ áp dụng, tức những con số về mức thâm nhập. Giờ đây, đó là ngôn ngữ của nền móng. Dưới những tên gọi như nền kinh tế thông minh và lực lượng sản xuất chất lượng mới, trí tuệ nhân tạo được định nghĩa không phải là công cụ của một ngành, mà là một lớp phủ ngang qua toàn bộ nền kinh tế. Trọng tâm đã chuyển từ chế tạo, sang gắn vào, rồi đến thẩm sâu.

Hai thói quen xuyên suốt sự thay đổi này. Một là thói quen viết mục tiêu bằng con số và mốc năm. Từ 150 tỷ nhân dân tệ đến 10 nghìn tỷ nhân dân tệ, từ tỷ lệ áp dụng 70 phần trăm đến 90 phần trăm, và từ tỷ trọng kinh tế số 10.5 phần trăm đến 12.5 phần trăm, các con số xuất hiện trong những văn kiện khác nhau nhưng trông như được viết bởi cùng một bàn tay. Thói quen còn lại là đặt phát triển và an ninh sát cạnh nhau, gần như từng câu một. Mở cửa và trật tự, đầu tư và hệ thống phê duyệt, công nghệ và đạo đức, chuyển đổi thông minh và việc làm luôn xuất hiện thành từng cặp. Ngữ pháp kép mà nghiên cứu của University of Virginia nhận diện trong

văn kiện năm 2017 vẫn vận hành trong kế hoạch chín năm sau.

Đây là lý do cuốn sách này không tách riêng doanh nghiệp và chính phủ. Qwen của Alibaba, ERNIE của Baidu và Seed của ByteDance vận động dưới tên các công ty, nhưng những đường kẻ của sân chơi mà chúng chạy trên đó trước hết đã được vạch ra bởi các văn kiện của State Council, Ministry of Industry and Information Technology và National Data Administration. Các mô hình của doanh nghiệp lấp đầy mục tiêu áp dụng 90 phần trăm, trong khi bộ dữ liệu, chuẩn kế toán và mạng lưới tính toán mà các mô hình ấy dựa vào được chuẩn bị bằng văn kiện của chính phủ. Chính phủ viết ra mục tiêu 1,000 tác nhân công nghiệp, nhưng doanh nghiệp cung cấp những bàn tay xây dựng chúng. Chính sách, doanh nghiệp và dữ liệu cùng quay như ba bánh răng.

Câu hỏi còn lại là bao nhiêu phần của bản thiết kế này sẽ đi từ những con số trên giấy thành hiệu suất thật. Kế hoạch năm 2017 đã ghi các mục tiêu cho năm 2030, và giờ đây, khi khoảng một nửa quãng thời gian đó đã ở phía sau, chúng ta có cả ngôn ngữ của mục tiêu lẫn hồ sơ từ hiện trường. Mô hình dự báo nhiệt độ trước lò cao và tỷ lệ thành công của robot trên dây chuyền lắp ráp là những ghi nhận đã được xác nhận. Ngược lại, các con số cho năm 2030 như tỷ lệ áp dụng 90 phần trăm hay tỷ trọng kinh tế số 12.5 phần trăm vẫn chỉ tồn tại như mục tiêu. Còn những tuyên bố xa hơn, chẳng hạn như thông minh hóa quân sự hoặc một bước nhảy tự chủ của trí tuệ nhân tạo, vượt quá những gì các nguồn này xác nhận.

Tài liệu mô tả rõ một cấu trúc hợp tác quân sự - dân sự, nhưng phần tư liệu này chưa cho phép nói năng lực nào đã được chuyển giao, hoặc chuyển giao đến mức nào. Khoảng trống đó là lý do từ chương tiếp theo trở đi, phần chính lần lượt đi theo từng cơ quan nhà nước và từng công ty. Nó cố đo khoảng cách giữa các mục tiêu được viết trong tài liệu và những sự kiện được xác nhận tại hiện trường, theo từng nguồn một. Nhan đề lạ "Mosu Gongzhen" trong danh sách thông báo của National Data Administration nằm ở đâu trong khoảng cách ấy sẽ rõ hơn khi xem xét vị trí của cơ quan này.

Chương 2. Vai trò của Quốc vụ viện, NDRC, CAC, MIIT và SAC

Mở hệ thống đăng ký tiêu chuẩn do Cục Quản lý Nhà nước về Điều tiết Thị trường và Cục Tiêu chuẩn hóa Trung Quốc cùng vận hành, bạn sẽ thấy các tài liệu đang được thẩm định hoặc lưu hành dưới dạng dự thảo trong nửa đầu năm 2026, xếp nối nhau theo số thứ tự. Ở đâu đó trong danh sách ấy xuất hiện nhan đề "Yêu cầu an toàn cơ bản đối với ứng dụng tác nhân thông minh." Mở tài liệu ra, một tên công ty quen thuộc hiện lên trong mục đơn vị soạn thảo. Một công ty viễn thông lớn đứng cạnh một viện nghiên cứu quốc gia và một trung tâm ứng phó khẩn cấp, cùng tham gia viết tiêu chuẩn. Đây là một văn bản quản lý, vậy mà một công ty thuộc diện bị quản lý đã bước vào chính văn bản ấy và đang tra chuốt từng câu chữ. Nếu cảnh này khiến bạn thấy lạ, bạn mới chỉ nhìn thấy một nửa cách Trung Quốc xây dựng hệ thống quản trị trí tuệ nhân tạo.

Riêng cảnh đó đã là một phiên bản thu nhỏ của câu chuyện trong chương này. Trí tuệ nhân tạo của Trung Quốc vận động dưới các tên công ty, như Qwen của Alibaba, ERNIE của Baidu, và các mô hình mở của DeepSeek. Người ta thường cố đọc lĩnh vực này qua bảng xếp hạng doanh nghiệp và điểm chuẩn kỹ thuật. Nhưng sân chơi nơi các công ty ấy chạy đua đã có vạch kẻ sẵn. Những vạch ấy không do một bàn tay duy nhất vẽ ra. Quốc vụ viện đặt hướng đi, Ủy ban Cải cách và Phát triển Quốc gia (NDRC) điều phối giữa các bộ, Cục Quản lý Không gian mạng Trung Quốc (CAC) xem xét mô hình nào được phép xuất hiện trước công chúng, còn Bộ Công nghiệp và Công nghệ Thông tin (MIIT) cùng Cục Tiêu chuẩn hóa Trung Quốc (SAC) đặt tiêu chí an toàn và tiêu chuẩn công nghiệp.

Tên gọi có thể xa lạ, vai trò có thể như chồng lên nhau, nhưng nếu lần theo từng văn kiện, bạn sẽ thấy mỗi cơ quan đứng ở đâu và nắm gì trong tay. Chương này giải thích năm cơ quan ấy qua các văn kiện chính sách và các trường hợp cụ thể, để khoảng cách giữa chính phủ và doanh nghiệp hiện ra rõ hơn.

Hãy bắt đầu với Quốc vụ viện. Quốc vụ viện là cơ quan hành pháp đứng đầu bộ máy hành chính Trung Quốc. Cơ quan này gồm thủ tướng, các phó thủ tướng và các ủy viên quốc vụ. Mỗi mùa xuân, trong kỳ họp Đại hội Đại biểu Nhân dân Toàn quốc, Quốc vụ viện công bố báo cáo công tác chính phủ, và cứ năm năm lại xác nhận Kế hoạch 5 năm, tức bản đồ kinh tế quốc gia. Trong câu chuyện trí tuệ nhân tạo, Quốc vụ viện gửi định hướng và con số từ trên xuống. Cơ quan này ấn định một công nghệ nhất định phải phát triển đến mức nào, không phải bằng khẩu hiệu, mà bằng chỉ tiêu tài chính về quy mô công nghiệp. Một khi mục tiêu đã được dịch sang ngôn ngữ của sổ sách kế toán, các bộ bên dưới, chính quyền địa phương và doanh nghiệp nhà nước không còn lựa chọn nào khác ngoài việc tiến về con số ấy.

Hướng đi này không hình thành ngay trong một lần. Nó đi qua nhiều văn kiện. Dòng chảy trở nên rõ khi theo dõi các báo cáo xu hướng chính sách của KIEP và phân tích của KISDI về chiến lược AI+. Trong Kế hoạch 5 năm lần thứ 13 năm 2016, thuật ngữ "artificial intelligence" lần đầu xuất hiện trong một văn kiện cấp quốc gia. Tháng 7 năm 2017, State Council ban hành bản quy hoạch tổng thể đầu tiên với một khung cấu trúc rõ ràng. Trong văn kiện mang tên "New Generation Artificial Intelligence Development Plan," State Council đặt mục tiêu đưa Trung Quốc trở thành trung tâm đổi mới artificial intelligence vào năm 2030.

Sau đó, văn kiện viết quy mô tăng trưởng bằng tiền. Đến năm 2020, 150 tỷ yuan trong các ngành cốt lõi và 1 nghìn tỷ yuan trong các ngành liên quan. Đến năm 2025, 400 tỷ yuan trong các ngành cốt lõi và 5 nghìn tỷ yuan trong các ngành liên quan. Đến năm 2030, 1 nghìn tỷ yuan trong các ngành cốt lõi và 10 nghìn tỷ yuan trong các ngành liên quan. Ở mỗi giai đoạn, văn kiện đều chốt một con số. Đặc điểm của các văn kiện State Council nằm ở việc biến mục tiêu từ khẩu hiệu định tính thành các chỉ tiêu số theo từng giai đoạn.

Nếu các văn kiện ban đầu đặt trọng tâm vào hạ tầng và lý thuyết nền tảng, thì hướng đi đã chuyển một lần vào mùa xuân năm 2024. Trong báo cáo công tác chính phủ tháng 3 năm đó, State Council đưa ra thuật ngữ "AI+." Điều này có nghĩa là họ sẽ không còn xem artificial intelligence chỉ là một ngành riêng cần phát triển, mà sẽ gắn nó vào các ngành hiện có như sản xuất, giáo dục và tiêu dùng. Hướng đi đó trở thành văn kiện vào tháng 8 năm 2025, "Opinions on Deepening the Implementation of the 'Artificial Intelligence Plus (AI+)' Action." Tại đây, State Council đưa ra các chỉ số thương mại hóa cho việc đưa artificial intelligence vào sáu lĩnh vực trải từ khoa học và công nghệ, phát triển công nghiệp đến đời sống người dân.

Mục tiêu là nâng tỷ lệ áp dụng các hệ thống thông minh, bao gồm thiết bị đầu cuối thông minh và agent, lên 70 percent vào năm 2027 và 90 percent vào năm 2030. Chuẩn đo mục tiêu đã chuyển từ kết quả phòng thí nghiệm sang mức độ công nghệ được lắp đặt rộng rãi trong thực tế. Chỉ tiêu về tỷ lệ áp dụng này nối trực tiếp với các con số về hạ tầng và tiêu chuẩn được bàn ở phần sau. Để đạt 90 percent, năng lực tính toán, dữ liệu và tiêu chuẩn công nghiệp phải được đặt xuống ở cùng một quy mô.

Đề cương Kế hoạch 5 năm lần thứ 15, được ban hành vào mùa xuân năm 2026, đưa bức tranh này lên thành một cuộc tái cơ cấu toàn bộ nền kinh tế quốc dân. Quốc vụ viện coi "lực lượng sản xuất chất lượng mới", trong đó có trí tuệ nhân tạo và các công nghệ tiên tiến, là một trụ cột tăng trưởng, đồng thời đưa vào một chỉ tiêu nâng tỷ trọng giá trị gia tăng của các ngành cốt lõi trong kinh tế số trong GDP lên 12.5 percent vào năm 2030. Văn kiện này cũng nêu con số yêu cầu đầu tư cho nghiên cứu và phát triển tăng hơn 7 percent mỗi năm. Đây là vị trí của Quốc vụ viện. Cơ quan này truyền định hướng và con số từ trên xuống, nhưng bộ nào sẽ cầm những con số ấy xuống thực địa, và cầm bằng cách nào, lại được quyết định ở tầng dưới. Nếu chỉ đọc văn kiện của Quốc vụ viện, trí tuệ nhân tạo Trung Quốc trông giống một bảng kế hoạch sạch sẽ. Khi

kế hoạch ấy đi vào tay các bộ thật sự và các máy chủ của doanh nghiệp, câu chuyện trở nên phức tạp hơn nhiều.

Bên dưới tầng đó là NDRC, ngồi ở vị trí điều phối. Ủy ban Cải cách và Phát triển Quốc gia ban đầu là cơ quan phụ trách quy hoạch kinh tế vĩ mô. Cơ quan này xử lý rộng các vấn đề đầu tư, giá cả và chính sách công nghiệp, nên nhiều người gọi đây là trung tâm quyền lực thật sự của chính sách. Theo phân tích của Ủy ban Đánh giá Kinh tế và An ninh Mỹ-Trung (USCC), việc NDRC nổi lên như một cơ quan điều phối chính sách trí tuệ nhân tạo là một thay đổi khá gần đây. Trước năm 2023, NDRC không giữ vai trò nổi bật trong lĩnh vực này. Kế hoạch tổng thể năm 2017 do Bộ Khoa học và Công nghệ dẫn dắt, còn từ năm 2021 đến năm 2023, CAC đứng ở tuyến đầu quản lý bằng cách lần lượt ban hành quy định về khuyến nghị thuật toán và biện pháp quản lý tổng hợp sâu. Trong giai đoạn này, quản lý trí tuệ nhân tạo dựa trên hai trục, thúc đẩy công nghệ và kiểm soát nội dung, được chia cho các bộ khác nhau.

Tác nhân làm thay đổi tình hình xuất hiện sau khi các mô hình tạo sinh dưới dạng chatbot lan rộng trên toàn thế giới. Một nhận định dần hình thành: chính sách trí tuệ nhân tạo không thể để mặc cho một cơ quan chủ yếu giỏi kiểm duyệt và kiểm soát nội dung. Để thắng trong cạnh tranh toàn cầu, Trung Quốc cần một lực đẩy về phía trước. Để bảo vệ sự ổn định của chế độ, nước này cũng cần một lực ép xuống. Một số phân tích ở nước ngoài kết luận rằng Trung Quốc cần cân bằng lại hai lực này qua lăng kính quy hoạch kinh tế vĩ mô, và NDRC được giao vào vị trí đó.

Tuy vậy, có một điểm cần thận trọng. Các văn kiện công khai không cho phép kết luận chắc chắn rằng NDRC một mình nắm toàn quyền hay chủ trì một ủy ban điều phối gắn kết nhiều bộ lại với nhau. Chỉ có thể suy luận rằng khả năng thứ hai cao hơn. Những tường thuật gắn vai trò điều phối này một cách dứt khoát với tên của một cá nhân cụ thể vẫn dựa trên nền tảng yếu. Thay vì nêu một người và chức danh như sự thật đã được xác lập, các tài liệu hiện có phù hợp hơn với cách chỉ ghi nhận sự dịch chuyển của cơ quan này sang một vị trí mới.

Việc lập trường của cơ quan này thay đổi cho thấy trọng tâm trong chính sách trí tuệ nhân tạo đã dịch chuyển đến đâu. Mục tiêu không chỉ là nâng hiệu năng mô hình, mà còn là quản lý những xáo trộn mà sự lan rộng của nó có thể gây ra trong việc làm và nền kinh tế thực. Mục tiêu xây dựng một hệ thống đánh giá toàn dụng lao động chất lượng cao trong Kế hoạch 5 năm lần thứ 15, được đặt trong cùng một văn kiện với mục tiêu phổ cập trí tuệ nhân tạo, cũng có thể được đọc như một dấu vết của sự phối hợp này. Đó là một cấu trúc cố gắng nắm cùng lúc cả lực đẩy tiến lên và hệ thống chống đỡ. Nếu trí tuệ nhân tạo lan rộng nhanh, việc làm bị chao đảo. Nếu nhà nước chỉ tập trung bảo vệ việc làm, quá trình áp dụng sẽ chậm lại. Việc đặt cả hai trong cùng một văn kiện cho thấy tính chất của vị trí mà NDRC đang nắm giữ.

Sự phối hợp này có thể nhìn thấy trên thực địa ở hai lĩnh vực: hạ tầng tính toán và dữ liệu. Để đạt mục tiêu của State Council về tỷ lệ áp dụng 90 phần trăm, Trung Quốc cần năng lực tính

toán ở quy mô tương ứng. Báo cáo so sánh giữa Hoa Kỳ và Trung Quốc của Samjong KPMG cho thấy NDRC xây dựng khung này thông qua một dự án quốc gia mang tên "Eastern Data, Western Computing." Đúng như tên gọi, dữ liệu phát sinh ở miền đông được xử lý ở miền tây. Dự án xây dựng các trung tâm dữ liệu tại tám cụm, gồm Jing-Jin-Ji, Yangtze River Delta, Guizhou và Inner Mongolia.

Nhưng dự án này không chỉ là xây dựng. Nó chặn đầu tư trùng lặp thiếu kiểm soát và áp đặt các quy tắc về hiệu quả sử dụng điện, qua đó cho phép chính phủ trung ương kiểm soát cách vốn và năng lượng tập trung. Đầu tư của chính quyền địa phương vào hạ tầng cũng được đưa vào hệ thống phê duyệt quốc gia và đặt dưới sự quản lý của NDRC. Hai bàn tay tương tự cũng xuất hiện ở đây: vừa đặt hạ tầng, vừa kiểm chế tình trạng quá nóng. Có bao nhiêu trung tâm dữ liệu được xây, xây ở đâu, và đạt mức hiệu quả điện năng nào, tất cả đều đi qua một cửa phê duyệt của cùng một cơ quan.

Ở phía dữ liệu, quy hoạch của NDRC cũng đi trước. Theo một báo cáo của USCC, vào tháng 3 năm 2025, NDRC cùng với National Data Administration, Ministry of Finance và Ministry of Human Resources and Social Security đã ban hành các ý kiến triển khai về phát triển ngành gán nhãn dữ liệu. Huấn luyện các mô hình quy mô lớn đòi hỏi dữ liệu được tổ chức tốt. Bức tranh ở đây là nhiều bộ phối hợp với nhau về lực lượng lao động gán nhãn tạo ra dữ liệu đó và về lưu thông dữ liệu công. Số liệu của National Data Administration nói rằng hơn 60 phần trăm hoạt động huấn luyện mô hình được thực hiện bên trong Trung Quốc sử dụng dữ liệu tiếng Trung.

Bên dưới các văn kiện này là quan điểm cho rằng dữ liệu được thu thập và lưu chuyển ở đâu, bằng cách nào, chính là vạch xuất phát của cuộc cạnh tranh mô hình. Tuy vậy, chính sách phát triển ngành gán nhãn này về sau còn phải được đọc qua một lăng kính khác. Bề mặt của một văn kiện về phát triển công nghiệp không cho thấy lao động đặt tên cho dữ liệu diễn ra trong điều kiện nào, hoặc ai là người gánh phần việc ấy. Một chương sau sẽ quay lại vấn đề này cùng với các bằng chứng đối lập.

Khi hướng đi lớn đã được xác lập và khung hạ tầng cùng dữ liệu đã có, chính sách đi xuống thực địa qua từng bộ ngành cụ thể. Bản tóm tắt của KIAT về "Kế hoạch thực hiện Kế hoạch hành động đặc biệt về Trí tuệ nhân tạo Plus Sản xuất" cho thấy tám bộ ngành, trong đó có MIIT, đã đặt mục tiêu cho năm 2027: đào sâu việc áp dụng ba hoặc bốn mô hình ngôn ngữ lớn trong sản xuất và đưa ra 1.000 tác nhân AI công nghiệp. Kế hoạch này muốn cắm các mô hình điều khiển tự chủ vào những ngành như thép và chế tạo thiết bị, lập vị trí giám đốc dữ liệu trong doanh nghiệp, và xây dựng một dòng chảy trong đó dây chuyền sản xuất tự tạo dữ liệu rồi tinh chỉnh mô hình.

Những con số mục tiêu như 1.000 tác nhân này cần được kiểm tra lại với văn bản gốc, nên ở đây chúng chỉ được xem là mục tiêu chính sách. Cách chính xác hơn là không làm mờ sự thật rằng đó là các chỉ tiêu do bộ ngành đặt ra, chưa phải kết quả đã được xác nhận.

Bây giờ chuyển sang CAC. Cục Quản lý Không gian mạng Trung Quốc là cơ quan giám sát quy định internet và kiểm duyệt. Có thể hiểu dễ nhất cơ quan này như người gác ở cánh cửa mà các mô hình phải đi qua để bước vào sử dụng công khai dưới dạng dịch vụ. Khi đọc cùng phần giải thích của Jeremy Daum về dự thảo biện pháp quản lý AI tạo sinh từ Paul Tsai China Center của Yale Law School và các công trình học thuật phân tích chiến lược tiêu chuẩn hóa của Trung Quốc, có thể thấy quyền kiểm soát của CAC không dừng ở việc kiểm duyệt phần mềm đã hoàn thiện. Cơ quan này xem toàn bộ vòng đời của trí tuệ nhân tạo, từ thu thập dữ liệu đến huấn luyện mô hình, triển khai và giám sát, là đối tượng quản lý. Khung quản lý vòng đời này là sợi chỉ nối các quy định khác nhau của CAC lại với nhau.

Tiếp theo là tầng mô hình. Điều 6 của dự thảo biện pháp về AI tạo sinh yêu cầu nhà cung cấp dịch vụ đăng ký thuật toán của mình với cơ quan quản lý. Dựa trên các quy định về quản lý khuyến nghị thuật toán trong dịch vụ thông tin internet, điều này buộc doanh nghiệp phải báo cáo trước cấu trúc và phương thức vận hành của thuật toán. Điều 7 quy định rằng nguồn dữ liệu dùng để huấn luyện phải hợp pháp. Quy định này thu hẹp cửa vào của dữ liệu bằng cách hạn chế việc thu thập từ những nguồn tùy ý, đồng thời buộc mô hình tạo ra trí tuệ bên trong một hệ tri thức đã được phê duyệt. Điều 8 cũng yêu cầu các quy tắc chi tiết cho quy trình con người gán nhãn dữ liệu bằng tay. Mục tiêu là ngăn thiên kiến của người gán nhãn trộn vào dữ liệu. Thiết kế này khiến dữ liệu phải đi qua nhiều cổng trước khi bước vào mô hình.

Trong các quy định này, Điều 15 có sức nặng riêng. Nếu phát hiện trí tuệ nhân tạo đã tạo ra nội dung bất hợp pháp, doanh nghiệp không thể chỉ xóa kết quả đó rồi coi như xong. Doanh nghiệp phải kiểm tra lại mô hình và huấn luyện lại để loại bỏ nguyên nhân gốc. Điều này không có nghĩa là lọc bỏ một câu trả lời có vấn đề. Nó đòi hỏi doanh nghiệp sửa cấu trúc tham số đã cho phép câu trả lời ấy xuất hiện. Theo nghĩa đó, quy định nhắm vào bên trong mô hình, chứ không chỉ nhắm vào đầu ra, và điều này có sắc thái khác với cách lọc sau sự việc ở phương Tây.

Dù vậy, thay vì kết luận rằng bên nào căn bản hơn bên nào, cách chính xác hơn là giữ sự phân biệt ở mức này: đối tượng điều chỉnh là khác nhau. Huấn luyện lại rất tốn kém, và chi phí đó rơi vào các doanh nghiệp đang cố tuân thủ quy định. Điều này nối với vấn đề gánh nặng doanh nghiệp được bàn ở phần sau.

Một khi mô hình đi qua cổng kiểm soát và đến với người dùng, CAC đặt trách nhiệm nặng nề lên nội dung. Điều 5 của các biện pháp này buộc nhà cung cấp dịch vụ chịu trách nhiệm về kết quả do AI tạo ra như thể họ là người sản xuất đã trực tiếp viết ra nội dung đó. Theo Điều 11, nhà cung cấp cũng chịu trách nhiệm xử lý thông tin cá nhân của người dùng. Theo Điều 12, mô hình không được tạo ra đầu ra mang tính phân biệt đối xử dựa trên giới tính, sắc tộc hoặc đức tin, và theo Điều 14, mô hình phải duy trì dịch vụ an toàn, ổn định. Theo Điều 13, nhà cung cấp cũng phải duy trì kênh tiếp nhận khiếu nại của người dùng và nhanh chóng sửa thông tin xâm phạm quyền. Trí tuệ nhân tạo có thể tạo ra ngôn ngữ giống con người, nhưng các quy định này

dựa trên tiền đề rằng người vận hành doanh nghiệp là con người phải chịu trách nhiệm về những lời đó.

Trong kiểm soát nội dung, nghĩa vụ gắn nhãn cần được nhìn thật kỹ. Điều 16 yêu cầu mọi nội dung do AI tạo ra phải có nhãn nhận dạng cho biết nội dung đó do máy tạo ra, đồng thời cấm người dùng tùy tiện gỡ bỏ hoặc che giấu nhãn này. Đây là một công cụ nhằm ngăn con người làm mờ ranh giới giữa sản phẩm do con người tạo ra và sản phẩm do máy tạo ra. Việc các nền tảng lớn như Douyin sớm ban hành quy tắc riêng về nội dung do AI tạo ra có thể được xem là một phần của cùng dòng chảy đó. Nhà nước đặt ra quy tắc, các nền tảng đưa chúng vào chuẩn mực nội bộ của mình, rồi những chuẩn mực ấy đi xuống tận màn hình của từng người dùng.

Mắt xích cuối trong kiểm soát của CAC là người dùng. Điều 9 yêu cầu xác minh danh tính đối với mọi người dùng dịch vụ AI tạo sinh. Dựa trên Luật An ninh mạng, cơ chế xác thực tên thật này đã trở thành một hệ thống chuẩn trong Trung Quốc, gắn hồ sơ hoạt động trực tuyến với giấy tờ định danh đã được xác minh. Kết quả là tính ẩn danh gần như bị xóa bỏ trên thực tế. Ngay cả khi một người dùng bút danh hoặc tên hiển thị trên diễn đàn, đó chỉ là hình thức xuất hiện trước công chúng. Nền tảng và cơ quan chức năng có thể dùng dữ liệu đăng ký ở hệ thống phía sau để biết ngay ai đã viết bài đăng đó.

Thông tin tên thật cũng được dùng để nhận diện người chưa thành niên, tạo cơ sở cho các nền tảng chặn thanh thiếu niên tiếp cận nội dung có hại. Nhà cung cấp dịch vụ phải xác định trước người dùng và mục đích sử dụng dự kiến của mình (Điều 10), đồng thời phải đặt ra các biện pháp kỹ thuật để ngăn người dùng dựa quá nhiều vào AI hoặc bị cuốn vào nó (Điều 18). Với người dùng vi phạm quy tắc, dịch vụ tài khoản phải bị dừng ngay lập tức (Điều 19). Từ điểm đầu vào của dữ liệu đến danh tính của người dùng, cấu trúc này nối từng biện pháp với biện pháp kế tiếp như một sợi xích.

Các biện pháp này cũng nêu rõ điều gì xảy ra khi những nghĩa vụ dày đặc ấy bị vi phạm. Nếu một luật khác đã quy định chế tài, luật đó được áp dụng trước. Nhưng ngay cả khi không có quy định rõ ràng, doanh nghiệp cũng không có lối thoát. Cơ quan chức năng có thể đưa ra cảnh báo, phê bình công khai, lệnh khắc phục, đình chỉ dịch vụ, và có thể phạt tiền từ 10.000 nhân dân tệ đến 100.000 nhân dân tệ. Bản thân mức phạt có thể chưa đủ lớn để tạo gánh nặng cho một công ty lớn. Nhưng bức tranh sẽ khác khi nó đi kèm với lệnh dừng toàn bộ dịch vụ.

Đối với một mô hình phục vụ công chúng, chính hoạt động kinh doanh sẽ dừng lại ngay khi cánh cửa bị đóng. Đây là lý do mang tính cấu trúc khiến các công ty như Alibaba, DeepSeek, Zhipu và MiniMax đầu tư nguồn lực vào kiểm duyệt và tuân thủ với mức độ cẩn trọng không kém gì việc nâng điểm chuẩn đánh giá. Các thông báo đăng ký dịch vụ cho thấy số lượng mô hình hoàn tất quy trình này và được đăng ký vẫn tiếp tục tăng, và mỗi mục trong danh sách đó cũng là bằng chứng cho thấy mô hình đã đi qua các quy định nêu ở trên.

Nếu CAC canh giữ cửa trước, thì MIIT và SAC đặt ra các thông số bên trong cánh cửa đó. Ministry of Industry and Information Technology và Standardization Administration of China là những cơ quan xây dựng tiêu chí an toàn và tiêu chuẩn công nghiệp. Từ "standard" nghe như một sổ tay kỹ thuật trung lập, nhưng trong quản trị AI của Trung Quốc, các tiêu chuẩn đóng vai trò như người gác cổng, quyết định ai đủ điều kiện bước vào thị trường. Nghiên cứu học thuật phân tích các văn bản tiêu chuẩn hóa gọi quy trình này là một "state-led, enterprise-driven pipeline".

Khi các ủy ban kỹ thuật thuộc SAC đặt chương trình nghị sự, các nhóm công tác thuộc MIIT soạn thảo và thử nghiệm tiêu chuẩn phản ánh nhu cầu từ thực địa, rồi cơ quan quản lý dùng các thông số kỹ thuật đó làm công cụ giám sát. Nhìn bề ngoài, chúng mang hình thức các quy tắc tự nguyện. Tuy nhiên, sau khi đi qua phê duyệt của nhà nước, chúng trở thành công cụ trên thực tế để kiểm soát việc gia nhập thị trường. Khoảng cách giữa từ "voluntary" và thực tế kiểm soát cho thấy vị trí của tiêu chuẩn kiểu Trung Quốc.

Cấu trúc hợp tác này được chống đỡ bằng các văn bản ban hành trong nhiều năm. Năm 2019, SAC công bố sách trắng về tiêu chuẩn hóa an toàn AI, nối việc đo lường kỹ thuật về an ninh với thực thi hành chính. Tháng 7 năm 2020, cơ quan này ban hành Guidelines for the Establishment of a National New Generation Artificial Intelligence Standards System. Các hướng dẫn này đề xuất xây dựng một hệ thống tiêu chuẩn theo tám trục, từ các lĩnh vực cơ bản và dùng chung như thuật ngữ, kiến trúc tham chiếu, kiểm thử và đánh giá, đến các lĩnh vực công nghệ cốt lõi như phát triển thuật toán và mô hình. Chúng đặt ra nguyên tắc áp dụng trước các tiêu chuẩn cấp bách, đồng thời thúc đẩy tích hợp liên ngành, và đặt đối mới độc lập bên cạnh hợp tác mở. Điều này có nghĩa là nhà nước không để tiêu chuẩn vận hành theo riêng logic thị trường. Nhà nước quyết định điều gì cần được tiêu chuẩn hóa trước, và theo thứ tự nào.

Nếu SAC định hình hệ thống, thì MIIT biến hệ thống đó thành các quy tắc cho xưởng sản xuất. Kế hoạch triển khai cho Kế hoạch hành động đặc biệt về "Artificial Intelligence + Manufacturing" đặt tiêu chuẩn lên trước, coi đó là cách để giữ vị thế dẫn đầu trong hệ sinh thái công nghiệp. MIIT làm việc với ủy ban kỹ thuật tiêu chuẩn hóa AI và ủy ban tiêu chuẩn hóa dữ liệu của mình, cùng với ủy ban kỹ thuật tiêu chuẩn hóa an ninh mạng trực thuộc SAC, để phân loại các ngành theo cấp độ và loại hình, rồi xây dựng quy chuẩn. Cơ quan này chuẩn bị các tiêu chuẩn nền tảng về an ninh, quản trị và đạo đức, các tiêu chuẩn chung cho hợp tác giữa phần mềm và phần cứng, cùng các chuẩn đo lường để đánh giá hiệu suất tại hiện trường. Yêu cầu doanh nghiệp thiết lập hệ thống giám đốc dữ liệu trưởng cũng gắn với nỗ lực này, vì nó tạo lớp nền quản trị dữ liệu để các tiêu chuẩn quốc gia có thể được tuân thủ.

Mức độ chi tiết của các tiêu chuẩn này hiện rõ trong những mẫu báo cáo mà chính phủ yêu cầu doanh nghiệp nộp. Chính phủ đính kèm các phụ lục có tiêu đề danh sách mô hình công nghiệp trọng điểm và danh sách bộ dữ liệu đánh giá, dùng biểu mẫu để quy định từng doanh

nghiệp phải điền những tham số và chỉ số hiệu suất nào. Trong ngành thép, một mô hình thép lớn với 13 tỷ tham số phải có khả năng dự báo xu hướng nhiệt độ gang lỏng trong lò cao và tối ưu hóa các tham số của quy trình cán, đồng thời độ trễ suy luận trung bình phải duy trì dưới 80 mili giây. Tài liệu này cũng nêu mục tiêu rằng mô hình đó phải nâng hiệu suất công việc tối ưu hóa của nhân sự quy trình lên năm lần.

Trong chế tạo thiết bị, một bộ dữ liệu đánh giá dùng để dự báo hỏng hóc máy công cụ phải đạt tiêu chuẩn trong bài kiểm tra gồm 500 cặp hỏi đáp: độ chính xác tối thiểu 95 phần trăm trong nhận diện lỗi văn bản và độ trễ suy luận không quá 50 mili giây. Một doanh nghiệp áp dụng tiêu chuẩn này phải chứng minh bằng hồ sơ rằng họ đã nâng độ chính xác dự báo mài mòn ổ trục chính từ 85 phần trăm lên 94 phần trăm. Vì tiêu chuẩn ghi mục tiêu bằng mili giây và phần trăm, doanh nghiệp cuối cùng sẽ lái hoạt động phát triển theo các con số đó. Một văn bản của chính phủ đặt ra các chỉ số hiệu suất xuống tận hiệu năng mô hình bên trong một nhà máy thép.

Tiêu chuẩn an toàn cũng thuộc trách nhiệm của họ. Trong các nhiệm vụ thiết lập an toàn theo kế hoạch triển khai, MIIT và các cơ quan trực thuộc xây dựng cơ chế quản trị an ninh để ngăn ngừa tai nạn. Họ xác định phát hiện tính xác thực của deepfake, an ninh thuật toán của mô hình công nghiệp, bảo vệ dữ liệu huấn luyện và đánh giá an ninh thiết bị đầu cuối thông minh là các công nghệ cốt lõi, rồi định hướng phát triển các công nghệ này. Để giảm rủi ro ảo giác trong các mô hình tạo sinh, họ yêu cầu doanh nghiệp tinh chỉnh cơ sở tri thức và sửa lỗi trong kho ngữ liệu huấn luyện. Các quy định yêu cầu gắn dấu nhận diện trên nội dung tổng hợp cũng đi theo cùng hướng đó. MIIT phân loại và xếp cấp AI dùng trong công nghiệp theo mức rủi ro, rồi ban hành các tiêu chuẩn an ninh bao gồm đánh giá, kiểm chứng và ứng phó khẩn cấp. Cơ quan này tách các ứng dụng rủi ro cao khỏi các ứng dụng rủi ro thấp hơn và áp dụng các mức độ quản lý khác nhau.

Ở tuyến đầu của quyền lực tiêu chuẩn đang mở rộng này là các tác tử công nghiệp. Tài liệu xuất hiện ở đầu chương này, Basic Safety Requirements for Smart Agent Applications, là một ví dụ. Hồ sơ đăng ký từ State Administration for Market Regulation và SAC cho thấy bản dự thảo cơ bản của tiêu chuẩn này đang được chuẩn bị với sự tham gia của CAC với tư cách một cơ quan, còn Cybersecurity Standardization Technical Committee thuộc SAC giữ vai trò chủ trì. Việc các viện nghiên cứu quốc gia, các trung tâm ứng phó khẩn cấp và những công ty viễn thông lớn nằm ở lõi của nhóm soạn thảo cho thấy rõ đường ống vận hành: nhà nước nắm chương trình nghị sự, đồng thời rút lấy hiểu biết kỹ thuật của doanh nghiệp để tạo ra tiêu chuẩn. Cấu trúc trong đó đối tượng bị điều chỉnh và người viết quy định cùng ngồi trong một văn bản chính là điều mà giới học thuật gọi bằng cụm từ 'state-led and enterprise-driven.'

Nội dung của tiêu chuẩn này rất chi tiết. Nó đặt toàn bộ vòng đời dưới sự kiểm soát, từ thiết kế các tác tử gọi công cụ cục bộ trên điện thoại di động hoặc trên đám mây để thực hiện nhiệm

vụ, cho đến phát triển, triển khai và bảo trì. Xác minh danh tính người dùng, lệnh gọi quyền hệ thống, và quy tắc về thu thập cũng như sử dụng dữ liệu đều được đưa vào đặc tả; các thao tác rủi ro cao đòi hỏi con người can thiệp trực tiếp để ngăn tác tử tự hành động. Tiêu chuẩn này cũng đặt ra tiêu chí chặn và tắt khẩn cấp khi xuất hiện tình huống bất thường, đưa vào đặc tả một cơ chế cho phép nhà chức trách hoặc nền tảng lập tức dừng các tác tử đã vượt khỏi tầm kiểm soát.

Điều này cho thấy đối tượng của việc đặt tiêu chuẩn đang chuyển từ chatbot tạo văn bản sang các hệ thống tự chủ vận hành công cụ thật. Khi những hệ thống tự phán đoán và gọi công cụ trở nên phổ biến hơn, Trung Quốc dường như đang cố đặt sẵn quy tắc về cách dừng chúng lại.

Dòng chảy tương tự đang lan sang các tiêu chuẩn trong những lĩnh vực khác. Trong giao thông vận tải, 'AI + Transportation Application Safety Guidelines' đang được soạn thảo, trong khi các tài liệu như hướng dẫn an toàn cho ứng dụng nhắm tới trẻ vị thành niên, hướng dẫn về ứng dụng công nghệ trí tuệ nhân tạo trong xe, và hướng dẫn lựa chọn, áp dụng mô hình lớn với các tiêu chí chọn foundation models đang đi qua quy trình ban hành cùng các số hiệu đã được giao. Thông qua một hình thức duy nhất là tiêu chuẩn, các quy tắc an toàn đang tỏa ra nhiều khu vực của xã hội. Khi những đặc tả khởi đầu từ tác tử công nghiệp lan sang giao thông, giáo dục và ô tô, một ngưỡng mới được đặt ra ở bất cứ nơi nào trí tuệ nhân tạo vươn tới.

Đây là lúc chúng ta cần nhìn vấn đề từ một góc khác trong giây lát. Bức tranh về năm cơ quan được phác họa đến đây trông giống một cấu trúc trơn tru, trong đó nhà nước đặt ra hướng đi và các bộ chia nhau phần thực thi. Nhưng có lý do khiến các nghiên cứu phân tích tài liệu tiêu chuẩn dùng cụm từ 'enterprise-driven'. Việc doanh nghiệp tham gia vào chính bàn tay viết nên nội dung thực tế của tiêu chuẩn cũng có nghĩa là quản trị không chỉ chảy từ trên xuống dưới. Doanh nghiệp hiểu các chi tiết kỹ thuật, và nhà nước phải mượn sự hiểu biết ấy để tạo ra các quy cách vận hành được trong thực tế. Nhà nước canh giữ cánh cửa, nhưng đo khung cửa cùng với doanh nghiệp. Nếu chỉ đọc sự tương tác này qua một chữ kiểm duyệt, chúng ta sẽ bỏ lỡ cả ảnh hưởng mà doanh nghiệp có được khi tham gia làm luật chơi, lẫn gánh nặng mà họ phải nhận.

Đến đây, chúng ta đã lần theo năm cơ quan qua các văn bản. The State Council truyền xuống định hướng và con số từ cấp trên. NDRC điều phối cả lực đẩy lẫn sự hỗ trợ giữa các bộ. CAC canh giữ lối vào bằng cách buộc mô hình, nội dung và người dùng vào cùng một chuỗi. MIIT và SAC đặt các quy cách bên trong cánh cửa ấy bằng mili giây, tỷ lệ phần trăm và tiêu chí tắt khẩn cấp. Nếu đọc cấu trúc này, nơi năm cơ quan ăn khớp với nhau như bánh răng, như bức tranh một cơ quan kiểm duyệt duy nhất ép xuống từ trên cao, chúng ta sẽ chỉ thấy một nửa.

Trên thực tế, bàn tay đặt hướng đi, bàn tay điều phối, bàn tay canh cửa và bàn tay đo quy cách được chia ra. Giữa các bàn tay ấy, doanh nghiệp đôi khi ngồi vào vị trí người viết tiêu

chuẩn. Khoảng cách giữa chính phủ và doanh nghiệp không phải là khoảng cách giữa một người quan sát ở xa và người bị quan sát. Nó gần hơn với khoảng cách giữa những cái tên được liệt kê cạnh nhau trong cùng một văn bản.

Có một điểm đáng chú ý: cấu trúc này đặt thúc đẩy và kiểm chế trong cùng một văn bản. Mục tiêu áp dụng 90 phần trăm của The State Council và hệ thống đánh giá toàn dụng lao động xuất hiện cùng nhau trong một kế hoạch. Việc NDRC mở rộng hạ tầng Eastern Data, Western Computing và kiểm chế đầu tư quá nóng xuất hiện cùng nhau trong một chương trình. Việc CAC mở dịch vụ và kiểm soát tên thật đứng cạnh nhau trong một biện pháp. Thiết kế cố nắm phát triển và an ninh trong cùng một bàn tay này là khuôn khổ của quản trị kiểu Trung Quốc. Mục tiêu làm cho một ngành tăng trưởng nhanh và sự kiểm soát nhằm giữ ngành ấy bên trong hệ thống không được viết trong các văn bản riêng, mà nằm ở những đoạn văn liền kề trong cùng một văn bản.

Dù vậy, vẫn có những điểm khó khăn định dứt khoát. Các tài liệu công khai không xác nhận rõ vai trò điều phối của NDRC là quyền chỉ huy độc lập hay là một hình thức ủy ban, cũng không cho biết cá nhân nào phụ trách việc điều phối đó. Những con số mục tiêu như 1,000 tác nhân sản xuất, cùng các chỉ số hiệu suất trong từng tiêu chuẩn ngành, là các giá trị cần được đối chiếu lại với văn bản gốc. Các chỉ số do State Council đưa ra, như tỷ lệ áp dụng và tỷ trọng giá trị gia tăng, cũng phải được đọc sao cho không làm mờ ranh giới giữa những con số được ghi như mục tiêu và kết quả thực tế đã đạt được. Và những gánh nặng mà các ngưỡng do năm cơ quan này đặt ra gây cho doanh nghiệp, mức độ kiểm duyệt và kiểm soát kéo dài đến đâu, cũng như việc gắn tên người vào dữ liệu diễn ra trong điều kiện nào, đều phải được xem xét cùng với bằng chứng từ phía đối lập.

Nếu chỉ đọc các tài liệu do những người đặt ra các ngưỡng viết, chúng ta không thể thấy chi phí để vượt qua chúng, cũng không thấy vị trí của những người bị đẩy ra ngoài các ngưỡng ấy. Câu chuyện đó tiếp tục khi chúng ta đọc lại chuẩn mực kiểu Trung Quốc, vốn nói về phát triển và an ninh cùng lúc, qua con mắt của kiểm soát dữ liệu và các bằng chứng phản bác nó.

Chương 3. Các quy tắc kiểm soát dữ liệu được xây dựng giữa phát triển và an ninh

Mở danh sách thông báo của Cục Quản lý Không gian mạng Trung Quốc, hay CAC, một văn bản được đăng vào mùa xuân năm 2026 sẽ nổi bật ngay. Tiêu đề của văn bản là "Thông báo về thông tin đăng ký đối với các dịch vụ trí tuệ nhân tạo tạo sinh, từ tháng 3 đến tháng 4 năm 2026." Bên trong, tên công ty và tên dịch vụ xuất hiện nối tiếp nhau. Danh sách cho biết công ty nào được phép phát hành chatbot và trình tạo hình ảnh dưới tên nào được đăng như một văn bản chính thức trên trang web của một cơ quan nhà nước.

Khi một dịch vụ AI mới xuất hiện ở Hàn Quốc hoặc Hoa Kỳ, công ty đưa ứng dụng lên mạng và người dùng tải xuống. Đó thường là bước đầu tiên. Ở Trung Quốc, trước đó còn có một cánh cửa khác. Trước khi một dịch vụ đi ra thị trường, công ty phải đăng ký thuật toán và dịch vụ với cơ quan chức năng, rồi đưa tên mình vào danh sách thông báo công khai này. Một dịch vụ không có tên trong danh sách thì không thể tồn tại trên thị trường, và nếu tên bị xóa khỏi danh sách, dịch vụ đó biến mất khỏi thị trường.

Thông báo đăng ký ngắn này là điểm xuất phát cho câu chuyện mà chương này theo dõi. Việc đưa một cái tên vào danh sách có thể trông giống một bước thủ tục giấy tờ, nhưng phía sau nó là một chuỗi yêu cầu dài, kéo từ thu thập dữ liệu đến huấn luyện mô hình, tạo nội dung và quản lý người dùng. DeepSeek, Alibaba, Zhipu và MiniMax đều phải đi qua chuỗi này trước khi cạnh tranh về điểm chuẩn benchmark. Và chuỗi này được xây trên một nguyên tắc.

Nguyên tắc đó là thúc đẩy phát triển nhưng không bao giờ buông lỏng an ninh. Chương này lần lượt theo dõi cách nguyên tắc ấy trở thành một ngưỡng cửa trong các văn bản chính thức, ai ngồi ở những vị trí tạo ra ngưỡng cửa đó, và điều gì hiện ra khi nhìn ngưỡng cửa ấy từ phía bên kia. Nó bắt đầu từ một thông báo đăng ký, rồi đi xuống các quy tắc, triết lý và những lập luận phản biện chống đỡ cho nó.

Các quy định về AI tạo sinh đòi hỏi gì ở doanh nghiệp

Một công ty phải chuẩn bị gì để được ghi tên vào danh sách đăng ký? Câu trả lời cơ bản nằm trong "Dự thảo Biện pháp quản lý dịch vụ trí tuệ nhân tạo tạo sinh" của CAC. Trước hết cần nói rõ một điểm. Văn bản này là bản dự thảo. Một số cách diễn đạt có thể đã được làm mềm đi, hoặc một số điều khoản có thể đã thay đổi giữa bản dự thảo và các quy định chính thức sau đó. Vì vậy, phần trình bày dưới đây nên được đọc như phần mô tả các yêu cầu ở giai đoạn dự thảo.

Những chi tiết như số điều khoản và mức tiền phạt nên được để lại như các mục cần đối chiếu lại với nguyên văn. Dù vậy, vẫn có lý do rõ ràng để theo dõi bản dự thảo này. Nó cho thấy bản đồ những điều cơ quan quản lý lo ngại và những trách nhiệm họ muốn chuyển sang cho doanh nghiệp. Độ cao của ngưỡng mà doanh nghiệp thật sự phải vượt qua có thể thay đổi từng chút

trong các quy định chính thức, nhưng bản dự thảo đã đánh dấu sẵn ngưỡng đó được đặt ở đâu.

Yêu cầu đầu tiên mà bản dự thảo này đặt lên doanh nghiệp liên quan đến ngưỡng đối với dữ liệu huấn luyện. Dữ liệu dùng để dạy một mô hình phải đến từ các nguồn hợp pháp. Trên mặt chữ, điều này có thể nghe như một lời kêu gọi tôn trọng bản quyền, nhưng trong bối cảnh quản lý của Trung Quốc, câu này được hiểu rộng hơn. Nó có nghĩa là doanh nghiệp phải ngăn vật liệu không mong muốn bị trộn vào văn bản web được thu thập tự động rồi thấm vào mạng nơ-ron của mô hình ngay từ đầu vào. Sau cùng, một mô hình sẽ giống với dữ liệu mà nó đã hấp thụ. Vì thế, quản lý không dừng ở việc kiểm tra đầu ra. Nó truy ngược đầu ra về cửa vào dữ liệu đã tạo ra đầu ra đó.

Quy trình gán nhãn, trong đó con người trực tiếp gán thẻ cho dữ liệu, cũng đi kèm yêu cầu rằng doanh nghiệp phải tự đặt ra và áp dụng các quy tắc của mình. Doanh nghiệp phải quản lý quy trình này sao cho phán đoán của từng người gán nhãn không đông cứng thành khuynh hướng của mô hình. Từ cửa vào của dữ liệu cho đến công việc do bàn tay con người thực hiện, bàn tay của quy phạm vươn sâu vào phần hậu trường. Khi các công ty phương Tây thu thập dữ liệu để nâng hiệu suất mô hình, các công ty Trung Quốc phải thu thập cùng loại dữ liệu đó, đồng thời kiểm tra xem nó có vượt qua cuộc kiểm tra tư tưởng hay không.

Ngưỡng thứ hai là đăng ký thuật toán. Các nhà cung cấp dịch vụ phải đăng ký phương thức vận hành và cấu trúc thuật toán của mình với cơ quan chức năng. Danh sách thông báo đăng ký mà chúng ta đã thấy ở trên là kết quả của thủ tục này. Cấu trúc này chỉ cho phép một dịch vụ bước vào thị trường sau khi nhà nước đã nắm được doanh nghiệp đã tạo ra gì và tạo ra nó như thế nào. Đây không phải là cách làm hậu kiểm, trong đó cơ quan chức năng điều tra sau khi dịch vụ ra mắt rồi phát sinh vấn đề. Đây là cách làm tiền kiểm, nhìn vào bên trong doanh nghiệp trước khi dịch vụ ra mắt.

Vì có cơ chế đăng ký trước này nên danh sách thông báo ở trên mới có thể tồn tại. Danh sách đó vừa là danh sách các dịch vụ được phê duyệt, vừa là danh sách các dịch vụ mà nhà nước đã biết. Việc cứ hai tháng lại có thông báo sắp xếp công ty nào đã hoàn tất đăng ký dưới tên nào và vào thời điểm nào cũng cho thấy quy trình này không phải là một ngưỡng chỉ vượt qua một lần, mà là một cánh cổng liên tục được làm mới.

Ngưỡng thứ ba nằm ở phía người dùng. Dựa trên Cybersecurity Law, dự thảo bắt buộc áp dụng hệ thống xác minh danh tính thật. Ngay cả khi người dùng dùng biệt danh trên giao diện, doanh nghiệp và cơ quan chức năng vẫn có thể xác định người dùng thực tế thông qua thông tin đăng ký ở phần hậu trường. Con đường dùng AI ẩn danh, trên thực tế, bị đóng lại. Bên cạnh đó, doanh nghiệp phải chuẩn bị các biện pháp kỹ thuật để ngăn người dùng dựa quá nhiều vào dịch vụ hoặc bị cuốn vào nó, đồng thời phải công bố dịch vụ dành cho những người dùng nào và được tạo ra để dùng vào việc gì.

Và nếu một người dùng đã được xác minh bằng tên thật vi phạm pháp luật và lạm dụng AI, công ty phải trực tiếp cắt dịch vụ đối với tài khoản đó. Bàn tay giám sát và chặn truy cập đi từ nhà nước đến công ty, rồi từ công ty đến người dùng. Thay vì để nhà nước xem xét từng người dùng một, cấu trúc này giao cho các công ty cửa sổ giám sát, rồi quản lý xem cửa sổ đó có vận hành đúng hay không. Công ty vừa là nhà cung cấp dịch vụ, vừa là người quản trị theo dõi người dùng thay cho nhà nước.

Ngưỡng thứ tư là trách nhiệm đối với các đầu ra do công ty tạo ra. Theo dự thảo, nhà cung cấp dịch vụ chịu cùng mức trách nhiệm đối với nội dung do AI tạo ra như một người sản xuất đã trực tiếp viết văn bản đó. Ngay cả khi mô hình bịa ra thông tin sai, trách nhiệm pháp lý vẫn rơi vào công ty. Dự thảo còn bổ sung yêu cầu rằng thông tin người dùng nhập vào prompt phải được bảo vệ, và đầu ra của mô hình không được phân biệt đối xử theo giới tính, sắc tộc hoặc đức tin. Điều khoản nặng nhất đối với các công ty xuất hiện ở đây. Nếu một AI bị phát hiện đã tạo ra nội dung bị đánh giá là bất hợp pháp, công ty không thể chỉ dừng ở việc lọc bỏ đầu ra đó. Công ty phải huấn luyện lại chính mô hình để loại bỏ gốc rễ của vấn đề.

Nếu nghĩ đến tài nguyên tính toán và thời gian cần để huấn luyện lại một mô hình ngôn ngữ lớn, ta có thể thấy yêu cầu huấn luyện lại này sẽ nặng nề thế nào đối với một công ty. Nếu một chatbot phải được dạy lại từ đầu chỉ vì nó tạo ra một câu sai, chi phí sẽ đủ lớn để đe dọa chính dịch vụ đó. Thêm một bộ lọc và xây dựng lại một mô hình thuộc hai cấp chi phí khác nhau. Cách đọc này dựa vào câu chữ của dự thảo, vì vậy quy định cuối cùng cần được xác nhận riêng để xem điều khoản này đã được chỉnh lại ra sao.

Ngưỡng thứ năm là việc đánh dấu tài liệu được tạo ra. Nội dung do máy tạo ra phải mang một nhãn có thể nhận biết, và người dùng bị cấm tự ý xóa hoặc che giấu nhãn đó. Khi nghĩa vụ gắn nhãn này mới xuất hiện, nó chưa được tuân thủ rộng rãi, nhưng người ta nói rằng nó chỉ thật sự bám rễ trong thực tế sau khi quy định và tiêu chuẩn kỹ thuật được xếp chồng lên nhau qua từng năm. Cách một quy tắc không kết thúc bằng một tuyên bố duy nhất, mà siết lại từng chút một trong nhiều năm, cho thấy chất liệu của quản lý nhà nước kiểu Trung Quốc.

Một yêu cầu lúc đầu trông giống như lời lẽ không có sức cưỡng hành thật sự, theo thời gian lại cứng lại thành một nghĩa vụ cụ thể. Cùng một mô thức cũng xuất hiện trong các quy định khác được bàn ở phần sau. Nếu một công ty nhìn vào một quy định rồi gạt đi vì cho rằng đó chỉ là khẩu hiệu, vài năm sau công ty ấy có thể phát hiện khẩu hiệu đó đã quay lại dưới dạng một điều kiện thật sự, làm thay đổi thiết kế dịch vụ của mình.

Dự thảo cũng nêu rõ cơ quan quản lý có thể làm gì với các công ty vi phạm các ngưỡng này. Các biện pháp hành chính sẽ được áp dụng, gồm cảnh cáo, phê bình công khai, lệnh khắc phục và đình chỉ dịch vụ, đồng thời có thể kèm tiền phạt. Phạm vi tiền phạt chính xác cần được đối chiếu lại với nguyên văn, nhưng dự thảo liệt kê các mức từ khoảng mười nghìn yuan đến khoảng một trăm nghìn yuan. Nặng hơn bản thân số tiền phạt là việc một dịch vụ có thể bị đưa

ra khỏi danh sách và biến mất khỏi thị trường. Việc một cái tên được đưa vào danh sách đăng ký và việc cái tên ấy bị xóa đi là hai mặt của cùng một cánh cửa. Nếu một công ty không vượt qua dù chỉ một trong năm ngưỡng này, phần phát triển mà họ đã đổ nguồn lực vào có thể dừng lại trước khi kịp ra thị trường. Với một công ty, các ngưỡng do quy định đặt ra trước hết là danh sách kiểm tra để sống còn, rồi sau đó mới là bảng tiền phạt.

Nếu chỉ nhìn đến đây, công ty có vẻ là một chủ thể thụ động đứng trước quy định. Nhưng khi tầm nhìn chuyển sang các cơ sở công nghiệp, bức tranh thay đổi một chút. Khi số lượng tác tử công nghiệp có thể tự mình điều khiển nhà máy và thiết bị tăng lên, các ngưỡng mà doanh nghiệp phải vượt qua mở rộng từ hệ tư tưởng sang các con số. "Implementation Opinions on the Standardized Application and Innovative Development of Agents" của State Council là văn kiện đã đưa tác tử công nghiệp vào ngôn ngữ chính sách.

Khi một chatbot trò chuyện với công chúng, quy định lo về nội dung. Khi một tác tử điều khiển máy móc, quy định lo về an toàn. Những tác tử tự đảm nhận quyền điều khiển vật lý phải được thiết kế sao cho con người có thể can thiệp trong các hoạt động rủi ro cao, và phải có cơ chế dừng khẩn cấp trong trường hợp tác tử vượt khỏi tầm kiểm soát.

Đây là nơi bàn tay phát triển nhằm nâng cao hiệu quả gặp bàn tay an ninh có thể dừng hệ thống bất cứ lúc nào, ngay trong cùng một văn bản tiêu chuẩn. Nghĩa vụ huấn luyện lại đối với chatbot công cộng và thiết bị dừng khẩn cấp dành cho tác nhân công nghiệp nhìn bề ngoài có vẻ là hai câu chuyện khác nhau, nhưng cả hai đều viết cùng một câu theo hai cách khác nhau. Chúng tôi nói rằng công nghệ phải có một chỗ để nhà nước có thể đưa tay vào và dừng nó bất cứ lúc nào. Dù là trong đối thoại hay trong điều khiển, yêu cầu ấy không đổi: nhà nước phải nắm công tắc cuối cùng.

Kiểu quản trị Trung Quốc nói cùng lúc về phát triển và an ninh

Đến đây phải đặt ra một câu hỏi. Vì sao Trung Quốc cố quản lý dữ liệu và mô hình chặt đến vậy? Khung trả lời đã được trình bày trong bài phân tích học thuật "China's AI Governance." Bài viết tóm lược cách tiếp cận của Trung Quốc bằng một công thức: thúc đẩy an ninh thông qua phát triển và bảo vệ phát triển thông qua an ninh. Khác với phương Tây, nơi coi AI là công cụ thị trường cho đổi mới và đặt nặng quyền tự chủ, Trung Quốc coi AI là vấn đề sống còn của quốc gia và duy trì chế độ. Vì vậy, quản lý và phát triển được buộc lại thành một dự án chung, thay vì giao cho những cơ quan khác nhau. Cấu trúc này không phải là bộ phận trách phát triển và bộ phận trách an ninh vận hành riêng rẽ rồi va chạm với nhau. Đó là cấu trúc trong đó hai mục tiêu biến nhau thành điều kiện ngay trong cùng một câu.

Bên dưới công thức này là cách giới lãnh đạo Trung Quốc nhìn AI. Họ coi AI là một công nghệ có sức lật đổ, có thể làm rung chuyển trật tự chính trị và kinh tế hiện có, đồng thời chuyển dịch cán cân xã hội. Nhưng nét định hình của kiểu quản trị này là nhà nước không tìm cách đàn áp

và xóa bỏ sức mạnh nguy hiểm ấy. Nhà nước cố thuần hóa nó và biến nó thành một lực bổ sung cho sức bền và quyền lực của chế độ. Đây không phải là thái độ lùi lại vì mối đe dọa đáng sợ. Đây là thái độ ôm lấy mối đe dọa vì tin rằng nó có thể được kiểm soát.

Các ngưỡng được xem xét ở phần trước là những công cụ cụ thể của tư thế này. Yêu cầu về tính hợp pháp của dữ liệu huấn luyện là bàn tay ngăn sức mạnh lật đổ thắm vào mô hình. Nghĩa vụ huấn luyện lại là bàn tay kéo sức mạnh ấy trở lại khi nó đã vượt qua ranh giới. Xác minh tên thật là bàn tay khiến nhà nước có thể nhận diện, vào bất cứ lúc nào, người đang dùng sức mạnh ấy. Mỗi ngưỡng có thể trông như một quy định rời rạc, nhưng phía sau chúng là một cách nhìn chung: sức mạnh lật đổ phải nằm trong tay nhà nước.

Tư thế này nằm trong khung rộng hơn mà ban lãnh đạo Tập Cận Bình gọi là khái niệm an ninh quốc gia toàn diện. Khái niệm an ninh này được xây thành nhiều tầng, lấy an ninh của nhân dân làm mục đích, an ninh chính trị làm gốc, và an ninh kinh tế làm nền tảng. Quản trị AI cũng vận động trong khung ấy. Văn kiện chỉ ra hai ý tưởng: rủi ro công nghệ không được xem như một lỗi biệt lập mà như một tập hợp rủi ro đan xen, và nỗi bất an mơ hồ mà con người cảm thấy trước AI được xử lý cùng với an toàn thể chế nhằm bảo vệ hạ tầng thực tế.

Có thể đọc văn kiện này như một nỗ lực đưa sức phá hoại của công nghệ vào một chiếc bình chuẩn tắc do nhà nước dẫn dắt. Từ góc nhìn này, ngay cả một câu trả lời riêng lẻ của chatbot cũng nằm trên cùng một bản đồ với an toàn hạ tầng. Việc quản lý các đầu ra nhỏ và sự ổn định của hệ thống lớn hơn được nối với nhau trong cùng một lối nghĩ. Việc nổi bất an tâm lý trong dân chúng và sự bảo vệ thể chế đối với hạ tầng được xử lý song song trong một văn kiện cho thấy kiểu quản trị này xem không chỉ hiệu năng công nghệ mà cả tâm lý xã hội là đối tượng quản lý.

Giữa hai trục phát triển và an ninh, phía phát triển được truyền xuống qua những chỉ tiêu số dày đặc. Nâng tỷ trọng kinh tế số lên đến đâu, phổ biến các hệ thống thông minh rộng đến mức nào, nhà nước viết các mục tiêu cụ thể vào văn kiện rồi gửi xuống dưới. Các chỉ tiêu chi tiết được xem xét ở nơi khác trong cuốn sách này trên cơ sở các văn kiện chính sách, nên chương này chỉ xác nhận phương hướng. Điều quan trọng là bàn tay chỉ đạo những mục tiêu phát triển áp đảo và bàn tay kiểm soát dữ liệu cùng nội dung không phải là hai bàn tay khác nhau.

Chỉ thị phát triển càng mạnh, các chuẩn mực quản lý sự phát triển đó càng dày lên để nó không đe dọa hệ thống. Hai trục này không kéo theo hai hướng ngược nhau. Chúng nâng đỡ nhau theo cùng một hướng. Điều này dễ hiểu hơn nếu ta nhớ lại các thông báo đăng ký ở phần trước. Một danh sách dài hơn là dấu hiệu của phát triển, vì thị trường đang lớn lên, và cũng là dấu hiệu của an ninh, vì nhiều dịch vụ hơn đã được nhà nước biết đến và quản lý. Khi dịch vụ tăng lên, các đối tượng phải đăng ký cũng tăng theo, và các ngưỡng phải quản lý cũng tăng lên.

Nhưng các chuẩn mực này không chỉ đi từ trên xuống. Ở đây, báo cáo nghiên cứu "China's AI Governance Is Not Only Top-Down" mở rộng bức tranh. Báo cáo lập luận rằng nếu ta chỉ nhìn

việc xây dựng quy tắc AI của Trung Quốc như một quá trình trong đó Đảng và nhà nước đơn phương đóng dấu ban hành quy tắc, ta sẽ bỏ lỡ cách nó thật sự vận hành. Các nhóm công tác của doanh nghiệp có mặt trong những căn phòng nơi tiêu chuẩn và hướng dẫn chi tiết được soạn thảo. Quá trình này gắn với một dây chuyền trong đó nhà nước đặt hướng đi, còn doanh nghiệp viết nội dung, với sự lãnh đạo của nhà nước và sáng kiến của doanh nghiệp trộn vào nhau. Các yêu cầu trong tiêu chuẩn tác nhân công nghiệp đã nói ở trên, như ranh giới đối với nhiệm vụ rủi ro cao cần con người can thiệp hoặc điều kiện đối với thiết bị dừng khẩn cấp, sẽ rất khó được tạo ra nếu không đi qua bàn tay của những doanh nghiệp hiểu lĩnh vực đó.

Một công ty đã chế tạo thiết bị biết nhiệm vụ nào có rủi ro cao và thiết bị dừng nên được đặt ở đâu. Bức tranh trong đó quản lý chỉ xuất hiện như sức ép một chiều bóp chặt doanh nghiệp được phủ thêm một mặt khác, trong đó doanh nghiệp tham gia làm ra quy tắc và dựng nên những ngưỡng quen thuộc với chính họ. Với một công ty đã góp phần viết quy tắc, các quy tắc ấy không trở thành một bức tường xa lạ, mà là những điều kiện được viết bằng thứ ngôn ngữ họ hiểu.

Khi đặt hai mặt này cạnh nhau, ý nghĩa của các danh sách thông báo đăng ký lại hiện rõ. Những công ty có tên trong danh sách là những công ty đã vượt qua các ngưỡng do nhà nước đặt ra, và cũng là những công ty, bằng cách này hay cách khác, đã đặt một chân vào quá trình tạo ra các ngưỡng đó. Một công ty đã góp phần viết một quy tắc sẽ thấy việc vượt qua quy tắc ấy dễ hơn đôi chút. Một công ty đứng xa quá trình xây dựng quy tắc sẽ cảm thấy cùng một ngưỡng ấy cao hơn. Cách quản trị nói về phát triển và an ninh cùng lúc không chia nhà nước và doanh nghiệp thành hai bên rạch ròi là bên quản lý và bên bị quản lý. Nó làm cho hai bên ăn khớp với nhau ngay trong cùng một văn bản.

Sự đan xen này là một lực đẩy phía sau tốc độ phát triển AI của Trung Quốc, đồng thời cũng là mục tiêu của các lập luận phản biện sẽ được xem xét ở phần sau. Giữa những công ty có khả năng tham gia viết luật chơi và những công ty không có khả năng đó, cùng một chuẩn mực có thể trở thành chiếc thang cho một số bên, nhưng lại là hàng rào đối với những bên khác.

Độc kiểm duyệt, trấn áp và gánh nặng doanh nghiệp như các lập luận phản biện

Phần thảo luận đến đây tập trung vào cách mô hình quản trị kiểu Trung Quốc buộc phát triển và an ninh đi cùng nhau. Nhìn từ phía ngược lại, cùng những chuẩn mực ấy lại được gọi bằng những tên khác. Kiểm soát được gọi là kiểm duyệt, quản lý được gọi là trấn áp, và các ngưỡng tuân thủ được gọi là gánh nặng doanh nghiệp. Một tài liệu thể hiện quan điểm đối lập này là "The Repressive Power of Artificial Intelligence" của tổ chức nhân quyền Freedom House. Điểm này cần được nói rõ ngay từ đầu. Vì tài liệu này phản ánh mạnh góc nhìn của một tổ chức vận động nhân quyền, các nhận định trong đó nên được đọc như một quan điểm và đối chiếu với các chứng cứ khác, thay vì chuyển thẳng thành kết luận về sự kiện.

Chương này xử lý vấn đề với sự thận trọng đó. Phân tích quản trị ở phần trước và phân tích trấn áp ở phần này là hai giọng nói gọi cùng một bộ quy tắc bằng những cái tên khác nhau. Thay vì quyết định trước rằng một bên đúng và bên kia sai, chương này theo dõi cách hai giọng nói ấy đặt tên khác nhau cho cùng những quy định.

Điểm xuất phát của cách nhìn này là một đánh giá về tự do internet. Trong các khảo sát của Freedom House, Trung Quốc trong nhiều năm đã bị nêu tên là một quốc gia có môi trường tự do internet nằm gần đáy bảng. Tổ chức này nhìn AI tạo sinh như một con dao hai lưỡi đối với các chính phủ độc đoán. Một mặt, chatbot có thể mở ra một lối vòng tới thông tin không bị kiểm duyệt cho công dân trong một môi trường khép kín. Khi một cỗ máy có thể tự tạo câu trả lời đi vào một không gian trước đây chỉ xuất hiện những câu trả lời được định sẵn, các khe hở trong kiểm soát có thể mở ra. Mặt khác, chẩn đoán của tài liệu là nhà nước nhanh chóng sửa luật và công nghệ để kiểm soát công cụ đó, đan AI vào bộ máy kiểm duyệt và đàn áp.

Nói cách khác, quy định được đưa ra để chặn tuyến đường ấy trước khi lối vòng kịp hình thành. Nhìn từ góc này, nghĩa vụ xác minh tên thật và yêu cầu dữ liệu huấn luyện phải hợp pháp, đã bàn ở trên, không được đọc như thủ tục phục vụ đổi mới, mà như công cụ đàn áp xóa bỏ tính ẩn danh và bảo vệ sự thuần khiết về ý thức hệ. Cùng một điều khoản, ở một phía mang tên an toàn, ở phía kia mang tên đàn áp. Câu chữ của quy tắc vẫn giống nhau, nhưng ý nghĩa của nó thay đổi tùy người đọc đứng ở đâu.

Nếu cũng những quy tắc ấy được nhìn từ sổ sách của doanh nghiệp, chúng mang tên gánh nặng. Hãy nghĩ lại về nghĩa vụ huấn luyện lại đã bàn ở trên. Nếu cả một mô hình phải được huấn luyện lại chỉ vì chatbot từng tạo ra một câu có vấn đề, chi phí đó trở thành sức nặng mà một công ty nhỏ gần như không thể gánh nổi. Trách nhiệm phải chịu với tư cách nhà sản xuất, nghĩa vụ gắn nhãn và quản lý nhãn, cùng việc cài một hệ thống xác minh tên thật vào bên trong dịch vụ, tất cả đều đổ lên doanh nghiệp. Từ đây nảy sinh phê phán rằng một phần vốn và tài nguyên tính toán lẽ ra phải dành cho đổi mới lại phải dùng để chứng minh sự tuân thủ với hệ thống.

Theo lập luận này, doanh nghiệp phải trả chi phí đáp ứng các yêu cầu kiểm duyệt của nhà nước trước khi có thể bắt đầu cạnh tranh với nhóm dẫn đầu thế giới. Phần trước nói rằng các doanh nghiệp cùng viết nên những quy tắc này, nhưng việc cùng đứng tên soạn thảo không làm gánh nặng nhẹ hơn. Giữa các công ty lớn có năng lực tham gia xây dựng quy tắc và các công ty nhỏ không có năng lực ấy, cùng một ngưỡng lại được cảm nhận như một độ cao khác nhau. Một quy tắc được viết cùng các công ty lớn vẫn là bức tường mà công ty nhỏ phải tự mình trèo qua.

Nhìn từ vị trí của lao động, một bóng đen khác hiện ra. Dữ liệu dùng để dạy AI cần nhãn do con người trực tiếp gắn vào, và công việc gắn nhãn được cho là được quản lý theo cách điều chỉnh năng lực nhận thức của con người theo nhịp của chương trình máy tính. Có người chỉ ra

rằng những phương pháp cũ từng kỷ luật cơ thể nay đã chuyển thành kiểu kiểm soát linh hoạt, dẫn dắt mô thức suy nghĩ của con người. Những người quan sát cũng ghi nhận rằng người lao động gắn nhãn chống lại sự kiểm soát này bằng cách tự giảm sản lượng hoặc thường xuyên đổi việc. Nếu nhớ lại rằng quy định được bàn ở phần trước yêu cầu doanh nghiệp tự lập quy tắc cho quy trình gắn nhãn, ta có thể thấy lao động con người đứng ở điểm cuối của những quy tắc ấy.

Các chuẩn mực kiểm soát dữ liệu vươn tới cả lao động của những người tạo ra dữ liệu. Một yêu cầu được chuyển xuống doanh nghiệp để quản lý dữ liệu biến thành những quy tắc quản lý một ngày làm việc của người gắn thẻ cho dữ liệu đó. Vấn đề lao động này sẽ được bàn kỹ trong các chương sau như một lập luận phản đối, nên ở đây chỉ cần xác nhận rằng bàn tay của chuẩn mực đã vượt ra ngoài mô hình và nội dung để đi vào nơi làm việc.

Khi mở rộng tầm nhìn ra toàn bộ thị trường lao động, ta thấy một nỗi sợ của nhà nước. Đó là nỗi sợ rằng khi AI lan rộng và việc làm biến mất, điều này có thể dẫn tới bất ổn xã hội, rồi dẫn tới bất ổn chính trị. Nếu nhớ rằng khái niệm an ninh quốc gia toàn diện được bàn ở phần trước xem an ninh kinh tế và ổn định xã hội là nền tảng duy trì chế độ, ta sẽ dễ hiểu hơn vì sao việc làm được xử lý bằng ngôn ngữ an ninh. Để đáp lại mối lo này, nhà chức trách Trung Quốc được cho là đã chuẩn bị hướng dẫn tiêu chuẩn quốc gia, yêu cầu đánh giá rủi ro việc làm trước khi áp dụng AI tạo sinh. Để một doanh nghiệp phát hành một dịch vụ đổi mới, một ngưỡng khác đã xuất hiện: trước hết doanh nghiệp phải ước tính và báo cáo công nghệ sẽ thay thế việc làm của con người đến mức nào.

Hướng dẫn này lúc đầu cũng được xem như một tuyên bố không có sức cưỡng hành thực tế, nhưng về sau được biết là đã được tinh chỉnh thành dạng một tiêu chuẩn thật sự. Chuyển động trong đó một quy tắc khởi đầu như tuyên bố, rồi qua nhiều năm trở thành một yêu cầu cụ thể, giống con đường mà các nghĩa vụ gắn nhãn và đánh dấu đã đi qua như đã bàn ở trên. Nếu chỉ nhìn vào ngôn ngữ lỏng lẻo ban đầu rồi gạt đi vì cho rằng nó không hiệu quả, vài năm sau người ta có thể phát hiện nó đã cứng lại thành một ngưỡng thật sự, níu chân doanh nghiệp. Một quốc gia thúc đẩy phát triển tiến lên cũng lo cho những người bị phát triển đẩy ra ngoài, và bắt doanh nghiệp gánh một phần nỗi lo ấy.

Nên đọc những lập luận đối lập này như thế nào? Ở một đầu là một hệ thống quản trị tinh vi, gắn phát triển với an ninh. Ở đầu kia là cách nhìn coi hệ thống quản trị ấy như một chuỗi kiểm duyệt, trấn áp và gánh nặng cho doanh nghiệp. Hai cách nhìn đặt những cái tên khác nhau cho cùng một nhóm văn bản. Một bên gọi danh sách trong thông báo đăng ký là bằng chứng của một thị trường có trật tự, bên kia gọi đó là bằng chứng của một không gian nơi chỉ lời nói được cho phép mới tồn tại. Một bên gọi xác minh danh tính thật là điều kiện để sử dụng có trách nhiệm, bên kia gọi đó là thiết bị giám sát xóa bỏ tính ẩn danh.

Cuốn sách này không chọn một trong hai cái tên ấy làm đáp án đúng. Nó đặt hai cái tên cạnh nhau và ghi nhận rằng chúng cùng chỉ vào một hệ thống quy tắc. Vì đánh giá của Freedom

House mang một góc nhìn mạnh, đọc chẩn đoán ấy cùng với phân tích quản trị sẽ cho thấy đồng thời hai mặt của quy định. Nếu chỉ nghe một tiếng nói rồi nghĩ rằng mình đã hiểu AI của Trung Quốc, ta sẽ bỏ lỡ cách chuẩn mực ấy tạo ra cả tốc độ lẫn rủi ro.

Có một điểm hiện rõ trong các văn bản. Các chuẩn mực kiểm soát dữ liệu của Trung Quốc bắt đầu từ phần hậu trường của doanh nghiệp, rồi đi qua dữ liệu huấn luyện, đăng ký thuật toán, xác minh danh tính thật, gắn nhãn nội dung được tạo ra và nghĩa vụ huấn luyện lại, vận hành trên toàn bộ vòng đời của dịch vụ. Tại bàn nơi các chuẩn mực này được soạn ra không chỉ có nhà nước, mà còn có các công ty từ thực địa. Nếu quy định chỉ là sự trấn áp một chiều, sẽ khó giải thích cảnh các công ty cùng tham gia viết văn bản. Nếu quy định hoàn toàn là hợp tác, cũng sẽ khó giải thích sức nặng mà các công ty cảm thấy trước nghĩa vụ huấn luyện lại.

Chuẩn mực này, được tạo ra giữa phát triển và an ninh, không xóa bỏ căng thẳng giữa hai mục tiêu ấy. Nó giữ căng thẳng đó ngay trong các văn bản. Giống như thiết bị dừng khẩn cấp được gắn vào một tác nhân công nghiệp đã đặt hiệu quả và kiểm soát trong cùng một câu, toàn bộ chuẩn mực cũng chuyển động trong khi mang theo sức căng giữa việc đẩy phát triển tiến lên và giữ an ninh đứng vững. Sức căng ấy không phải là lỗi. Đó là cách chính hệ thống quản trị này vận hành.

Vậy còn lại một câu hỏi. Khi các công ty vừa tham gia vào việc xây dựng quy tắc, vừa phải gánh sức nặng của chính các quy tắc ấy, cấu trúc này theo thời gian sẽ trở thành một ngưỡng mà họ học được cách vượt qua dễ dàng hơn, hay sẽ vẫn là một bức tường dày hơn khi hệ thống ngày càng vạch thêm những ranh giới mà nó phải bảo vệ? Việc các dấu nhãn và đánh giá rủi ro việc làm đều cứng lại, từ cách diễn đạt ban đầu còn lỏng thành các tiêu chuẩn thật sự, cho thấy một khi chuẩn mực đã được đặt ra, nó thường trở nên kém linh hoạt hơn, chứ không mỏng đi.

Các công ty có tên trong danh sách đăng ký vào mùa xuân năm 2026 hiện đang đi qua chính giữa câu hỏi đó. Danh sách này được cập nhật hai tháng một lần. Mỗi lần như vậy, những cái tên mới xuất hiện và một số cái tên biến mất. Mở danh sách ấy ra cũng đồng nghĩa với việc mỗi lần lại phải kiểm tra trạng thái hiện tại của các chuẩn mực kiểu Trung Quốc, nơi phát triển và an ninh bám chặt lấy nhau mà chưa nghiêng hẳn về bên nào.

Chương 4. Cục Dữ liệu Quốc gia, chính sách bộ dữ liệu và mối liên kết giữa mô hình và dữ liệu

Khi cuộn xuống các thông báo trên trang web của Cục Dữ liệu Quốc gia, một thông báo đăng vào đầu năm 2026 lập tức gây chú ý. Tiêu đề của nó là "Thông báo về Hành động Cộng hưởng Mô hình-Dữ liệu năm 2026." Nghe giống một khẩu hiệu nghiên cứu của một công ty công nghệ, nhưng bên gửi không phải là công ty. Đó là một cơ quan chính phủ. Trong thuật ngữ, tức cộng hưởng mô hình-dữ liệu, chữ đầu chỉ mô hình, chữ thứ hai chỉ dữ liệu. Lý do đặt hai thứ này cạnh nhau rất rõ. Mục tiêu là làm cho mô hình và dữ liệu kéo nhau lại gần, thay vì vận hành trong hai thế giới riêng. Mở thông báo ra, chính phủ bắt đầu bằng câu chuyện về mô hình, rồi lập tức nói đến dữ liệu, sau đó lại đi từ dữ liệu trở về các hiện trường công nghiệp.

Cách mô hình, dữ liệu và nhà máy lần lượt xuất hiện trong cùng một đoạn văn cho thấy thái độ của chính phủ: họ muốn buộc cả ba vào một mạch vận hành chung. Chương này bắt đầu từ chính thông báo ấy. Nó lần theo thứ tự văn bản để xem một văn bản do một cơ quan ban hành được chuyển thành kế hoạch thực hiện ở cấp tỉnh và cấp tập đoàn như thế nào, rồi thành các biểu mẫu báo cáo, và cuối cùng chạm tới dữ liệu cảm biến bên trong nhà máy cùng hồ sơ nhân sự của người lao động ra sao.

Trước hết cần dừng lại ở tên gọi Cục Dữ liệu Quốc gia. Đây là một cơ quan còn xa lạ với độc giả Việt Nam. Nếu chỉ đọc chính sách trí tuệ nhân tạo của Trung Quốc qua bảng xếp hạng doanh nghiệp, rất khó nhìn thấy cơ quan này. Những câu chuyện về DeepSeek cắt giảm chi phí, hay Qwen của Alibaba hỗ trợ nhiều ngôn ngữ, thường đi qua tên các công ty. Nhưng trên sân chơi nơi những công ty ấy cạnh tranh, chính phủ đã kẻ sẵn các đường biên. Một cơ quan thẩm định xem một mô hình có được cung cấp như dịch vụ công cộng hay không, một cơ quan khác đặt ra chuẩn an toàn. Cục Dữ liệu Quốc gia là nơi đường biên dữ liệu được vạch ra giữa rất nhiều đường biên ấy.

Việc này khác về tính chất so với thẩm định mô hình hay lọc nội dung. Công việc của cơ quan này là gom dữ liệu phân tán, đặt ra chuẩn chất lượng, và quyết định bằng văn bản dữ liệu nào sẽ chảy vào mô hình nào trong ngành nào. Bản thân việc thông báo về cộng hưởng mô hình-dữ liệu xuất hiện trong mục thông báo của Cục Dữ liệu Quốc gia đã báo hiệu rằng dữ liệu đã đi vào kết cấu chính của chính sách trí tuệ nhân tạo. Một chính sách nói về mô hình trước và một chính sách nói về dữ liệu trước có trọng tâm khác nhau. Khi Cục Dữ liệu Quốc gia bước ra tuyến đầu, sức nặng chính sách nghiêng về dữ liệu.

Chúng ta hãy phân tích kỹ hơn một chút sự chuyển dịch trọng tâm ấy. Trong trí tuệ nhân tạo, sự chú ý hầu như luôn nghiêng về mô hình. Công ty nào phát hành mô hình lớn hơn, mô hình nào đạt bao nhiêu điểm trong một bài kiểm tra, những chuyện đó trở thành tin tức. Dữ liệu ẩn phía sau nên khó nhìn thấy. Nhưng ngay khi chính phủ đưa cụm từ cộng hưởng mô hình-dữ

liệu vào tiêu đề của một văn kiện chính thức, dữ liệu đã được đặt ở phần đầu văn kiện, bằng cỡ chữ ngang với mô hình. Cách đặt ấy không phải ngẫu nhiên.

Mô hình lớn lên nhờ được nuôi bằng dữ liệu, và những bản ghi mới mà các mô hình ấy tạo ra trong thực địa lại trở thành dữ liệu, rồi quay về mô hình tiếp theo. Chính phủ đã coi vòng phản hồi này như một mạch kín và đặt tên cho mạch ấy. Đó là điểm mà chương này sẽ quay lại nhiều lần. Một khi mạch đã có tên, cơ quan, biểu mẫu và lịch trình quản lý nó sẽ đi theo. National Data Administration đứng ở tuyến đầu của hệ thống quản lý gắn với cái tên ấy.

Các câu trong thông báo không dừng ở tuyên bố. Văn kiện chính sách của Trung Quốc có một cầu thang đi từ trên xuống dưới. Khi trung ương ban hành thông báo, các tỉnh và tập đoàn phải biến nó thành kế hoạch triển khai phù hợp với khu vực và ngành của mình. Chính phủ đã có sẵn một văn kiện đặt khung cho các kế hoạch triển khai ấy. Đó là "XX Province/Group Parameter-Data Resonance Action Implementation Plan Template." Từ template rất đáng chú ý. Chính quyền địa phương hoặc các tập đoàn doanh nghiệp không tự do lập kế hoạch theo ý mình. Họ điền số liệu của khu vực mình vào một biểu mẫu có ô trống do trung ương gửi xuống. Theo các mục đã cố định, họ viết ngành nào sẽ được chọn, kịch bản nào sẽ làm trước, bao nhiêu bộ dữ liệu sẽ được tạo ra, và doanh nghiệp nào sẽ phụ trách.

Mẫu kế hoạch triển khai yêu cầu viết riêng mục tiêu, nhiệm vụ, lịch trình và biện pháp hỗ trợ. Ô mục tiêu chứa cam kết sẽ đạt được điều gì vào thời điểm nào. Ô nhiệm vụ chứa các công việc cụ thể được tách ra từ cam kết ấy. Ô lịch trình chứa thời gian biểu chia theo từng quý. Ô biện pháp hỗ trợ nêu rõ tổ chức, ngân sách và nhân sự sẽ được cung cấp ra sao. Khi điều này xảy ra, các kế hoạch do những tỉnh khác nhau lập ra sẽ cùng dùng chung một khung. Dù ngành nghề và điều kiện mỗi vùng khác nhau, biểu mẫu chung giúp trung ương dễ dàng gom các kế hoạch vào một bảng và so sánh cạnh nhau. Một kế hoạch địa phương nhìn có vẻ tự chủ, nhưng thực ra vận hành bên trong biểu mẫu do trung ương đặt ra.

Cần nói lại một lần nữa định dạng này có nghĩa gì. Nếu mỗi địa phương viết kế hoạch bằng những cách diễn đạt khác nhau, trung ương sẽ khó so sánh các kế hoạch ấy với nhau. Trung ương không thể nhìn thoáng qua là biết tỉnh nào đi trước, tỉnh nào tụt lại. Nhưng nếu trung ương gửi xuống một biểu mẫu có cùng các ô trống, những văn bản được thu lại từ bên dưới sẽ có cùng các mục ở cùng vị trí. Có bao nhiêu bộ dữ liệu đã được tạo ra, bao nhiêu kịch bản đã được mở, và doanh nghiệp nào đóng vai trò chủ thể chính, tất cả đều xuất hiện trong cùng những ô tương ứng.

Trung ương có thể đo tiến độ của từng khu vực chỉ bằng cách rà qua những ô đó. Vì vậy, biểu mẫu vừa là sợi xích trói tay chân chính quyền địa phương, vừa là thiết bị trở thành đôi mắt của trung ương. Hai đặc điểm này chồng lên nhau trong cùng một mẫu văn bản. Quyền tự do lập kế hoạch thu hẹp lại, nhưng đổi lại, thành tích của địa phương trở thành một tập hợp con số mà trung ương có thể đọc bất cứ lúc nào.

Khi kế hoạch triển khai hoàn chỉnh được đưa vào thực hiện, kết quả của nó sau đó phải được gửi ngược lên trên. Biểu mẫu dùng cho báo cáo ấy là "Parameter-Data Resonance Action Situation Summary Report Template." Văn bản này đáng được xem kỹ trong chương này, vì mẫu báo cáo tổng kết cho thấy, qua các con số và hạng mục, chính phủ Trung Quốc đòi hỏi gì về dữ liệu. Báo cáo đi theo năm đề mục chính. Trước hết, nó yêu cầu trình bày tình hình triển khai tổng thể.

Báo cáo yêu cầu tóm tắt toàn bộ quá trình thực hiện Parameter-Data Resonance Action của khu vực hoặc doanh nghiệp, nêu các lợi ích kinh tế và xã hội mà hành động này tạo ra, và giải thích trí tuệ nhân tạo đã thúc đẩy công nghiệp hóa mới như thế nào. Tiếp đó, báo cáo yêu cầu nhập các thành quả vào một số bảng cố định. Các kịch bản giá trị cao, mô hình công nghiệp, mô hình chuyên dụng và các tác nhân đặc sắc, bộ dữ liệu chất lượng cao, không gian Parameter-Data Resonance, và bộ dữ liệu đánh giá, mỗi loại được đưa vào một bảng riêng.

Bảy bảng phụ lục chia không gian báo cáo theo loại thành quả. Phần tiếp theo của mục lục yêu cầu cơ quan báo cáo giải thích họ đã xử lý dữ liệu ra sao, đã mở dữ liệu cho bên ngoài như thế nào, và đã áp dụng những biện pháp bảo vệ an toàn nào. Nếu họ xử lý dữ liệu quan trọng, họ phải công bố riêng quy trình đó. Phần thứ tư yêu cầu họ mô tả cách xây dựng tổ chức và thúc đẩy các nhiệm vụ trọng điểm, còn phần cuối yêu cầu họ ghi lại những ràng buộc gặp phải trong quá trình thực hiện và các khuyến nghị cho tương lai. Đây không phải là văn bản chỉ để ghi lời khen. Nó buộc cơ quan báo cáo phải công bố dữ liệu đã đi đâu, họ đã bảo vệ những gì, và họ đã vấp phải rào cản ở đâu.

Khi mở lần lượt các bảng được đính kèm làm phụ lục cho mẫu báo cáo này, quy mô và độ sâu của lượng dữ liệu mà chính phủ yêu cầu trở nên rất cụ thể. Bảng đầu tiên là danh sách các bộ dữ liệu đa dụng chất lượng cao cho các ngành công nghiệp then chốt. 'Đa dụng' ở đây có nghĩa là tri thức nền tảng có thể áp dụng rộng rãi trong toàn ngành. Bảng đưa ra các trường hợp mẫu làm ví dụ, và trong đó có một bộ dữ liệu về quy trình luyện kim thép. Ngành được ghi cho bộ dữ liệu này là thép, còn một tập đoàn thép và một viện nghiên cứu luyện kim được nêu cạnh nhau với tư cách là các đơn vị sản xuất dữ liệu và vận hành dịch vụ. Việc ghép hai chủ thể này truyền đi thông điệp rằng ngành công nghiệp, giới học thuật và các tổ chức nghiên cứu phải cùng nhau xây dựng dữ liệu. Cột quy mô tài nguyên chứa những con số gây chú ý: 500.000 tham số quy trình, 50.000 hình ảnh vi cấu trúc, và 10.000 trang tài liệu tiêu chuẩn kỹ thuật.

Đây không phải là văn bản được thu thập từ web. Đó là dữ liệu vật lý từ mặt bằng nhà máy. Hình thức dữ liệu pha trộn giữa số liệu chuỗi thời gian, hình ảnh và văn bản, còn nội dung bao phủ toàn bộ quy trình từ luyện gang bằng lò cao đến luyện thép bằng lò chuyển, đúc liên tục, cán nóng và cán nguội. Nó phải bao gồm phạm vi được chấp nhận của các tham số quy trình thông dụng, danh mục có minh họa về các lỗi thường gặp, các đường cong tiêu chuẩn cho vận hành thiết bị, quy chuẩn kỹ thuật công nghiệp, và quy tắc an toàn. Điều mà bảng này yêu cầu

không phải là chỉ dẫn mua và dùng phần mềm. Nó gần với một yêu cầu buộc các nhà máy biến những hồ sơ và bản vẽ tương tự mà họ đã tích lũy trong nhiều năm thành tài sản dữ liệu theo khuôn dạng do chính phủ đặt ra, rồi đưa chúng ra sử dụng. Một con số do cảm biến để lại, một trang trong tài liệu tiêu chuẩn, trở thành vật liệu lắp vào một ô trong danh sách do nhà nước xác định.

Đến điểm này, tính chất của dữ liệu thay đổi. Bên trong nhà máy, 500.000 tham số vốn chỉ là những ghi chép còn lại trong khi thiết bị vận hành. Nhưng ngay khi các ghi chép ấy được sắp xếp và đưa vào một danh sách do nhà nước xác định, chính những con số đó biến thành tài sản của chính sách công nghiệp. Việc tập đoàn thép và viện nghiên cứu luyện kim được liệt kê cùng nhau với tư cách là các đơn vị chịu trách nhiệm phù hợp với sự thay đổi này. Cách sắp xếp ấy không để việc tạo dữ liệu cho riêng một công ty. Nó nối các công ty tại xưởng sản xuất với các viện nghiên cứu, và cố gắng nâng những hồ sơ của họ thành tri thức nền tảng mà toàn ngành có thể dùng.

Đó là lý do cụm từ 'mục đích chung' được gắn vào đây. Mục tiêu là xây dựng một nền tảng có thể hoạt động ở bất cứ đâu trong ngành thép, chứ không phải một bộ dữ liệu chỉ dùng được cho một nhà máy. Một bộ dữ liệu được tạo ra theo cách này trở thành nguyên liệu đầu vào cho mô hình tiếp theo, rồi những bản ghi mới mà mô hình ấy tạo ra tại một nhà máy khác lại quay trở về bộ dữ liệu. Vòng mạch mà cụm từ moushu gongzhen, cộng hưởng tham số-mô hình, vẽ ra bắt đầu mang những tên ngành và con số cụ thể trong chính bảng này, rồi bắt đầu vận hành.

Cùng một biểu mẫu báo cáo ấy còn có danh sách các bộ dữ liệu đánh giá. Bảng này nêu rõ một tác tử phải vượt qua những chuẩn kiểm tra nào trước khi được triển khai tại nơi làm việc công nghiệp. Trong ví dụ về dự báo hỏng hóc thiết bị, một bộ dữ liệu gồm 500 cặp hỏi đáp dạng văn bản phải đáp ứng bốn tiêu chí. Độ chính xác nhận diện lỗi phải đạt ít nhất 95 phần trăm, hệ thống phải chịu được tiếng ồn trong xưởng, thời gian suy luận không được vượt quá 50 mili giây, và hệ thống còn được đánh giá theo mức độ phù hợp của các tham số quy trình. Độ chính xác, sức chịu đựng, tốc độ và độ khó: cả bốn đều phải có trong một bảng.

Biểu mẫu không dừng lại ở đó. Trong trường hợp ứng dụng, doanh nghiệp phải nhập các con số cải thiện thực tế. Chẳng hạn, một công ty có thể báo cáo rằng họ đã phát hiện trước tình trạng mài mòn ở ổ trục chính và nâng độ chính xác cảnh báo sớm từ 85 phần trăm lên 94 phần trăm. Đây là một phương pháp buộc doanh nghiệp phải cố định điểm xuất phát và điểm kết thúc bằng con số, thay vì mô tả kết quả bằng lời. Theo định dạng này, một tuyên bố rằng mô hình hoạt động tốt sẽ không được chấp nhận. Kết quả chỉ được tính khi doanh nghiệp điền số liệu vào bảng: từ bao nhiêu phần trăm lên bao nhiêu phần trăm, hoặc mô hình đưa ra câu trả lời trong bao nhiêu mili giây. Bảng này là nơi lời khoe định tính biến thành hồ sơ định lượng.

Nhìn kỹ hơn các chuẩn định lượng này sẽ thấy chúng là ngôn ngữ của hiện trường công nghiệp. Con số 50 mili giây không phải là chuẩn để viết bài trong phòng thí nghiệm. Khi thiết bị

nhà máy gửi tín hiệu bất thường, hiện trường coi mô hình là chậm nếu nó mất nhiều thời gian hơn mức đó để đọc tín hiệu và phản hồi. Độ chính xác 95 phần trăm cũng vậy. Nếu hệ thống bỏ sót một lỗi trong hai mươi lỗi, một lần bỏ sót ấy có thể dẫn đến sản phẩm lỗi, nên ngưỡng này được đặt ra. Khả năng chịu nhiễu được yêu cầu riêng vì nhà máy không phải là một phòng thí nghiệm yên tĩnh.

Mô hình phải đưa ra câu trả lời ổn định ngay cả khi rung động, nhiễu điện từ và thay đổi nhiệt độ cùng xuất hiện. Một bảng đặt bốn tiêu chuẩn này cạnh nhau cho thấy lập trường của chính phủ: họ sẽ chấm điểm mô hình bằng ngôn ngữ của công nghiệp. Họ không đòi hỏi một mô hình tốt theo nghĩa trừu tượng, mà đòi hỏi một mô hình đứng vững được tại hiện trường. Vì yêu cầu đó được cố định bằng con số, các công ty phải tinh chỉnh cả mô hình lẫn dữ liệu của mình để đạt những con số ấy.

Một phụ lục khác của mẫu báo cáo nhắm vào không gian vật lý nơi toàn bộ dữ liệu này thực sự được thu thập và tính toán. Đó là danh sách các không gian cộng hưởng tham số. Ví dụ được nêu là một không gian sản xuất thiết bị tại một tỉnh nhất định. Trường phần cứng yêu cầu một cụm tính toán cấp 180P, 4 petabyte lưu trữ và một nền tảng truy cập Internet of Things công nghiệp. Trường phần mềm yêu cầu một nền tảng song sinh số cho thiết bị, phần mềm trung gian kết nối dữ liệu từ nhiều nhà sản xuất, một hệ thống huấn luyện mô hình liên kết và một nền tảng quản lý quyền sở hữu trí tuệ. Các mục quản lý dữ liệu phương thức hợp tác dữ liệu và phương thức xác nhận quyền dữ liệu đã được thiết lập hay chưa. Vì dữ liệu đi qua nhiều bên, biểu mẫu kiểm tra xem quy trình hành chính đã đăng ký dữ liệu đó thuộc về ai và đã xác nhận các quyền ấy hay chưa.

Các mục ứng dụng ghi nhận những kết quả như kết nối hơn 2.000 nút thiết bị, bao gồm máy công cụ điều khiển số và robot công nghiệp. Trường trong bảng này nhìn xa về phía trước là kết nối liên thông. Câu ví dụ nêu rằng việc ghép nối kỹ thuật với hạ tầng dữ liệu quốc gia đã hoàn tất vào tháng 3 năm 2026, một cuộc thử nghiệm kết nối các danh mục dữ liệu đang được tiến hành, và khả năng vận hành liên thông đầy đủ của toàn bộ dữ liệu sẽ đạt được vào khoảng tháng 5 năm 2026. Chỉ một trường này cho thấy các không gian dữ liệu do từng công ty hoặc từng tỉnh xây dựng cuối cùng được thiết kế để kết nối dưới mạng dữ liệu quốc gia. Một cụm tính toán trong một nhà máy không được dựng lên để chỉ chạy bên trong chính nó. Lịch trình của nó được đặt sao cho, vào một ngày đã định, nó kết nối với hạ tầng quốc gia ở tầng trên.

Trường kết nối liên thông đáng được chú ý kỹ trong chương này. Một yêu cầu nói phải cung cấp phần mềm trung gian kết nối dữ liệu từ nhiều nhà sản xuất. Một yêu cầu khác nói phải nhập trước ngày ghép nối với hạ tầng dữ liệu quốc gia. Cả hai cùng xuất hiện trong một bảng. Yêu cầu thứ nhất liên quan đến việc gom dữ liệu từ các thiết bị khác nhau trong nhà máy vào một hệ thống. Yêu cầu thứ hai liên quan đến việc đưa dữ liệu đã được gom theo cách đó ngược lên mạng quốc gia. Đọc cùng nhau, hai yêu cầu này cho thấy hướng mà dữ liệu được dự định sẽ

chảy theo.

Đó là một cầu thang đi lên từ thiết bị đến nhà máy, từ nhà máy đến không gian của tỉnh, rồi từ không gian cấp tỉnh đến hạ tầng quốc gia. Nhu cầu về một quy trình xác nhận quyền đối với dữ liệu cũng khớp với dòng chảy này. Khi dữ liệu đi qua nhiều bên và được đẩy lên cao hơn, tranh chấp về quyền sẽ phát sinh về sau nếu từng công đoạn không ghi rõ đó là dữ liệu của ai. Để ngăn những tranh chấp ấy từ trước, chính phủ cố định việc xác nhận quyền thành một thủ tục hành chính.

Ở điểm này, ba loại văn bản được xâu lại thành một đường. Thông báo đặt ra hướng đi, mẫu kế hoạch thực hiện biến hướng đi đó thành kế hoạch cho các địa phương và doanh nghiệp, còn mẫu báo cáo tổng kết thu thập kết quả của các kế hoạch ấy vào một bảng cố định. Cấu trúc qua lại này, đi từ trên xuống dưới rồi từ dưới quay trở lại trên, trao hình thức hành chính cho cụm từ 'cộng hưởng mô hình-dữ liệu'. Cuộc gặp giữa mô hình và dữ liệu không vận hành bằng khẩu hiệu. Nó chạy qua những thủ tục lặp đi lặp lại: điền vào chỗ trống và nộp bảng biểu. Một thông báo của một bộ trở thành các kế hoạch do địa phương và doanh nghiệp viết ra, rồi các kế hoạch ấy quay lại thành những báo cáo đầy số liệu. Chu trình đó mới là con đường thực sự để chính sách chạm xuống mặt đất.

Một văn bản khác hỗ trợ dòng chảy này từ bên cạnh. Đó là kế hoạch thực hiện nhằm thúc đẩy xây dựng các bộ dữ liệu công nghiệp chất lượng cao. Nếu thông báo về cộng hưởng mô hình-dữ liệu vẽ ra bức tranh rộng về việc mô hình gặp dữ liệu, thì văn bản này tách riêng phía bộ dữ liệu trong bức tranh đó để xử lý. Việc chuẩn bị dữ liệu mà nó yêu cầu vừa thô ráp vừa thực tế. Nó kêu gọi nghiên cứu các công nghệ tự động làm sạch dữ liệu bị trộn lẫn qua nhiều định dạng, các quy trình lọc lỗi, bản trùng và khiếm khuyết để đưa dữ liệu về trạng thái sẵn sàng cho phân tích, cùng các công nghệ gắn nhãn thông minh để gắn ý nghĩa, tất cả nhằm xây dựng các bộ dữ liệu chuẩn hóa. Nó cũng yêu cầu dữ liệu tổng hợp và phương pháp kiểm thử đối kháng để tạo các bộ dữ liệu thử nghiệm mô phỏng điều kiện cực đoan, như lưu lượng giao thông tăng vọt hoặc liên lạc bị đứt.

Mục tiêu là tạo trước bằng dữ liệu những tình huống xấu có khả năng xảy ra ngoài hiện trường, rồi kiểm tra xem mô hình có đứng vững được trong các điều kiện ấy hay không. Khi mã nguồn mở được dùng làm dữ liệu, kế hoạch yêu cầu các pipeline lọc bỏ xung đột giấy phép và lỗi hỏng bảo mật. Đây là chỉ thị nhằm sàng lọc không chỉ chất lượng dữ liệu, mà cả rủi ro pháp lý và an ninh ẩn bên trong nó. Việc thu thập dữ liệu và bảo vệ dữ liệu xuất hiện trong cùng một câu là dấu hiệu của các văn kiện chính sách Trung Quốc, nơi phát triển và an ninh luôn được đặt cạnh nhau.

Kế hoạch xây dựng bộ dữ liệu này cũng đi vào cơ cấu tổ chức bên trong doanh nghiệp. Nó yêu cầu doanh nghiệp lập vị trí giám đốc dữ liệu chuyên trách về dữ liệu và tuân theo các tiêu chuẩn quốc gia dùng để đo mức độ trưởng thành của năng lực quản lý dữ liệu. Khi việc này diễn

ra, các hồ sơ vốn rải rác giữa các phòng ban được gom lại dưới sự chỉ đạo của một người, và những hồ sơ đó chỉ trở thành thức ăn cho mô hình sau khi đã được tinh chỉnh để đáp ứng các tiêu chuẩn chất lượng do nhà nước đặt ra. Toàn bộ quá trình một doanh nghiệp tự thu thập, làm sạch, mở rộng và gán nhãn dữ liệu được sắp xếp để vận hành theo các quy cách mà chính phủ yêu cầu.

Ở đây, ta thấy vì sao các mục tiêu bằng con số về việc tạo ra một vài bộ dữ liệu lại gây sức ép lên hiện trường. Mục tiêu được treo từ trên xuống, còn các tổ chức, tiêu chuẩn và biểu mẫu cần để đạt mục tiêu ấy đi xuống và trở thành một phần trong công việc hằng ngày của doanh nghiệp. Việc bổ nhiệm một lãnh đạo phụ trách dữ liệu và đáp ứng tiêu chuẩn trưởng thành ban đầu có thể là lựa chọn của doanh nghiệp, nhưng rồi chúng chuyển thành yêu cầu chính sách. Một văn bản này cho thấy quá trình quản trị dữ liệu doanh nghiệp đi từ phạm vi tự chủ sang phạm vi chính sách.

Nhìn kỹ hơn vào vai trò của giám đốc dữ liệu, ta thấy yêu cầu này có sức làm thay đổi sơ đồ tổ chức của doanh nghiệp. Công việc liên quan đến dữ liệu ban đầu nằm rải rác ở nhiều phòng ban. Bộ phận sản xuất nắm dữ liệu thiết bị, bộ phận chất lượng nắm hồ sơ kiểm tra, còn bộ phận nhân sự nắm thông tin nhân viên. Sự phân tán này khiến doanh nghiệp khó kết nối dữ liệu thành một tài sản thống nhất. Vì vậy, chính phủ đang yêu cầu doanh nghiệp đặt một lãnh đạo phụ trách dữ liệu đứng trên các chức năng rải rác ấy.

Khi dữ liệu của các phòng ban được tổ chức dưới quyền lãnh đạo này theo cùng một tiêu chuẩn, dữ liệu đó cuối cùng mới trở thành nguyên liệu thô cho những bộ dữ liệu mà nhà nước đòi hỏi. Một cuộc tái tổ chức doanh nghiệp bắt đầu từ một câu trong văn bản của chính phủ. Cùng logic ấy giải thích vì sao nhà nước gắn thêm một tiêu chuẩn quốc gia để đo mức độ trưởng thành. Nếu chính phủ đo việc quản lý dữ liệu của doanh nghiệp bằng thang đo của mình, doanh nghiệp phải bổ sung nhân sự và quy trình để nâng điểm trên thang đo đó.

Đến đây, cần lùi lại một bước và đọc cấu trúc này từ phía ngược lại. Việc làm sạch dữ liệu để khớp với quy cách và nối dữ liệu ấy vào mạng lưới quốc gia giúp nuôi các mô hình, nhưng cũng đặt gánh nặng lớn lên những doanh nghiệp phải trực tiếp làm việc đó. Cần người và tiền để sắp xếp 500.000 tham số, biến 10.000 trang tài liệu tiêu chuẩn thành dữ liệu, xây dựng một đường ống lọc giấy phép và lỗ hổng, chỉ định một lãnh đạo chịu trách nhiệm về dữ liệu, và đáp ứng tiêu chuẩn quốc gia. Các biểu mẫu báo cáo đã có sẵn thủ tục giải thích những biện pháp an toàn nào đã được áp dụng khi dữ liệu được cung cấp ra bên ngoài.

Vì doanh nghiệp vừa phải điền kết quả vào các bảng đã được quy định, vừa phải giải trình về an toàn, họ cùng lúc mang gánh nặng sản xuất và gánh nặng canh giữ. Nếu chỉ đọc gánh nặng này như hiệu quả chính sách, ta sẽ bỏ lỡ một nửa bức tranh. Khoảnh khắc dữ liệu trở thành tài sản, chi phí tạo ra, quản lý và giải trình về tài sản ấy cũng xuất hiện. Doanh nghiệp lớn còn có khoảng trống để gán các chi phí đó, nhưng với doanh nghiệp nhỏ, cùng những biểu mẫu ấy lại

trở thành một ngưỡng nặng nề. Việc cùng một quy cách nhưng không có cùng sức nặng đối với mọi người là phần bóng tối cần luôn được giữ trong tầm mắt khi đọc chính sách này.

Việc dùng dữ liệu vượt ra ngoài hàng rào nhà máy. Trong khi Cục Dữ liệu Quốc gia biến dữ liệu thành ngôn ngữ của chính sách công nghiệp, các bộ khác dùng cùng ngữ pháp ấy trong lĩnh vực của mình. Ý kiến thực hiện về phát triển ứng dụng Trí tuệ nhân tạo Plus trong nhân lực, do Bộ Nhân lực và An sinh xã hội ban hành, là một văn kiện như vậy. Tên gọi có thể nghe lạ, nhưng ý nghĩa không khó nắm. Đó là kế hoạch đưa trí tuệ nhân tạo vào nền hành chính tuyển người, phân công việc, quản lý bảo hiểm xã hội, và đào tạo lại những người mất việc. Nhân lực và an sinh xã hội là những lĩnh vực nơi thông tin cá nhân nhạy cảm di chuyển. Một khi hồ sơ tuyển dụng, lịch sử làm việc, tài liệu đánh giá, và hồ sơ đóng bảo hiểm được biến thành dữ liệu rồi đưa vào mô hình, những vấn đề từng thấy trước đó với các bộ dữ liệu công nghiệp cũng đi theo.

Những câu hỏi cũ lại xuất hiện: dữ liệu nào được phép đưa vào, ai nắm quyền đối với dữ liệu ấy, và phải gắn kèm những biện pháp an toàn nào. Các câu hỏi từng đặt ra với dữ liệu quy trình thép nay lặp lại trước hồ sơ nhân sự con người. Khoảnh khắc quản lý nhân lực bước vào chính sách dữ liệu, việc dùng dữ liệu chuyển từ máy móc sang con người. Khi ngữ pháp được mài giũa trên sàn sản xuất đi vào những nơi xử lý con người, sức nặng của điều đang bị đặt cược cũng thay đổi.

Một điểm mà văn kiện này giao nhau với chính sách dữ liệu là việc làm. Khi trí tuệ nhân tạo tiếp quản công việc của con người, sẽ có người mất chỗ đứng. Các ý kiến triển khai về công tác nhân sự đưa vào trong bộ máy hành chính một thủ tục để xem xét trước cú sốc đó. Hướng đi là buộc các công ty đánh giá rủi ro việc làm khi họ đưa trí tuệ nhân tạo vào sử dụng, đồng thời dùng dữ liệu để quản lý các con đường đào tạo và tái việc làm cho những người có thể mất việc. Dữ liệu vừa trở thành công cụ đẩy con người ra ngoài, vừa là công cụ đo sự dịch chuyển ấy. Đây cũng là nơi chính sách dữ liệu phải được đọc mà không tô đẹp nó, vì cùng một đường ống dữ liệu có thể được dùng để nâng năng suất và để theo dõi lao động chặt hơn.

Việc văn kiện đưa trí tuệ nhân tạo vào quản trị nhân sự đồng thời thiết lập ngưỡng đánh giá rủi ro việc làm cho thấy chính phủ Trung Quốc cũng ý thức được hai mặt này. Đặt ra một thủ tục để đo rủi ro trước cũng có nghĩa là thừa nhận rủi ro tồn tại. Nhưng chỉ dựa vào câu chữ của văn kiện thì khó nói liệu đánh giá rủi ro đó thật sự sẽ hoạt động như một tấm khiên cho người lao động, hay chỉ còn là giấy tờ biện minh cho việc áp dụng. Điểm này nằm giữa một thủ tục đã được xác nhận và một kết quả chưa được xác nhận.

Hãy đi thêm một chút theo ngữ pháp của dữ liệu khi nó đi vào quản trị nhân sự. Trong sản xuất, các tập đoàn thép đã mở dữ liệu từ chính quy trình của họ. Ngay cả khi dữ liệu đó bị dùng sai, thiệt hại chủ yếu rơi vào công ty. Dữ liệu nhân sự thì khác. Khi hồ sơ của những người bị từ chối tuyển dụng, tài liệu của những người nhận điểm đánh giá thấp, và lịch sử của những người

mất việc đi vào một mô hình, người bị tổn hại khi dữ liệu đó bị dùng sai là từng cá nhân. Vì vậy, dù các ý kiến triển khai về công tác nhân sự dùng cùng một ngữ pháp dữ liệu, bản chất rủi ro gắn với ngữ pháp đó không giống dữ liệu sản xuất.

Các thủ tục xác nhận quyền đối với dữ liệu và giải thích biện pháp an toàn cũng đã được yêu cầu tại các cơ sở sản xuất, nhưng ở nơi thông tin cá nhân đi qua đi lại, những thủ tục đó phải có trọng lượng lớn hơn. Việc văn kiện tạo ra ngưỡng gọi là đánh giá rủi ro việc làm có thể được đọc như kết quả của một mức độ ý thức nào đó về trọng lượng ấy. Chương này chỉ xác nhận sự thật rằng ngưỡng đó đã được đặt ra. Ngưỡng ấy vận hành ra sao tại hiện trường vẫn nằm ngoài văn kiện.

Đây là bản đồ dữ liệu được vẽ ra từ các văn kiện của chính phủ. Muốn hiểu vì sao bản đồ này có hình dạng như vậy, cần nhìn vào mục tiêu lớn hơn được đặt ở phía trên nó. Thuật ngữ nền kinh tế thông minh đã bám rễ trong ngôn ngữ chính sách của Trung Quốc. Tín hiệu mà chính phủ phát đi về nền kinh tế thông minh là gì và phải xây dựng nó như thế nào không hề mơ hồ. Hướng đi là không để trí tuệ nhân tạo chỉ còn là một công nghệ nằm trong phòng thí nghiệm, mà biến nó thành nền tảng của toàn bộ nền kinh tế, trải từ sản xuất, tiêu dùng đến hành chính. Trong bức tranh này, mô hình đứng ở phía trước, nhưng dữ liệu mới là thứ giữ cho mô hình ấy sống. Mô hình hiện ra bên ngoài, còn dữ liệu chảy phía sau nó.

Đây cũng là lý do thuật ngữ cộng hưởng tham số-dữ liệu buộc mô hình và dữ liệu vào cùng một cụm từ. Vì mục tiêu của nền kinh tế thông minh nằm ở phía trên, các đường ống dẫn dữ liệu phải được đặt trước để đi tới mục tiêu đó. Công việc của Cục Quản lý Dữ liệu Quốc gia trong việc đặt ra các bộ dữ liệu, biểu mẫu báo cáo và lịch trình liên thông có thể được hiểu là công việc đặt những đường ống ấy. Tuyên bố lớn rằng toàn bộ nền kinh tế sẽ được làm cho thông minh và biểu mẫu nhỏ yêu cầu một bộ dữ liệu thép điền 500.000 tham số đều chỉ về cùng một hướng. Con đường mà chương này đã đi theo là quá trình trong đó một mục tiêu lớn được dịch thành những biểu mẫu nhỏ.

Bên cạnh bức tranh lớn hơn về một nền kinh tế thông minh còn có một loại ngôn ngữ khác: mục tiêu giành trước vị trí cao trong các ứng dụng công nghiệp. Cuộc cạnh tranh không kết thúc ở việc xây dựng một mô hình tốt. Câu hỏi thật sự là địa điểm công nghiệp nào sẽ nhận mô hình đó trước. Cách nhìn này giải thích tính chất của các văn kiện đã bàn ở trên. Mẫu báo cáo "cộng hưởng mô hình-dữ liệu" yêu cầu các nhóm tách riêng kịch bản, mô hình công nghiệp và bộ dữ liệu. Kế hoạch xây dựng bộ dữ liệu yêu cầu họ làm sạch dữ liệu chuyên ngành theo tiêu chuẩn. Sau cùng, cả hai động tác đều cố gắng dùng văn kiện để quản lý việc vị trí cao cho ứng dụng sẽ được xây dựng trước ở đâu.

Thái độ ở đây không dừng lại ở việc khoe khoang về hiệu năng của một mô hình, mà kiểm tra bằng con số xem hiệu năng ấy gắn với các bối cảnh cụ thể như quy trình thép, dự báo thiết bị và quản lý nhân sự tốt đến mức nào. Sự nhấn mạnh vào ứng dụng công nghiệp này đã khiến chính

sách dữ liệu trở nên chi tiết như vậy. Xuyên suốt các văn kiện là một phép tính: cuộc đua không hẳn nằm ở việc quốc gia nào phát hành mô hình lớn hơn trước, mà nằm ở việc quốc gia nào gắn mô hình ấy vào hiện trường trước và đưa kết quả quay trở lại dữ liệu.

Bây giờ chúng ta quay lại thông báo đầu tiên. Thông báo duy nhất về cộng hưởng mô hình-dữ liệu được đăng trên bảng tin của National Data Administration không phải là một văn bản đứng riêng. Phía sau nó là mẫu kế hoạch triển khai để các nhóm thực hiện điền vào, mẫu báo cáo tóm tắt để thu thập kết quả, một kế hoạch xây dựng riêng nhằm đưa các bộ dữ liệu vào dạng chuẩn, và các ý kiến triển khai từ phía nhân sự, mang cùng một ngữ pháp ấy vào quản trị con người. Những văn bản này nối với nhau và tạo thành một dòng chảy.

Khi thông báo đi từ trên xuống, kế hoạch đi từ dưới lên. Khi các kế hoạch ấy được thực hiện, báo cáo lại được gửi lên. Dữ liệu trong các báo cáo ấy sau đó kết nối với mạng dữ liệu quốc gia. Mô hình thu hút sự chú ý ở phía trước dòng chảy này, nhưng lực thực sự giữ cho dòng chảy vận hành đến từ tiêu chuẩn dữ liệu, biểu mẫu và lịch trình. Những gì vẫn vô hình khi chúng ta chỉ nhìn theo tên công ty bắt đầu hiện ra khi chúng ta nhìn theo thứ tự của các văn bản.

Độc giả Việt Nam nên rút ra khoảng ba điều từ chương này. Một là chính sách AI của Trung Quốc xem mô hình và dữ liệu như một cơ thể thống nhất. Cụm từ cộng hưởng mô hình-dữ liệu cô đọng thái độ ấy. Hai là thái độ này không dừng lại ở khẩu hiệu. Nó đi xuống thành các biểu mẫu có ô trống và tiêu chí định lượng. Những con số như độ chính xác 95 percent, độ trễ suy luận 50 milliseconds, 500,000 tham số và 2,000 nút đã được viết sẵn trong mẫu báo cáo. Điều cuối cùng là dòng dữ liệu này không dừng ở nhà máy. Nó lan sang quản trị nhân sự, sang việc làm và an sinh xã hội của con người.

Khi dữ liệu trở thành tài sản, bàn tay quản lý tài sản ấy cũng lớn lên. Dù bàn tay đó là chief data officer trong một công ty, một cán bộ báo cáo trong chính quyền cấp tỉnh, hay một lịch trình liên thông từ National Data Administration, cuối cùng nó đều dẫn về một đường ống. Đường ống ấy càng trơn tru, mô hình càng lớn nhanh. Nhưng phía sau mặt sáng của chính sách còn có một sự thật khác: gánh nặng đặt lên các công ty phải lấp đầy đường ống ấy tăng lên, và lượng thông tin cá nhân được đưa qua đó cũng tăng theo.

Câu hỏi còn lại là đau kia của dương ông ay nam o đau. Mục kết nối liên thông trong mẫu báo cáo đã ghi sẵn ngay dữ liệu của từng công ty sẽ được đưa vào hạ tầng dữ liệu quốc gia. Khi ngày đi qua đi và dữ liệu thực sự được nối vào một hệ thống, khoảng cách giữa các công ty xây dựng mô hình và các bộ, ngành định nghĩa dữ liệu sẽ ngắn hơn hiện nay. Khoảng cách đó sẽ thu hẹp đến mức nào, ai sẽ vẫn là chủ sở hữu dữ liệu trong không gian đã thu hẹp ấy, và cái giá nào sẽ quay lại cho chi phí doanh nghiệp khi lập data ông ay cũng như cho thông tin của người lao động được chuyển qua đó, tất cả vẫn nằm ngoài các văn bản.

Chương này đã đi theo các văn bản đến dung ranh giới mà chúng vẽ ra. O đau xa của dòng chảy này, nơi 500,000 tham số tu quy trình luyện tập đi vào mạng lưới quốc gia và hồ sơ nhân

su của người lao động trở thành dữ liệu theo cùng một ngữ pháp, nhưng biểu mẫu tiếp theo sẽ dán ghi vào các báo cáo mà chúng thu thập việc mô hình và dữ liệu sẽ được ghép với nhau chặt chẽ đến mức nào.

Chương 5. DeepSeek và những con hổ AI: hai vòng lặp do mã nguồn mở tạo ra

Bất cứ ai mở trang xếp hạng lượt tải xuống của Hugging Face vào tháng 12 năm 2025 đều sẽ thấy một cảnh tượng lạ. Trong suốt tháng 11 và tháng 12 năm đó, bảy trong mười mô hình được tải xuống nhiều nhất trên nền tảng mã nguồn mở toàn cầu này là mô hình Trung Quốc. Các nhà phát triển cũng đang tải lên các mô hình phái sinh, lấy mô hình của người khác, sửa đổi rồi đăng lại, với tốc độ gần gấp đôi so với các mô hình dựa trên mô hình của Mỹ. Chỉ không lâu trước đó, các mô hình mở của Trung Quốc vẫn bị đánh giá là chậm hơn các mô hình tuyến đầu của Mỹ khoảng bảy hoặc tám tháng về hiệu năng. Trên bảng xếp hạng, khoảng cách ấy đã bị đảo ngược.

Điểm khởi đầu của cảnh tượng này là một báo cáo kỹ thuật: DeepSeek-V3 Technical Report của DeepSeek, do DeepSeek () công bố. Tài liệu phát hành dưới tên công ty này đặt các con số về tài nguyên huấn luyện lên trước mọi lời tự hào về hiệu năng mô hình. Những con số ấy làm lung lay lẽ thường của ngành. Chương này bắt đầu từ những con số đó, rồi lần theo con đường DeepSeek đã mở ra, cách Moonshot, MiniMax và Zhipu lao theo con đường ấy, và hai vòng phản hồi khác nhau đang vận hành phía sau hậu trường. Nếu chỉ nhìn AI Trung Quốc như một bảng xếp hạng công ty, chúng ta sẽ bỏ lỡ một nửa bức tranh, và nửa bị bỏ lỡ ấy nằm trong chính hai vòng phản hồi này.

Những Con Số Đã Thay Đổi Cuộc Tranh Luận

Trước khi DeepSeek xuất hiện, niềm tin chi phối ngành AI toàn cầu là cái gọi là Scaling Laws. Đây là một phát hiện dựa trên thực nghiệm: hiệu năng của một mô hình cải thiện theo cách có thể dự đoán khi số lượng tham số, lượng tính toán dùng trong huấn luyện và quy mô bộ dữ liệu tăng lên. Niềm tin đó tự nhiên dẫn tới kết luận rằng càng lớn càng tốt, và kết luận ấy lại biện minh cho một mô hình phát triển khép kín cần nguồn vốn khổng lồ.

Để duy trì hàng nghìn bộ xử lý đồ họa hoạt động trong thời gian dài, các công ty còn phải trả tiền cho năng lực dự phòng trước những lỗi phần cứng chắc chắn sẽ xuất hiện trên đường vận hành. Chỉ một số ít công ty Mỹ có thể chịu được chi phí đó. Lập luận rằng năng lực chịu đựng chi phí tính toán đắt đỏ trong các đợt chạy dài tự nó đã là năng lực kỹ thuật đang dần trở thành lý do để để AI nằm trong tay một số ít bên.

Báo cáo DeepSeek V3 đã va thẳng vào tiền đề đó. Mô hình này dùng kiến trúc Mixture-of-Experts, hay MoE, với 671 tỷ tham số, 671B. Nhưng khi xử lý một token đơn lẻ, chỉ 37 tỷ tham số, 37B, thật sự được kích hoạt. Thiết kế này giữ cho toàn bộ mô hình có quy mô rất lớn, trong khi phần được dùng ở mỗi lần xử lý vẫn nhỏ. Trên nền đó, DeepSeek bổ sung Multi-head Latent Attention, MLA, và kiến trúc DeepSeekMoE, rồi dùng cách tiếp cận không có auxiliary loss

để phân bổ tải đều trong quá trình huấn luyện mà không cần một hàm mất mát phụ.

Phân bổ tải đều có nghĩa là cân bằng quá trình huấn luyện để công việc không dồn vào chỉ vài expert, trong khi các expert còn lại ngồi không. Các hệ thống trước đây thêm một thành phần loss riêng để giữ cân bằng đó, nhưng thành phần này có tác dụng phụ là làm hao hụt hiệu năng cốt lõi của mô hình. DeepSeek viết rằng họ đã tìm ra cách giữ cân bằng ngay cả khi không có thành phần ấy. Dữ liệu dùng cho tiền huấn luyện lên tới 14,8 nghìn tỷ token, tức 14.8T. Dù vậy, toàn bộ lượt huấn luyện chỉ mất 2,788 triệu giờ trên GPU H800.

Điểm lặng lẽ nhưng nặng ký trong báo cáo là độ ổn định khi huấn luyện. Khi một mô hình được huấn luyện ở quy mô này, sự cố xảy ra là chuyện thường: giá trị loss có thể đột ngột tăng vọt đến mức không thể cứu, hoặc các kỹ sư có thể phải quay lại một checkpoint trước đó. Một sự cố như vậy có thể làm lãng phí vài ngày tính toán. DeepSeek viết rằng không có sự cố nào như thế xảy ra trong toàn bộ quá trình huấn luyện. Nói cách khác, họ giữ hệ thống ổn định bằng kỹ thuật, chứ không phải bằng cách đổ thêm vốn.

Vì không còn không gian để xây dựng lớn hơn, họ tìm cách xây dựng tiết kiệm hơn, và phương pháp ấy được ghi lại trong báo cáo bằng các con số. Cần giữ một điểm trong đầu ở đây. H800 là bộ xử lý đồ họa đã bị giảm hiệu năng để Hoa Kỳ có thể xuất khẩu sang Trung Quốc. Việc những con số này xuất hiện trong điều kiện DeepSeek không thể có được phần cứng cao cấp nhất khiến báo cáo có thêm một tầng ý nghĩa.

Khi những con số này đến với thị trường, phản ứng đã hiện ra trong các chỉ báo tài chính. Theo báo cáo của Samjong KPMG, ngay sau khi DeepSeek công bố mô hình R1 vào tháng 1 năm 2025, tranh luận bùng lên quanh tuyên bố rằng R1 đã được huấn luyện bằng khoảng 5,6 triệu USD tài nguyên tính toán, khác với các đối thủ Mỹ vốn thường chi hơn 100 triệu USD để xây dựng các mô hình tuyến đầu. Ngay sau khi thông tin được công bố, Nvidia, công ty đứng ở trung tâm hệ sinh thái phần cứng AI, chứng kiến giá cổ phiếu giảm hơn 15 phần trăm. Cổ phiếu của các công ty liên quan, gồm Oracle và Constellation Energy, từng được xem là bên hưởng lợi từ nhu cầu điện của trung tâm dữ liệu, cũng giảm.

Cú sốc không dừng lại ở một nhà sản xuất bán dẫn. Nó lan ra toàn bộ chuỗi giá trị hạ tầng. Câu hỏi của thị trường rất sắc. Nếu một mô hình được xây dựng với 5,6 triệu USD có thể làm gần như cùng công việc với một mô hình 100 triệu USD, thì hàng trăm tỷ USD chi tiêu vốn mà các công ty Mỹ đang đổ vào hạ tầng có còn hợp lý không? Câu hỏi ấy đã nổi lên. Dù vậy, con số 5,6 triệu USD gần với cách đếm tài nguyên tính toán được dùng cho một lần huấn luyện cuối cùng. Nên nhớ rằng đó không phải tổng chi phí bao gồm phần nghiên cứu trước đó, các lần thử thất bại và chi phí chuẩn bị dữ liệu.

Khoảng cách về chi phí suy luận cũng rất rộng. Theo các số liệu được trích trong một báo cáo của U.S.-China Economic and Security Review Commission, hay USCC, các mô hình lớn của Mỹ có mức giá kết hợp từ 6 đến 10 USD cho mỗi triệu token, trong khi các mô hình Trung Quốc

thường được định giá dưới 1 đến 3 USD. Khác biệt còn lớn hơn khi nhìn vào từng con số trong báo cáo của Samjong KPMG. Chi phí đầu ra cho mỗi triệu token là 60,0 USD đối với o1 của OpenAI, 15,0 USD đối với Claude 3.5 Sonnet của Anthropic, và 3,5 USD đối với Llama 3.1 405B của Meta. DeepSeek R1, được đánh giá là có hiệu năng suy luận tương tự o1, có chi phí đầu ra là 2,2 USD.

Cái giá để bán cùng một loại trí tuệ đã chênh nhau theo cấp số lớn. Khi giá giảm sâu đến mức này, các công ty vừa và nhỏ có nền vốn mỏng, cùng những ngành bám rễ trong nền kinh tế vật chất, cũng có thể đưa AI vào tay mình. Vì thế, nhiều người đọc tin hiệu này như dấu hiệu cho thấy rào cản gia nhập đang hạ xuống. Chi phí suy luận là khoản tiền phải trả lặp đi lặp lại mỗi khi người dùng chạy mô hình sau khi quá trình phát triển đã hoàn tất, nên một con số thấp ở đây không dừng lại ở một thành tích một lần. Nó tích lũy thành năng lực cạnh tranh hằng ngày.

Một lối rẽ bên trong các bài báo kỹ thuật

Mức chi phí thấp này không đến từ việc cắt giảm chi phí lao động hay ép phần cứng vượt quá giới hạn của nó. Nó đến từ một lối rẽ đã thay đổi chính cấu trúc. DeepSeek đưa dự đoán nhiều token, tức MTP, vào mục tiêu huấn luyện của mình. Thay vì học cách dự đoán chỉ token kế tiếp tại từng thời điểm, mô hình học cách nhìn trước nhiều token cùng lúc. Hồ sơ đáng chú ý ở đây đến từ một đối thủ cạnh tranh. Các nhà nghiên cứu đứng sau GLM-5 của Zhipu đã xem xét thiết kế này trong bài báo của họ và để lại một nhận xét. Phương pháp MTP ban đầu có một điểm yếu: trong quá trình suy luận, nó phải đi qua nhiều lớp, nên khi các bước giải mã suy đoán tăng lên, mức dùng bộ nhớ cho tham số và bộ đệm key-value, tức KV cache, cũng tăng tuyến tính.

Họ viết rằng DeepSeek V3 đã cho ba lớp MTP dùng chung tham số với nhau, nhờ đó nâng tỷ lệ token được chấp nhận mà không dùng thêm bộ nhớ. Đây là cảnh một tài liệu của đối thủ thừa nhận thiết kế của bên cạnh tranh và học từ nó. Sự thừa nhận như vậy không phải phép lịch sự thông thường trong thế giới mã nguồn mở. Nó là bằng chứng cho thấy một thiết kế công khai đã thật sự đến được tay công ty kế tiếp và chạy được ở đó.

Các thiết kế attention thừa mà DeepSeek công bố sau đó cũng đi theo hướng này. Khi xử lý ngữ cảnh dài tới 128K, attention dày đặc truyền thống phải gánh một chi phí lớn: câu càng dài, lượng tính toán tăng theo bình phương. Nếu ngữ cảnh dài gấp đôi, lượng tính toán tăng gấp bốn. DeepSeek giảm gánh nặng này bằng cách đưa vào một phương pháp thưa, phân bổ tài nguyên attention rất chi tiết theo nội dung và mức độ quan trọng của từng token. Thay vì so sánh mọi token với mọi token khác, ý tưởng là chỉ dồn tính toán vào những liên kết quan trọng.

Cần nói rõ một điểm ở đây. Các bài báo của DeepSeek nằm rải trong nhiều tài liệu, nên khi đưa tên gọi và con số của một kỹ thuật cụ thể vào một nguồn đã chốt, phải kiểm tra lại nguyên văn. Vì vậy, chương này chỉ xem là đã xác lập một sự thật: thiết kế ấy đi theo hướng hạ chi phí tính toán của ngữ cảnh dài. Chương này không nêu các con số chi tiết, chẳng hạn chi phí đã

giảm bao nhiêu lần.

Tầm quan trọng của cách xử lý này trở nên rõ hơn khi nhìn vào cách một công ty khác tiếp nhận nó. Zhipu nhận định rằng xử lý ngữ cảnh dài sẽ đòi hỏi phải thay đổi kiến trúc dày đặc hiện có, nhưng huấn luyện lại từ đầu sẽ tốn một khoản khổng lồ. Vì vậy, công ty chọn đặt thiết kế attention thưa mà DeepSeek đã mở mã nguồn lên trên mô hình nền dày đặc của chính mình, rồi nối hai phần đó thông qua Continued Pre-Training.

Thay vì xây lại từ gốc, Zhipu có được năng lực xử lý ngữ cảnh dài bằng cách gán kiến trúc của mình vào một thiết kế mà đối thủ đã mở ra. Đó là một quyết định linh hoạt trong sức ép từ các biện pháp kiểm soát xuất khẩu bán dẫn của Hoa Kỳ, vốn đã trói chặt tài nguyên tính toán. Việc một công ty công bố kỹ thuật đã mua thời gian cho một công ty khác, và khoảng thời gian vay mượn ấy trở thành khung cho câu chuyện tiếp theo.

Chu kỳ đầu tiên do tính mở tạo ra

Chi phí không phải là điều duy nhất DeepSeek đã thay đổi. Nó đã đẩy một cấu trúc quản trị khép kín sang hướng các mô hình trọng số mở. Báo cáo của USCC giải thích sự khác nhau giữa mô hình khép kín kiểu Mỹ và mô hình trọng số mở kiểu Trung Quốc. Một mô hình khép kín chỉ có thể được dùng thông qua giao diện do nhà phát triển xây dựng, còn một mô hình trọng số mở cho phép người dùng tải xuống và tự nắm giữ các tham số đã được huấn luyện. Cách thứ nhất mở cửa, nhưng không cho bạn bước vào căn phòng. Cách thứ hai trao luôn chiếc chìa khóa của căn phòng. Người nhận chìa khóa có thể dịch chuyển đồ đạc bên trong và sơn lại tường. Sự khác biệt này đã tạo ra cơ chế phản hồi đầu tiên trong chương này, vòng lặp phản hồi của các mô hình mở.

Cơ chế đó vận hành như sau. Một công ty Trung Quốc hàng đầu phát hành miễn phí trọng số của một mô hình hàng đầu trên một nền tảng như Hugging Face. Các nhà phát triển trên khắp thế giới tải nó xuống, tinh chỉnh cho những trường hợp sử dụng hẹp của riêng họ, rồi tải các mô hình phái sinh đó trở lại nền tảng. Các bản ghi sử dụng và những sửa lỗi từ cộng đồng tích tụ qua quá trình này sau đó trở thành nguồn lực để cải thiện mô hình gốc. Mỗi lần vòng lặp này hoàn thành một vòng, trọng tâm của hệ sinh thái lại dịch thêm một chút về phía các mô hình Trung Quốc. Nhiên liệu ở đây là sự tham gia, không phải vốn. Nếu chiến lược khép kín của Mỹ giấu mô hình sau một giao diện và thu phí truy cập, thì chiến lược mở của Trung Quốc phát hành mô hình rồi kéo hệ sinh thái mọc lên trên nó về phía mình.

Các số liệu cho thấy sức mạnh của vòng lặp này. Theo dữ liệu dự án ATOM, các mô hình Trung Quốc đã vượt qua các mô hình Mỹ về tổng lượt tải xuống trên Hugging Face vào tháng 8 năm 2025. Đến tháng 10, các công ty Trung Quốc như Zhipu (Z.ai), Alibaba, DeepSeek, Moonshot AI và Meituan đã vươn lên nhóm đầu trên LMArena, nơi đo lường sở thích của con người, đẩy các mô hình của Meta sang một bên.

Điều đó dẫn tới những con số đã nhắc ở trên: bảy trong mười lượt tải lên đứng đầu trong tháng 11 và tháng 12, và tốc độ tải lên các mô hình phái sinh tăng gấp đôi. Các báo cáo trong ngành tại Trung Quốc cũng mô tả Trung Quốc năm 2026 như một cực của bản đồ công nghệ, và độc chuyển động này theo cùng hướng. Dù vậy, số lượt tải xuống không có nghĩa là được dùng thật trong các nhà máy và công trường công nghiệp. Nhiều mô hình được tải xuống, thử nghiệm, rồi bị bỏ lại.

Cách tri thức di chuyển qua vòng lặp này là chưng cất tri thức. DeepSeek R1 đã tự học được năng lực suy luận của mình, tức chuỗi suy nghĩ cho thấy các bước logic mà nó đi qua để đi tới câu trả lời. Năng lực đó được cấy sang các mô hình dense nhỏ hơn, từ 1,5 tỷ tham số (1.5B) đến 70 tỷ tham số (70B), xây trên Qwen của Alibaba hoặc Llama của Meta.

Khi trí thông minh của một mô hình lớn được đưa vào một mô hình nhỏ hơn, một hệ sinh thái gọn nhẹ mở ra, nơi các mô hình suy luận mạnh có thể chạy với chi phí thấp trên laptop hoặc máy chủ nội bộ. Chưng cất dạy mô hình nhỏ bằng cách dùng các câu trả lời do mô hình lớn tạo ra, cùng với quá trình dẫn tới các câu trả lời đó, làm người thầy. Mô hình nhỏ hơn có thể kế thừa phần lớn năng lực mà không cần sao chép chính bản gốc.

Một trường hợp cụ thể cho thấy rõ chuyển động này. Cuối năm 2025, mô hình đứng đầu về lượt tải xuống trên Hugging Face là mô hình nhỏ tạo chú thích video của ByteDance (, ByteDance), 'Tarsier2-Recap-7b.' Nó không được xây dựng từ con số không. Nó là một phái sinh của mô hình mã nguồn mở 'Qwen2-VL-7B' của Alibaba, được ByteDance tinh chỉnh cho các video ngắn của chính mình. Trên nền một mô hình công khai của một công ty Trung Quốc, một công ty Trung Quốc khác đặt lên đó một mô hình chuyên biệt, và mô hình ấy trở thành mô hình được tải xuống nhiều nhất thế giới.

Thay vì xây dựng một mô hình đa dụng từ đầu, phương pháp này đặt các năng lực chuyên biệt lên trên một nền tảng công khai mạnh. Đó là nhịp điển hình của vòng lặp mã nguồn mở. Alibaba đã để cánh cửa mở, nên ByteDance bước qua. Mô hình chuyên biệt mà ByteDance đặt lên trên sau đó lại kéo các nhà phát triển trên khắp thế giới quay về kiến trúc nền của Alibaba.

Một câu mà các nhà nghiên cứu Zhipu để lại trong bài báo của họ cho thấy rõ tính chất của vòng lặp này. Ý của họ là vì DeepSeek phát hành các tham số của mình một cách ẩn danh, mô hình có thể vượt qua thiên kiến địa chính trị, và cộng đồng toàn cầu chấp nhận nó vì nó hoạt động tốt trong thực tế. Đó là một sự thừa nhận rằng hiệu năng đã làm con người dịch chuyển trước cả ranh giới quốc gia. Chiến lược mở vận hành theo cách này: nó xóa dấu hiệu quốc tịch, trong khi dòng chảy do mô hình không mang nhãn quốc gia tạo ra lại âm thầm quay về phía Trung Quốc.

Ba công ty vạch ra đường bắt kịp nhanh

Ba công ty tiến nhanh trên con đường mà DeepSeek mở ra là Moonshot AI (, Moonshot AI), MiniMax (, MiniMax), và Z.ai (, Z.ai), thường được xếp vào nhóm gọi là AI Tigers. Bản thân cách gom nhóm này gần với một nhãn gọi tiện dùng của truyền thông và các báo cáo ngành hơn, nhưng tên pháp lý bằng tiếng Trung và tên sản phẩm của từng công ty đều có thể được xác nhận trong tài liệu công khai. Vì vậy, cuốn sách này đặt các công ty thật, sản phẩm của họ, và cách họ đi theo con đường DeepSeek đã mở ra lên trước biệt danh đó.

MiniMax (, MiniMax) là một công ty mô hình nền tảng bắt đầu hoạt động vào đầu năm 2022. Công ty này vẫn mang dấu ấn của các nhà nghiên cứu đến từ SenseTime (, SenseTime), và đã tiến nhanh trong thương mại hóa cũng như mở rộng ra nước ngoài. Tháng 1 năm 2026, công ty niêm yết trên Sở Giao dịch Chứng khoán Hong Kong, huy động 619 triệu đô la, và giá cổ phiếu tăng gấp đôi trong ngày giao dịch đầu tiên. Tuy vậy, quy mô của đợt niêm yết này là một chi tiết cần được kiểm tra lại với nguồn gốc, vì các tài liệu hỗ trợ có phần chông lẩn giữa nhiều bản tường thuật.

Dòng sản phẩm mà MiniMax giới thiệu trên trang web công ty trải dài từ mô hình ngôn ngữ, tạo video, đến công cụ agent. Mô hình tạo video miễn phí Hailuo của công ty được đánh giá là có khả năng bắt chước môi trường vật lý ở mức có thể so sánh với Sora của OpenAI. Tầm bao phủ từ ngôn ngữ đến video cho thấy hướng đi của công ty: trước hết giành được sự nhận biết thông qua các mô hình công khai, sau đó chuyển sự nhận biết tên tuổi ấy thành sản phẩm thương mại.

Những thay đổi trong kiến trúc attention mà MiniMax đã dùng cho thấy các mối quan tâm kỹ thuật của công ty. Ở các mô hình thế hệ trước, công ty pha trộn Lightning Attention, nhưng đến dòng M2 về sau lại quay về full multi-head attention để giữ ổn định. Đây là quá trình một công ty lùi lại một bước, rồi tiến lên tiếp, giữa một thiết kế tìm kiếm tốc độ và một thiết kế bảo vệ độ ổn định. Một kiến trúc mới không phải lúc nào cũng chỉ đi về phía trước. Khi sự bất ổn xuất hiện trong vận hành thực tế, chọn quay lại một phương pháp đã được chứng minh cũng là một phần của năng lực kỹ thuật.

Ở thế hệ M2.7 về sau, các báo cáo mô tả những chuyển động có thể được đọc như tín hiệu ban đầu của self-evolution, tức việc một mô hình tự cải thiện chính nó. Điều này chỉ nên được xem là một hướng cần kiểm chứng sau, không phải một thành tựu đã được xác lập. Tuyên bố rằng một mô hình có thể tự sửa mình nghe hấp dẫn, nhưng các tài liệu hiện có chỉ đủ nâng đỡ cụm từ thận trọng "tín hiệu ban đầu." Cụm từ "một mô hình tự tiến hóa" gần với ngôn ngữ của rủi ro chưa được kiểm chứng hơn, và cuốn sách này không đặt nó vào vị trí của sự thật đã được xác nhận.

Moonshot AI (, Moonshot AI) là một thế lực mới nổi, tạo được tên tuổi nhờ các mô hình mở được tải xuống rộng rãi. Công ty đã lớn đủ để nhắm tới mức định giá 10 tỷ USD trong một vòng đầu tư mới. Dòng Kimi chủ lực của hãng đặt trọng tâm vào xử lý ngữ cảnh dài và trí tuệ

tác nhân thị giác, còn cái gọi là "mô hình tư duy", được cho là có khoảng 1 nghìn tỷ (1T) tham số, đã làm rung chuyển hệ sinh thái mã nguồn mở.

Nhìn vào các kho GitHub của Moonshot có thể thấy dấu vết của một công ty đang đi vượt ra ngoài chatbot để bước vào nhịp làm việc hằng ngày của lập trình viên. Công ty đã phát hành điểm khởi đầu cho một tác nhân lập trình chạy các tác nhân thể hệ mới trong terminal, mở mã nguồn các kernel hiệu năng cao để tốc độ không tụt lại trong ngữ cảnh dài, và vận hành một cửa sổ hướng dẫn có nhiệm vụ đưa câu hỏi qua nhiều sản phẩm tới đúng nơi cần đến.

Công ty không dừng lại ở việc phát hành mô hình. Họ đã mở ra những con đường để lập trình viên có thể dùng các mô hình ấy trong công việc thật. Khi nói rằng một vòng lặp mã nguồn mở quay về phía cộng đồng lập trình viên, điều đó đi cùng với việc đặt sẵn những con đường này. Phát hành một mô hình và làm cho mô hình ấy dễ dùng là hai dạng lao động khác nhau, và Moonshot đã gánh cả hai.

Zhipu (, Z.ai) là một công ty bắt rễ từ các nhà nghiên cứu của Đại học Thanh Hoa và đã tiếp tục dòng GLM. Ngày 16 tháng 1 năm 2025, công ty bị đưa vào Entity List của Bộ Thương mại Hoa Kỳ và chịu một cú đánh, nhưng thay vì đối đầu trực diện, họ chọn con đường đi vòng đã nói ở phần trước. Dòng GLM-5 mới nhất đã vượt ra ngoài việc tạo câu và chuyển trọng tâm sang lập trình cùng các tác nhân tự trị, và nhan đề bài báo của họ nắm bắt trực tiếp hướng đi đó. Câu chuyện ở đây là bước chuyển từ viết mã bằng trực giác sang một giai đoạn mà tác nhân nắm quyền điều hành quy trình.

Khi các lệnh trừng phạt trói chặt tài nguyên tính toán, lựa chọn của công ty là nổi thiết kế đã công bố của một đối thủ cạnh tranh với quá trình tiếp tục tiền huấn luyện và năng lực xử lý ngữ cảnh dài an toàn. Điều đó cho thấy vòng lặp mã nguồn mở có thể hạ thấp bức tường trừng phạt ra sao. Trái với ý định của những người đã khóa cửa, một thiết kế mở lại tạo khoảng thở cho một công ty bị nhốt phía sau cánh cửa ấy. Các biện pháp kiểm soát xuất khẩu nhắm vào phần cứng, nhưng những thiết kế được phát hành dưới dạng mã nguồn mở đã mở ra một con đường để sự khéo léo của phần mềm bù đắp cho tình trạng thiếu phần cứng.

Đây là nơi cần thận trọng. Các bảng xếp hạng so sánh những mô hình này với các mô hình hàng đầu của Hoa Kỳ vẫn liên tục xuất hiện, nhưng giới phê bình cũng chỉ ra rằng các chỉ số dùng để đo mức độ lan tỏa vẫn chưa ổn định thành một chuẩn chung. Vì vậy, thay vì tuyên bố rằng một mô hình đã "tiến sát rất nhanh" tới mô hình khác hoặc đã "vượt qua" nó, cách an toàn hơn là đọc xem dòng tải xuống và các bản phái sinh được tải lên đang nghiêng về phía nào. Dòng V3.2 của DeepSeek, được cho là phát hành trong khoảng từ cuối tháng 11 đến đầu tháng 12 năm 2025, đã được mô tả là đạt tới trình độ của một thí sinh đoạt huy chương vàng Olympic Toán học Quốc tế, nhưng ngay cả ngày phát hành cũng khác nhau giữa các nguồn, với cả ngày 29 tháng 11 và ngày 1 tháng 12 cùng xuất hiện. Vì vậy, văn bản này đặt hai ngày cạnh nhau, thay vì trình bày thành tựu đó như một sự thật đã được xác lập.

Những con số trong các bảng xếp hạng trông rất ấn tượng, nhưng chúng ta cũng nên nhớ rằng vẫn chưa có một thẩm quyền duy nhất được đồng thuận để đo lường chúng. Vì cùng lý do đó, cuốn sách này tránh những tuyên bố rằng các thiết bị tính toán dòng Ascend do Trung Quốc sản xuất đứng ngang hàng với các thiết bị hàng đầu của Hoa Kỳ.

Vòng lặp thứ hai

Trọng tâm của khung phân tích mà báo cáo USCC dùng để giải thích AI của Trung Quốc nằm ở chỗ có hai vòng phản hồi, chứ không phải một. Vòng mã nguồn mở đã bàn ở trên vận hành giữa các nhà phát triển phần mềm trên internet. Vòng thứ hai, vòng phản hồi của nền kinh tế vật chất, vận hành tại các địa điểm thực tế nơi nhà máy, hệ thống logistics và robot hoạt động.

Khi AI phương Tây đặt trọng tâm vào lưu lượng web và dịch vụ chatbot, Trung Quốc đưa mô hình thẳng vào hạ tầng sản xuất lớn nhất thế giới. Sau đó, nước này thu được dữ liệu về quy trình và vận hành mà không thể có được bằng cách quét web, rồi dùng dữ liệu đó để sửa lại các mô hình công nghiệp. Bản ghi hội thoại từ chatbot và bản ghi điều khiển từ thiết bị nhà máy là hai loại tài nguyên khác nhau, và loại sau không thể được tạo ra bên ngoài chính nhà máy đó.

Vòng thứ hai này khác vòng thứ nhất vì nó vận hành trên nền hướng dẫn hành chính của chính phủ. Chính phủ Trung Quốc đã đưa vào kế hoạch hành động sản xuất các mục tiêu phát hành những AI agent có thể dùng ở quy mô lớn trong môi trường công nghiệp và xây dựng các bộ dữ liệu chất lượng cao; các báo cáo ngành trong nước cũng xem mục tiêu này là một trục trong bức tranh công nghệ của Trung Quốc. Nếu vòng mã nguồn mở vận hành nhờ sự tham gia tự nguyện của các nhà phát triển, thì vòng kinh tế vật chất vận hành khi các công ty cung cấp dữ liệu theo những tuyến mục tiêu đã được văn kiện chính sách vạch sẵn từ trước.

Một bên lớn lên từ dưới lên, còn bên kia đi xuống từ trên xuống. Các con số chi tiết và cách thực hiện mục tiêu này cần được đối chiếu với văn bản gốc trong các chương sau về chính sách dữ liệu và chính sách công nghiệp. Ở đây, điểm duy nhất được dùng làm cơ sở là vòng thứ hai vận hành theo sự hướng dẫn của nhà nước.

Đọc riêng hai vòng lặp

Để đọc chính xác chuyển động của AI Trung Quốc, chúng ta cần giữ rõ một sự thật: hai vòng lặp này khác nhau về tính chất và mục đích. Dữ liệu của chúng đến từ các nguồn khác nhau. Vòng lặp mã nguồn mở chạy trên mã công khai và dữ liệu web, rồi mở rộng hệ sinh thái phần mềm bằng cách biến tính mở, thứ ai cũng có thể tiếp cận, thành vũ khí của mình. Trái lại, vòng lặp kinh tế thực chạy trên dữ liệu vật lý độc quyền từ cảm biến trên sàn nhà máy, nhiệt độ của sắt nóng chảy, và động cơ điều khiển của robot.

Ngay cả khi các công ty Mỹ thu thập dữ liệu từ web và tiến lên trong vòng lặp mã nguồn mở, dữ liệu điều khiển đang đổ ra theo thời gian thực bên trong các dây chuyền sản xuất của Trung Quốc vẫn không thể bị sao chép hay thu thập từ bên ngoài. Khung phân tích này xem dữ liệu đó

là một rào cản gia nhập khó thách thức. Dù vậy, tài liệu hiện có vẫn chưa đủ để xác định quy mô của rào cản ấy thành một con số cố định, nên ở đây tôi chỉ ghi nhận hướng lập luận.

Hai vòng lặp này cũng gặp các cơ quan nhà nước khác nhau. Các mô hình trong vòng lặp mã nguồn mở thuộc phạm vi kiểm duyệt nội dung và thủ tục đăng ký của Cyberspace Administration of China, hay CAC, và khi xuất hiện đầu ra có vấn đề, doanh nghiệp phải gánh trách nhiệm huấn luyện lại mô hình. Cơ quan này xem xét một mô hình được phát hành như dịch vụ công cộng có thể đưa ra những câu trả lời nào. Các mô hình công nghiệp trong vòng lặp kinh tế thực phải tuân theo các tiêu chuẩn an toàn do Ministry of Industry and Information Technology, hay MIIT, và Standardization Administration of China, hay SAC, đặt ra. Chúng phải vượt qua các tiêu chuẩn vật lý cho phép con người can thiệp vào công việc rủi ro cao và cho phép hệ thống dừng lại khi việc điều khiển gặp lỗi.

Hình thức kiểm soát thứ nhất hướng vào nội dung và hệ tư tưởng, còn hình thức thứ hai hướng vào máy móc và an toàn. Những bàn tay khác nhau của cùng một chính phủ chạm riêng vào hai vòng lặp này. Sự đối lập ấy sẽ quay lại ở phần sau, khi văn bản lần lượt bàn đến từng cơ quan nhà nước.

Công bằng mà nói, ở đây cũng phải nhìn thấy bóng tối ở phía bên kia. Sự mở cửa và chi phí thấp đã tạo ra một mặt sáng, nhưng chúng cũng kéo theo gánh nặng ở phía sau. Khi các mô hình trong vòng lặp mã nguồn mở phải vượt qua thủ tục kiểm duyệt và phải huấn luyện lại mỗi khi xuất hiện đầu ra có vấn đề, doanh nghiệp tiếp tục trả các chi phí tuân thủ không liên quan đến hiệu năng. Lợi ích từ việc phát hành mô hình và mở rộng hệ sinh thái bị buộc chặt với chi phí liên tục điều chỉnh các mô hình ấy để chúng không vi phạm quy định. Trong vòng lặp kinh tế thực, yêu cầu cung cấp các tập dữ liệu lớn và vượt qua tiêu chuẩn an toàn cũng tạo ra một ngưỡng khó đối với các doanh nghiệp ở tuyến đầu.

Thu thập dữ liệu và gắn nhãn dữ liệu đều cần lao động của con người, nhưng điều kiện diễn ra thứ lao động ấy không xuất hiện trên các bảng xếp hạng. Đằng sau các điểm benchmark được đánh bóng và các bảng xếp hạng lượt tải xuống là những chi phí kiểu này, và tốt hơn là không nên bỏ qua chúng chỉ vì nhìn vào bảng điểm. Ở phần sau, trong chương bàn về các phản biện, cuốn sách sẽ xem xét gánh nặng quản lý này và cái bóng của lao động theo đúng nghĩa của chúng.

Khi hai vòng lặp này được tách riêng ra, nửa phần biến mất khi AI Trung Quốc chỉ được đọc qua bảng xếp hạng doanh nghiệp sẽ hiện ra. Những cái tên DeepSeek, Moonshot, MiniMax và Zhipu xuất hiện trên màn hình dưới dạng số lượt tải xuống và điểm benchmark. Những con số ấy là bảng điểm của vòng lặp mã nguồn mở. Nhưng bên dưới sân chơi nơi chính các công ty ấy cạnh tranh còn có một vòng lặp thứ hai, nơi dữ liệu nhà máy chảy qua, và vòng lặp ấy vận hành trong những đường ranh được các văn kiện nhà nước vạch ra trước. Tựa đề của chương này, hai vòng lặp do mã nguồn mở tạo ra, có nghĩa là một chiến lược, sự mở cửa, đang đồng thời

vận hành hai chu kỳ có tính chất khác nhau.

Một vòng lặp là vòng lặp phần mềm, kéo các nhà phát triển trên khắp thế giới về phía các mô hình trong nước. Vòng lặp kia là vòng lặp phần cứng, hút dữ liệu lên từ chính các nhà máy của quốc gia đó. Điều DeepSeek mở ra bằng một báo cáo duy nhất là kết quả được gọi là chi phí thấp, nhưng điều mà kết quả ấy khởi động lại là hai chu trình khác nhau này.

Vẫn còn một câu hỏi bỏ ngỏ. Nếu thế giới chấp nhận mô hình vì các tham số của nó được công bố ẩn danh và xóa đi ranh giới quốc tịch, thì ai quyết định chu trình nào sẽ nhận dòng dữ liệu và công việc phái sinh mà mô hình tạo ra? Các con số trên bảng xếp hạng thay đổi hàng tháng, nhưng hướng quay của hai vòng lặp bên dưới chúng không thay đổi thường xuyên như vậy. Các công ty mô hình mới nổi được đọc qua trụ sở doanh nghiệp và địa chỉ đăng ký

DeepSeek (, DeepSeek) có hồ sơ công khai mỏng hơn so với các công ty niêm yết thuộc loại khác. Công ty đã công bố nhiều báo cáo kỹ thuật và kho mô hình, nhưng không có cấu trúc của một công ty niêm yết, nơi trụ sở và chi tiết đăng ký được trình bày đầy đặc trong các báo cáo định kỳ. Dựa trên thông tin doanh nghiệp công khai, pháp nhân vận hành của DeepSeek được ghi là Hangzhou DeepSeek Artificial Intelligence Basic Technology Research Co., Ltd. (). Địa chỉ đăng ký được nêu là Room 1201, West Building 1, Huijin International Building, No. 169 Huancheng North Road, Gongshu District, Hangzhou, Zhejiang Province (16911201).

Địa chỉ này cho thấy DeepSeek khởi đầu bên trong nguồn lao động internet và công nghệ tài chính của Hangzhou. Việc địa điểm là Hangzhou, chứ không phải Zhongguancun của Beijing hay Nanshan của Shenzhen, là điều đáng kể. Thử nghiệm mô hình chi phí thấp đã xuất hiện từ một thành phố của điện toán đám mây và thương mại điện tử do Alibaba xây dựng, một thành phố nơi vốn công nghệ tư nhân gặp những con người được đào tạo về toán học và kỹ thuật.

Moonshot AI (, Moonshot AI) được công khai gắn với Beijing Moonshot AI Technology Co., Ltd. () với tư cách pháp nhân đối ngoại của công ty. Địa chỉ liên hệ trên website chính thức của công ty chỉ tới tầng 13, Tòa nhà 1, JD Technology Building, số 76 đường Zhichun, quận Haidian, Bắc Kinh (76113). Haidian là nơi Đại học Thanh Hoa (), Đại học Bắc Kinh (), Viện Hàn lâm Khoa học Trung Quốc (), và Zhongguancun () nằm gần nhau.

Vị trí địa lý này cũng nằm phía sau việc Moonshot nhấn mạnh các mô hình ngữ cảnh dài và dịch vụ Kimi. Hình ảnh một startup đặt trong tòa nhà trụ sở của một nền tảng lớn cho thấy các công ty AI mới của Trung Quốc không bắt đầu từ những nhà để xe hoàn toàn độc lập. Họ lớn lên trong khoảng cách vật lý rất ngắn với các nền tảng hiện có, các trường đại học, và mạng lưới đầu tư.

MiniMax (, MiniMax) tự mô tả trong phần giới thiệu công ty chính thức là một công ty mô hình nền tảng được thành lập vào đầu năm 2022. Tài liệu niêm yết và tài liệu nhà đầu tư nêu các văn phòng tại số 1 đường Xitucheng, quận Haidian, Bắc Kinh (1), đường Guiqing, quận Xuhui, Thượng Hải (), và Tháp D, China Resources Building, quận Nanshan, Thâm Quyển (D). Cách bố trí này có thể được hiểu là dấu hiệu cho thấy công ty tách riêng và phối hợp nhân tài nghiên cứu, đội ngũ thương mại hóa, và các đầu mối thị trường vốn.

Bắc Kinh gần hơn với ngôn ngữ của nghiên cứu và chính sách. Thượng Hải gần hơn với cánh cửa thị trường vốn và kinh doanh toàn cầu. Thâm Quyển gần hơn với hệ sinh thái phần cứng và sản phẩm. Không phải ngẫu nhiên mà một công ty phát hành dịch vụ tạo video Hailuo (), cùng với các sản phẩm giọng nói, âm nhạc, và agent, lại hoạt động trên cả ba thành phố này.

Các tài liệu liên quan đến niêm yết công khai của Zhipu (, Z.ai) cũng ghi tên Beijing Zhipu Huazhang Technology Joint Stock Company Limited (, Knowledge Atlas Technology Joint Stock Company Limited). Văn phòng đăng ký và trụ sở chính tại Trung Quốc được ghi là tầng 10, Tòa nhà 9, số 1 Zhongguancun East Road, quận Haidian, Bắc Kinh (1910). Địa chỉ này nằm trong khu Zhongguancun, gần Tsinghua University.

Lý do dòng GLM của Zhipu vừa mang ngôn ngữ học thuật của các phòng thí nghiệm đại học, vừa mang ngôn ngữ API của sản phẩm doanh nghiệp, gắn chặt với địa điểm này. Từ Zhipu () trong tên công ty cũng liên quan đến mối quan tâm ban đầu của họ về việc kết nối đồ thị tri thức với mô hình. Nếu chỉ nhìn công ty qua việc bị đưa vào danh sách trừng phạt của Hoa Kỳ, ta mới thấy một nửa bức tranh. Zhipu là một công ty bị trừng phạt, đồng thời là một công ty nằm ở trung tâm hệ sinh thái đại học và viện nghiên cứu của Bắc Kinh.

Liang Wenfeng

Nếu chỉ nhìn Liang Wenfeng () như nhà sáng lập DeepSeek, điểm xuất phát của ông sẽ bị làm mờ. Ban đầu, ông được biết đến nhiều hơn với tư cách đồng sáng lập quỹ phòng hộ High-Flyer (, High-Flyer). Sân khấu đầu tiên của ông là công việc đưa toán học và điện toán vào thị trường tài chính để nắm bắt những chênh lệch giá rất nhỏ. Kinh nghiệm đó cũng để lại dấu ấn trong thiết kế mô hình của DeepSeek. Nó có nghĩa là tiết kiệm năng lực tính toán thay vì chi tiêu lớn, tìm các điểm nghẽn và thay đổi cấu trúc thay vì tiến lên bằng trực giác con người.

Việc báo cáo kỹ thuật DeepSeek V3 nhấn mạnh chi phí huấn luyện, độ ổn định, xử lý token và kiến trúc mixture-of-experts hơn là quy mô mô hình không tách rời khỏi xuất thân của nhà sáng lập. Trong tài chính, lợi nhuận không đến từ những tuyên bố lớn lao. Nó đến từ việc chạy khối lượng tính toán rẻ thật lớn, ngăn thua lỗ và liên tục xếp chồng những khác biệt nhỏ. Chiến lược mô hình chi phí thấp của DeepSeek cho thấy một bản năng tương tự.

Lý do ông có vẻ xa lạ trong ngành AI Trung Quốc không chỉ là vì ông ít nói. Các nhà sáng lập công ty công nghệ lớn của Trung Quốc thường xây dựng nền tảng, gom người dùng rồi mở rộng kinh doanh bằng cách thêm quảng cáo hoặc thanh toán. Liang Wenfeng đi theo một con đường khác. Ông đến với cảm nhận về nghiên cứu và tài nguyên tính toán trước khi có cảm nhận về dịch vụ tiêu dùng. Từ thời High-Flyer, ông đã gom các bộ xử lý đồ họa hiệu năng cao và xây dựng cụm tính toán của riêng mình, rồi những tài sản tính toán đó sau này nuôi dưỡng nghiên cứu mô hình của DeepSeek.

Khi các biện pháp kiểm soát xuất khẩu của Hoa Kỳ thu hẹp khả năng tiếp cận chip của các công ty Trung Quốc, lượng tài sản tính toán đã tích lũy sẵn và kỹ thuật tiết kiệm chi phí trở thành lá chắn của DeepSeek. Ở điểm này, Liang Wenfeng hơi khác cả các nhà sáng lập startup kiểu Silicon Valley lẫn các nhà sáng lập nền tảng Trung Quốc. Ông gần với kiểu người nhìn vào sổ cái của tính toán trước khi bán một mô hình.

DeepSeek thu hút sự chú ý của thế giới nhờ R1 và V3. Nhưng phía sau chúng là nhiều năm tích lũy. Câu trong báo cáo kỹ thuật nói rằng các lần chạy thất bại không bị quay lui trong quá trình huấn luyện mô hình quy mô lớn có thể nghe khô khan với độc giả ngoài phòng thí nghiệm. Nhưng với những người hiểu huấn luyện mô hình, câu đó có nghĩa là chi phí và thời gian đã được giữ lại trọn vẹn. Nếu một lần huấn luyện đã chạy nhiều ngày rồi sập đổ một lần, hàng trăm nghìn đô la sẽ tan biến vào hư không.

DeepSeek xây dựng cấu trúc của mình để giảm kiểu lãng phí đó, và diễn đạt các thiết kế như mô hình mixture-of-experts và multi-head latent attention bằng ngôn ngữ cắt giảm chi phí. Vai trò của Liang Wenfeng không giống một anh hùng đơn độc tự mình tạo ra một phát minh chói sáng, mà giống một nhạc trưởng buộc các nhà nghiên cứu phải nhìn thẳng vào bức tường chi phí. Ai cũng có thể nói về giấc mơ làm cho mô hình lớn hơn. Hỏi làm thế nào để đi xa hơn với cùng lượng tài nguyên thì kém hào nhoáng hơn, nhưng bền hơn.

Lần sóng mà ông tạo ra được đọc theo những cách khác nhau trong và ngoài Trung Quốc. Ở Trung Quốc, ông được đón nhận như một nhân vật đã tìm ra con đường tiến lên ngay cả dưới các biện pháp kiểm soát xuất khẩu của Mỹ. Bên ngoài Trung Quốc, sự trỗi dậy của một mô hình Trung Quốc chi phí thấp được xem là một sự kiện làm lung lay logic đầu tư đứng sau hạ tầng AI của Mỹ. Cả hai cách đọc đều có cơ sở. Nhưng nếu Liang Wenfeng chỉ được xem như một biểu tượng trong một câu chuyện dân tộc chủ nghĩa, chúng ta sẽ bỏ lỡ kết cấu kỹ thuật mà DeepSeek thật sự đã thay đổi. Lựa chọn của ông gần với việc kéo các nhà phát triển và nhà nghiên cứu vào bằng các mô hình open-weight hơn là kiểm tiền trước bằng các dịch vụ đóng.

Việc phát hành là một tín hiệu chính trị, nhưng cũng là một lựa chọn kỹ thuật nhằm chia sẻ gánh nặng chi phí. Khi các nhà phát triển trên khắp thế giới tải mô hình xuống, thử nghiệm, sửa lỗi và xây dựng các mô hình phái sinh, công ty có thể dựa vào sức mạnh của một hệ sinh thái mà không phải tự mình xây dựng mọi ứng dụng. Điều Liang Wenfeng mở ra không chỉ là cánh

cửa của riêng DeepSeek. Đó là cánh cửa dẫn tới một vòng lặp mô hình mở kiểu Trung Quốc.

Sự thận trọng quanh Liang Wenfeng cũng nên được giữ lại trong bản thảo. Các báo cáo kỹ thuật và hiệu năng mô hình của DeepSeek có thể được kiểm tra, nhưng tài liệu công khai không cho chúng ta thấy chi tiết quá trình ra quyết định nội bộ của công ty, cách phân bổ tài nguyên tính toán, hay khoảng cách kế toán thực tế giữa High-Flyer và DeepSeek. Vì vậy, tốt hơn nên đọc hồ sơ của ông qua hướng đi của những lựa chọn có thể xác nhận, thay vì thổi phồng thành huyền thoại. Những gì có thể xác nhận là một kiến trúc mô hình giúp hạ chi phí, việc phát hành open weights, một tổ chức nghiên cứu đặt tại Hangzhou, và cảm quan tính toán được mang sang từ kỹ thuật tài chính.

Kết hợp lại, bốn yếu tố này đã tạo nên tính cách của DeepSeek. Ý nghĩa của Liang Wenfeng không nằm gần cụm từ "nhà sáng lập thiên tài" bằng việc ông là "người đã biến chi phí tính toán thành một câu hỏi chính trị và công nghiệp". Câu hỏi ông đặt ra vẫn còn đó. Muốn xây dựng một mô hình tốt có thật sự cần lượng vốn gần như vô tận không? Sau DeepSeek, ngành AI toàn cầu khó có thể né tránh câu hỏi ấy.

Yang Zhilin

Yang Zhilin () là gương mặt đại diện trước công chúng của Moonshot AI. Ông học tại Đại học Thanh Hoa và Carnegie Mellon University, thuộc thế hệ từng tích lũy kinh nghiệm nghiên cứu tại Google Brain và Meta AI. Tên ông thường được nhắc cùng các công trình về mô hình ngữ cảnh dài như Transformer-XL và XLNet. Khi biết nền tảng này, ta dễ hiểu hơn vì sao Moonshot lần đầu tạo tên tuổi bằng Kimi, một chatbot ngữ cảnh dài.

Moonshot không tự giới thiệu mình như một công ty muốn trải rộng mọi chức năng trên một bề mặt thật rộng. Công ty đưa ra một mô hình có thể đọc nhiều tài liệu cùng lúc, ghi nhớ chúng trong thời gian dài, và theo kịp mạch làm việc của người dùng. Ngữ cảnh dài không phải là khẩu hiệu tiếp thị. Đó là kết quả của việc con đường nghiên cứu của nhà sáng lập trở thành sản phẩm của công ty. Một vấn đề từng được nghiên cứu trong các bài báo dưới tên độ dài ngữ cảnh, khi đi vào dịch vụ doanh nghiệp, đã trở thành năng lực đọc báo cáo, hợp đồng và kho mã cùng một lúc.

Sự nghiệp của Yang Zhilin cũng cho thấy một lát cắt về cách nhân tài AI luân chuyển ở Trung Quốc. Một nhà nghiên cứu Trung Quốc từng làm việc về nghiên cứu mô hình tuyến đầu tại các viện nghiên cứu của Mỹ đã trở về Trung Quốc, lập công ty, rồi nhanh chóng gọi vốn trong mạng lưới đầu tư và đại học ở Haidian, Bắc Kinh. Dòng chảy này không thể được giải thích chỉ bằng các văn kiện của chính phủ Trung Quốc. Nhà nước đặt ra phương hướng và vốn đi theo, nhưng những người biến cấu trúc đó thành các mô hình thực tế là các nhà nghiên cứu đã hấp thụ ngôn ngữ của các bài báo quốc tế và hệ sinh thái mã nguồn mở.

Yang Zhilin là người đã chuyển ngôn ngữ đó thành một sản phẩm cho thị trường Trung Quốc. Đó cũng là lý do Kimi cố gắng bước vào quy trình làm việc của các nhà phát triển và người làm tri thức. Một mô hình ngữ cảnh dài giống một công cụ có thể giữ lại các tài liệu khô khan trong thời gian dài hơn là một chatbot giải trí. Trong thị trường doanh nghiệp, chính sự khô khan đó lại trở thành sức mạnh.

Moonshot đã tăng trưởng nhanh, và sự tăng trưởng đó cũng mang theo bất ổn. Các công ty mô hình lớn phải đối mặt cùng lúc với chi phí máy chủ, chi phí nhân tài và chi phí dữ liệu. Khi người dùng tăng, chi phí suy luận cũng tăng theo. Khi mô hình lớn hơn, chi phí huấn luyện lại nhảy vọt. Dưới áp lực này, Yang Zhilin thúc đẩy cả các mô hình mở như Kimi K2 lẫn các dịch vụ tiêu dùng. Mô hình mở giúp nuôi một hệ sinh thái nhà phát triển, nhưng không phải là con đường tạo doanh thu ngay lập tức.

Dịch vụ tiêu dùng xây dựng thương hiệu, nhưng đồng thời cũng đẩy chi phí lên. Vì vậy, nhiệm vụ của ông là di chuyển giữa niềm tự hào của một nhà nghiên cứu và sổ sách của một nhà quản lý. Việc Moonshot nằm trong mạng lưới nghiên cứu của Bắc Kinh và trong các tòa nhà của những nền tảng đã có sẵn vừa làm dịu áp lực này, vừa trói công ty lại. Công ty có thể tiến gần hơn đến vốn và nhân tài, nhưng không thể đi quá xa khỏi kỳ vọng của thị trường và chính sách.

Điều quan trọng khi đánh giá Yang Zhilin là ông đã xây dựng một công ty mô hình lớn mà không có một nền tảng khổng lồ đứng sau. Alibaba và Tencent đã có sẵn hạ tầng đám mây, người dùng và mạng lưới thanh toán. Moonshot được biết đến trước hết nhờ chính mô hình của mình và cảm quan sản phẩm, chứ không dựa trên loại nền móng đó. Vì lý do này, thử nghiệm của Moonshot đã trở thành phép thử xem một công ty mô hình nền tảng độc lập có thể trụ được bao lâu ở Trung Quốc. Điểm mạnh của Yang Zhilin là chiều sâu nghiên cứu và sự rõ ràng trong sản phẩm. Điểm yếu của ông là hạ tầng và dòng tiền. Sức căng giữa hai mặt này sẽ quyết định chương tiếp theo của Moonshot. Một nhà nghiên cứu làm việc với ngữ cảnh dài giờ đây phải chịu được khung thời gian dài của kinh doanh. Đó là vị trí của Yang Zhilin hiện nay.

Việc Kimi của Moonshot đặt ngữ cảnh dài lên tuyến đầu cho thấy rõ tính khí của Yang Zhilin. Trong thị trường chatbot đại chúng, những câu trả lời ngắn gọn và dí dỏm dễ thu hút ánh nhìn, nhưng với doanh nghiệp và nhà nghiên cứu, điều quan trọng hơn là khả năng giữ trọn các tài liệu dài, mã nguồn, hợp đồng và bài nghiên cứu cho đến cuối. Yang Zhilin xây dựng công ty của mình đúng trên nhu cầu đó. Lựa chọn này nổi bật ngay cả trong hệ sinh thái AI của Trung Quốc. Trong khi DeepSeek làm rung chuyển bài toán chi phí, Alibaba mở rộng hệ sinh thái đám mây và mô hình mở, còn Tencent nhúng mô hình vào WeChat, Moonshot đưa trí nhớ dài và công cụ cho nhà phát triển lên phía trước.

Rủi ro của Yang Zhilin là sản phẩm càng rõ ràng thì thị trường có thể càng hẹp. Để biến ngữ cảnh dài thành tiền thật, công ty phải đi sâu vào các công việc có giá trị cao như luật, tài chính, nghiên cứu và phát triển, và lập trình. Thị trường đó di chuyển chậm hơn các ứng dụng tiêu

dùng, nhưng một khi một công ty đã bén rễ ở đó, nó sẽ tồn tại lâu dài. Liệu Yang Zhilin có thể biến chiều sâu của một nhà nghiên cứu thành sức bền của một nhà quản lý hay không là phép thử tiếp theo của Moonshot.

Yan Junjie

Yan Junjie () được giới thiệu là người sáng lập MiniMax, đồng thời là chủ tịch, giám đốc điều hành và giám đốc công nghệ của công ty. Việc một người nắm cả chức danh quản lý lẫn kỹ thuật nói lên khá nhiều về tính cách của công ty này. MiniMax đang cùng lúc đẩy mạnh ngôn ngữ, video, giọng nói, âm nhạc và các agent. Một danh mục sản phẩm rộng có thể cho thấy hướng đi bị nhòe, hoặc cũng có thể cho thấy năng lực nhanh chóng đưa các mô hình nền tảng vào nhiều dạng truyền thông khác nhau. Sự nghiệp của Yan Junjie ủng hộ cách hiểu thứ hai. Ông lấy bằng tiến sĩ về trí tuệ nhân tạo tại Institute of Automation, Chinese Academy of Sciences, hoàn thành nghiên cứu sau tiến sĩ tại Tsinghua University, và từng lãnh đạo một tổ chức nghiên cứu tại SenseTime. Ông đã đi qua cả bài báo học thuật, tổ chức nghiên cứu lẫn sản phẩm thương mại.

Nền tảng của ông tại SenseTime giải thích màu sắc của MiniMax. SenseTime lớn lên nhờ nhận diện khuôn mặt và thị giác máy tính, nhưng giữa làn sóng mô hình lớn, công ty này phải chuyển hướng sang trí tuệ không gian và mô hình tạo sinh. Yan Junjie nhiều khả năng đã học được khoảng cách giữa nghiên cứu và sản phẩm ngay ở trung tâm của quá trình chuyển đổi đó. MiniMax không dừng lại ở các mô hình văn bản. Công ty nhanh chóng tung ra các dịch vụ tạo video như Hailuo (), tổng hợp giọng nói, tạo nhạc và sản phẩm agent, một mô hình gắn với nền tảng của một nhà sáng lập có cảm nhận mạnh về thị giác và AI đa phương thức. Phần giới thiệu chính thức của công ty nói đến hơn 200 triệu người dùng và hơn 100.000 khách hàng doanh nghiệp cùng nhà phát triển. Điều đó cũng có nghĩa là MiniMax muốn tự định nghĩa mình không phải như một mô hình trong phòng thí nghiệm nghiên cứu, mà như một bó sản phẩm.

Nhiệm vụ của Yan Junjie là cân bằng tốc độ và niềm tin. Tạo video và tạo giọng nói gây ấn tượng mạnh với người dùng, nhưng khách hàng doanh nghiệp lại phải đối mặt với các câu hỏi về tính nhất quán và quyền lợi. Những mô hình tạo nhạc, làm video và nói giống con người kéo theo cùng lúc các vấn đề về bản quyền, quyền đối với hình ảnh cá nhân, thông tin sai lệch và an toàn. MiniMax càng tung sản phẩm mới nhanh, các câu hỏi này càng lớn nhanh. Vì vậy, vai trò của nhà sáng lập, khi kết hợp với vai trò giám đốc công nghệ, trở nên quan trọng. Ông không chỉ phải cải thiện chất lượng mô hình, mà còn phải quyết định tính năng nào được phát hành trước ở thị trường nào, đặt lớp bảo vệ ở đâu, và sản phẩm nào nên giữ lại ở giai đoạn nghiên cứu. Một cấu trúc trong đó nhà sáng lập có nền tảng nghiên cứu cũng điều hành quản lý giúp quyết định nhanh hơn. Nó cũng mang rủi ro ngược lại: cơ chế kiểm tra yếu hơn và phân công lao động kém rõ hơn.

Cách MiniMax trình bày địa chỉ tại ba thành phố cũng cho thấy thách thức của Yan Junjie. Beijing gần với nhân tài và chính sách, Shanghai gần với vốn và liên kết quốc tế, còn Shenzhen gần với sản phẩm và hệ sinh thái phần cứng. Nếu một công ty muốn hoạt động xuyên qua cả ba nơi, nhà sáng lập của nó phải hiểu ngôn ngữ của giới nghiên cứu, nhà đầu tư và quản lý sản phẩm. Vì Yan Junjie đã di chuyển giữa bài báo nghiên cứu và sản phẩm doanh nghiệp, ông có vẻ có khả năng thực hiện sự chuyển dịch này. Thành công của MiniMax phụ thuộc ít hơn vào một bản demo gây ấn tượng, và nhiều hơn vào năng lực chuyển dịch giữa các thế giới ấy. Làm cho một mô hình tạo ra một video thuyết phục là một chuyện. Xây dựng một hệ thống mà các công ty trả tiền để dùng hằng tháng lại là chuyện khác. Yan Junjie phải thu hẹp khoảng cách giữa hai việc đó.

Một điểm dễ bị bỏ sót trong hồ sơ của Yan Junjie là “tốc độ biến đa phương thức thành sản phẩm” của ông. Không phải nhà nghiên cứu nào công bố nhiều bài báo cũng làm ra sản phẩm tốt. Cũng không phải công ty nào tung sản phẩm nhanh cũng có nền tảng nghiên cứu sâu. MiniMax nhắm vào khoảng ở giữa hai bên đó. Tạo video Hailuo, tổng hợp giọng nói, tạo nhạc và các công cụ agent có thể trông như những thị trường khác nhau, nhưng tất cả đều là công cụ để máy móc tạo ra hoặc hỗ trợ cách con người biểu đạt. Yan Junjie đang cố gom những công cụ biểu đạt này vào trong một công ty mô hình nền tảng.

Con đường này hấp dẫn, nhưng nhiều rủi ro. Chất lượng của một video đơn lẻ có thể khiến người dùng kinh ngạc, nhưng khách hàng doanh nghiệp đòi hỏi chất lượng lặp lại được, xử lý quyền sử dụng, bảo mật và kiểm soát chi phí. Vì vậy Yan Junjie phải liên tục đi qua đi lại giữa phòng thí nghiệm nghiên cứu và văn phòng hoạch định sản phẩm. Số phận của MiniMax phụ thuộc ít hơn vào việc các mô hình của họ có thể làm gì, mà phụ thuộc nhiều hơn vào việc công ty có thể biến những năng lực đó thành một gói sản phẩm ổn định hay không.

Câu hỏi còn lại đối với nhân vật này là tổ chức đứng sau tốc độ đó. Một công ty xử lý nhiều loại truyền thông cùng lúc rất dễ chứng kiến ngôn ngữ của các nhóm nghiên cứu, sản phẩm và bán hàng tách rời nhau. Vai trò kết hợp của Yan Junjie, vừa là giám đốc điều hành vừa là giám đốc công nghệ, là một cơ chế để giảm sự tách rời đó. Phán đoán của một người trở nên nhanh hơn, nhưng công ty luôn bị thử thách ở chỗ liệu tranh luận nội bộ có đủ hay không.

Zhang Peng

Zhang Peng () xuất hiện trong các tài liệu công khai với tư cách đồng sáng lập và CEO của Zhipu (, Z.ai). Gốc rễ của Zhipu đi ngược về các nghiên cứu kỹ thuật tri thức tại Tsinghua University. Tên tiếng Trung Zhipu Huazhang (), nằm trong tên công ty, mang ý tưởng về nên một bản đồ tri thức, và mối quan tâm ban đầu của công ty là nối đồ thị tri thức với các mô hình tiền huấn luyện quy mô lớn. Vai trò của Zhang Peng là biến nền tảng học thuật ấy thành một tổ chức kinh doanh. Các bài báo từ một phòng thí nghiệm của Tsinghua University và các bản kế hoạch kinh doanh từ Zhongguancun ở Beijing nói bằng hai ngôn ngữ khác nhau. Đứng

giữa hai thế giới đó, ông biến dòng mô hình GLM thành sản phẩm và mở đường đến APIs, agents, và các mô hình lập trình.

Zhipu chịu sức ép nặng nề khi bị đưa vào danh sách trừng phạt của Hoa Kỳ. Sự việc đó thử thách năng lực quản trị của Zhang Peng. Với một công ty mô hình lớn, trừng phạt không chỉ chặn việc mua chip. Nó còn làm lung lay quyền tiếp cận các dịch vụ đám mây ở nước ngoài, tâm lý nhà đầu tư, và đánh giá rủi ro của khách hàng. Trong hoàn cảnh ấy, Zhang Peng làm rõ hơn hướng đi của công ty, nghiêng về các mô hình mở và nhu cầu công nghiệp trong Trung Quốc. Dòng GLM tìm cách tự xác lập như một nền tảng đem lại cho nhà phát triển và khách hàng doanh nghiệp các chức năng tiếng Trung, lập trình, và agent, trong khi nghiên cứu tiếp tục tập trung vào việc cải thiện xử lý ngữ cảnh dài bằng cách tiếp nhận thiết kế mở của DeepSeek.

Điều hiện ra ở đây là lựa chọn ưu tiên công dụng thực tế hơn thể diện. Học từ thiết kế mở của một đối thủ rồi đưa nó vào mô hình của mình là việc khó làm nếu đặt lòng tự ái lên trước. Dưới sự dẫn dắt của Zhang Peng, Zhipu chọn hấp thụ nhanh tri thức mở để tìm đường đi vòng qua bức tường trừng phạt.

Sức mạnh của ông nằm ở khả năng giữ cho hệ sinh thái học thuật và hệ sinh thái kinh doanh cùng vận hành. Zhipu đứng tại nơi mạng lưới các nhà nghiên cứu từ Tsinghua University, môi trường chính sách của Beijing, và nhu cầu của khách hàng doanh nghiệp gặp nhau. Zhang Peng phải điều chỉnh ngôn ngữ của công ty để ba lực này không va chạm. Với giới nghiên cứu, hiệu năng mô hình và đóng góp cho bài báo là điều quan trọng. Với khách hàng, APIs ổn định và bảo mật là điều quan trọng. Trong môi trường chính sách, an toàn và kiểm soát là điều quan trọng. Đưa ba yêu cầu ấy vào một sản phẩm mới là công việc quản trị thật sự của một công ty mô hình lớn Trung Quốc. Nếu GLM của Zhipu muốn trở thành không chỉ một cái tên kỹ thuật mà còn là sản phẩm bán được cho khách hàng doanh nghiệp, Zhang Peng phải đưa nhịp độ nghiên cứu khớp với nhịp quản trị.

Zhang Peng cho thấy rõ cách AI Trung Quốc đi từ đại học vào doanh nghiệp. Bức tranh này khác với kiểu Mỹ, nơi một nhà sáng lập bước ra từ Big Tech cùng một nền tảng khổng lồ. Ở đây, công nghệ mô hình ngôn ngữ và đồ thị tri thức được xây trong phòng thí nghiệm đại học, đi qua một địa chỉ công ty ở Zhongguancun và các hồ sơ cho thị trường vốn Hong Kong, rồi trở thành sản phẩm doanh nghiệp. Zhang Peng là người vận hành quá trình ấy. Ông không hẳn là nhà sáng lập ngôi sao của một dịch vụ tiêu dùng hào nhoáng, mà giống một nhà quản lý đưa mô hình phòng thí nghiệm thành hệ thống mà khách hàng sẵn sàng trả tiền. Phép thử tiếp theo của Zhipu nằm ở sức bền hơn là công nghệ. Liệu công ty có thể tiếp tục biến dòng GLM thành các sản phẩm sống được giữa lệnh trừng phạt, thị trường vốn và cạnh tranh nguồn mở hay không? Câu hỏi ấy đi cùng tên Zhang Peng.

Một câu hỏi khác gắn với Zhang Peng là liệu Zhipu có thể vượt ra khỏi giá trị của cái tên Tsinghua University hay không. Uy tín của một phòng thí nghiệm đại học tạo ra niềm tin ban

đầu. Nhưng theo thời gian, khách hàng nhìn vào bài báo ít hơn, và nhìn nhiều hơn vào khả năng ứng phó sự cố, giá cả, bảo vệ dữ liệu và tốc độ cập nhật mô hình. Zhang Peng phải chấp nhận thứ ngôn ngữ thị trường lạnh lùng này. Việc Zhipu gõ cửa thị trường vốn Hong Kong và tiếp cận giới phát triển toàn cầu dưới cái tên ngắn hơn Z.ai là một phần của thay đổi đó.

Trong khi tên công ty vẫn mang sắc thái học thuật, tên sản phẩm lại thay đổi theo các chu kỳ ngắn hơn và nhanh hơn. Quản lý cùng lúc hai gương mặt ấy là công việc của Zhang Peng. Ông phải phát hành các mô hình và agent mà khách hàng có thể dùng ngay, nhưng không đánh mất chiều sâu của hệ sinh thái nghiên cứu Tsinghua University. Lệnh trừng phạt, cạnh tranh và ánh mắt của thị trường vốn đều đang đặt lên ông. Tương lai của Zhipu phụ thuộc vào việc Zhang Peng có thể biến niềm tin học thuật thành doanh thu kinh doanh lặp lại hay không.

Sức mạnh ông cần là quan hệ khách hàng bền lâu. Các công ty mô hình nhận được tiếng vỗ tay trong ngày ra mắt, nhưng khách hàng doanh nghiệp tìm người nhắc máy khi hệ thống gặp lỗi. Nếu Zhang Peng muốn biến Zhipu thành một công ty AI doanh nghiệp thật sự, ông phải làm cho hệ thống hỗ trợ, giá cả và cam kết bảo mật của công ty vững hơn các bằng hiệu năng của GLM. Khoảnh khắc niềm tin từ phòng thí nghiệm trở thành niềm tin vào vận hành chính là nơi cuộc đấu của ông sẽ được định đoạt.

Chương 6. Alibaba, Tencent, Baidu, ByteDance và iFLYTEK trong cuộc đua toàn bộ chuỗi công nghệ

Mở bản trình bày kết quả của Alibaba Group cho năm tài chính 2026, trang đầu tiên không bắt đầu bằng thương mại điện tử. Nó bắt đầu bằng điện toán đám mây và trí tuệ nhân tạo. Tiêu đề trong thông cáo của công ty viết như sau: doanh thu của Alibaba Cloud đã tăng nhanh lên 40 phần trăm, và động lực phía sau là cách chiến lược trí tuệ nhân tạo của họ đang chuyển thành kết quả. Một câu đã chứa hướng đi của công ty. Trí tuệ nhân tạo không còn là một bài tập trong phòng thí nghiệm. Nó là một dòng mục trên bảng doanh thu. Việc tài liệu lợi nhuận của một công ty nói như vậy cho thấy công ty muốn tự gọi mình là gì. Họ muốn được biết đến không phải như một công ty thương mại điện tử, mà như một công ty hạ tầng.

Hãy đi theo các con số. Trong quý kết thúc vào tháng 3 năm 2026, doanh thu đám mây đạt mức khoảng 41,6 tỷ nhân dân tệ, và doanh thu từ các sản phẩm liên quan đến trí tuệ nhân tạo tiếp tục tăng ba chữ số so với cùng kỳ năm trước trong quý thứ mười một liên tiếp. Tỷ trọng của trí tuệ nhân tạo trong doanh thu đám mây bên ngoài tăng lên khoảng 30 phần trăm. Tổng giám đốc Eddie Wu nói rằng khoản đầu tư trí tuệ nhân tạo toàn tầng của công ty đã vượt qua giai đoạn ươm tạo và bước vào thương mại hóa quy mô lớn, và trên cơ sở đó công ty đã phê duyệt cổ tức tiền mặt 2,5 tỷ đô la.

Cổ tức là một tín hiệu cho thấy, theo đánh giá của chính công ty, mảng kinh doanh trí tuệ nhân tạo không còn là chiếc hũ không đáy. Điều đó có nghĩa là mảng kinh doanh này đã đi từ giai đoạn phải rót tiền vào sang giai đoạn có thể phân phối tiền ra. Cụm từ mười một quý cũng đáng để dừng lại suy nghĩ. Nó nói rằng doanh thu đã tăng theo một hướng trong gần ba năm, và nó mang theo tuyên bố của công ty rằng điều này nên được đọc như một xu hướng, không phải như một tia lóe trong một quý.

Chương này đặt năm công ty lớn của Trung Quốc cạnh nhau: Alibaba, Tencent, Baidu, ByteDance và iFlytek. Nếu DeepSeek và những công ty được gọi là AI Tigers đã bàn ở chương trước là những con thuyền nhỏ, nhanh, rẽ dòng nước trước tiên, thì năm công ty này là những đội tàu lớn sở hữu cảng, tàu và tuyến đường. Họ đang cố xây toàn bộ tầng công nghệ bên trong một công ty, từ nền tảng chip và trung tâm dữ liệu, đến bộ não của các mô hình nền tảng, rồi đến đầu ngón tay của những dịch vụ mà người tiêu dùng và doanh nghiệp thật sự dùng. Cấu trúc nhiều lớp đó được gọi là toàn tầng. Năm công ty dùng cùng một thuật ngữ, nhưng mỗi công ty xây nó theo cách riêng của mình.

Đọc ra những khác biệt ấy là nhiệm vụ của chương này. Chúng ta cũng cần nhớ rằng ranh giới của sân chơi nơi năm công ty này cạnh tranh ban đầu không phải do doanh nghiệp vẽ ra, mà do các văn kiện của chính phủ xác lập. Những mô hình nào có thể được đưa ra thành dịch vụ công, dữ liệu phải được xử lý ra sao, và những ngành nào nên được tiếp nhận trí tuệ nhân

tạo trước, đều là những vấn đề đã được đề cập trong các văn kiện của Quốc vụ viện và các bộ ngành liên quan. Chiến lược của năm công ty là những bức tranh được tô bằng các màu khác nhau trên nền các văn kiện ấy.

Trước hết là Alibaba. Sức mạnh của Alibaba nằm ở độ dày của nền tảng. Theo một báo cáo của Samjong KPMG so sánh năng lực cạnh tranh trí tuệ nhân tạo của Mỹ và Trung Quốc, Alibaba (, Alibaba) công bố vào tháng 2 năm 2025 rằng họ sẽ đầu tư ít nhất 380 tỷ nhân dân tệ vào hạ tầng điện toán đám mây và trí tuệ nhân tạo trong ba năm tiếp theo. Quy mô này vượt các kế hoạch đầu tư của những công ty Big Tech Trung Quốc khác vào cùng thời điểm. Khoản tiền đó không chỉ chảy vào các trung tâm dữ liệu trong nước, mà còn vào các trung tâm dữ liệu mới ở Đông Nam Á, trong đó có Malaysia và Philippines.

Điện toán đám mây không phải là sản phẩm chỉ bán trong biên giới quốc gia. Nó phải bám rễ ra bên ngoài biên giới, vì vậy Alibaba đã chọn trồng trước các căn cứ tính toán ở nước ngoài. Đằng sau lựa chọn ấy có một phép tính. Khi các biện pháp kiểm soát xuất khẩu bán dẫn của Mỹ làm rung chuyển nguồn cung trong nước, công ty đang cố gắng trải rộng hạ tầng vật lý càng xa càng tốt, để tự chữa chỗ chuyển sang một tuyến khác nếu một con đường bị chặn. Hạ tầng không thể xây trong một đêm, nên bên xây trước sẽ mua được thời gian.

Trong câu chuyện về nền móng này có một điểm cần thận trọng. Công ty nói về chip của chính mình. Họ nói rằng các chip do công ty con thiết kế bán dẫn của Alibaba chế tạo đã được triển khai trên quy mô lớn trong đám mây công cộng của họ, và những chip này hỗ trợ các hãng xe cùng các công ty xe tự lái. Điều này phải được nêu rõ là lời kể của chính công ty. Tài liệu hiện có không đủ để khẳng định rằng chip Trung Quốc ngang với các chip hàng đầu của Mỹ về hiệu năng. Những sự kiện đã được xác nhận chỉ đi đến mức này: công ty đã nhắc tới việc triển khai chip của chính mình trong một buổi công bố kết quả kinh doanh chính thức, và họ đang tiến tới tỷ trọng tính toán Trung Quốc cao hơn dưới các biện pháp kiểm soát xuất khẩu của Mỹ.

Hiệu năng tốt hơn hay kém hơn là một câu hỏi tách biệt với hướng đi đó, và lúc này nói thật hơn thì nên để ngỏ. Việc công ty nhắc đến chip trong buổi trình bày kết quả kinh doanh tự nó đã có ý nghĩa. Thông điệp gửi tới nhà đầu tư rất rõ. Alibaba muốn cho thấy mình không bị buộc vào dây xích của chip Mỹ, và ngay cả nền móng hạ tầng của mình cũng nằm trong tay mình. Cử chỉ đó được hiệu năng thực tế chống lưng đến đâu thì chỉ có thể phán đoán khi có dữ liệu riêng xuất hiện.

Bộ não đặt trên nền móng ấy là mô hình Qwen. Các tài liệu so sánh Mỹ-Trung và phân tích bên ngoài cho thấy Qwen đã vượt ra ngoài phạm vi dùng nội bộ trong công ty, trở thành một mô hình mà các nhà phát triển trên khắp thế giới tải xuống, sửa đổi và xây dựng tiếp. Alibaba chọn mở mô hình thay vì khóa chặt nó. Khi một mô hình được phát hành, các nhà phát triển trên toàn thế giới tạo ra các mô hình phái sinh trên nền mô hình đó cho mục đích riêng của họ. Khi số lượng mô hình phái sinh tăng lên, mô hình gốc tiến gần hơn tới vị trí tiêu chuẩn trên thực

tế. Đây được gọi là vòng lặp mã nguồn mở. Ở nơi công ty không thu tiền trực tiếp, nó xây một tài sản khác: ảnh hưởng.

Dù vậy, các con số chi tiết như số lượng chính xác của mô hình phái sinh hay thứ hạng tải xuống thay đổi tùy nguồn và tiêu chuẩn, nên văn bản này chỉ xem hướng chung, rằng Qwen thuộc nhóm các mô hình được tải xuống rộng rãi trên thế giới, là sự thật và không đi xa hơn mức đó. Cần nhìn rõ vì sao chiến lược mở có lợi cho công ty. Khi các nhà phát triển xây công cụ trên Qwen, nơi các công cụ ấy chạy rất dễ trở thành Alibaba Cloud. Cấu trúc là thế này: công ty mở mô hình miễn phí, rồi kéo nhu cầu tính toán do mô hình đó tạo ra vào đám mây của chính mình. Vì vậy bàn tay mở ra và bàn tay bán hàng không hành động tách rời nhau.

Chiến lược mở nối với chiến lược bán hàng. Alibaba mở ví doanh nghiệp bằng cách bán mô hình dưới dạng dịch vụ, trên thực tế là bán một mô hình như thể nó là một dịch vụ. Trong buổi trình bày kết quả kinh doanh, công ty nói doanh thu định kỳ hằng năm từ mô hình và dịch vụ ứng dụng đang tăng nhanh, và đặt mục tiêu nâng mạnh doanh thu gộp từ điện toán đám mây và trí tuệ nhân tạo trong vài năm tới. Bức tranh của Alibaba khép lại ở đây. Với 380 tỷ nhân dân tệ, họ đặt nền móng cho trung tâm dữ liệu và chip, đặt bộ não Qwen lên trên, mở bộ não ấy ra thế giới trong khi nhắm tới vị trí tiêu chuẩn, rồi bán chính bộ não đó cho doanh nghiệp để biến nó thành doanh thu.

Bàn tay mở ra và bàn tay bán hàng cùng gắn với một cơ thể. Để cấu trúc này vận hành được, ba tầng phải nuôi lẫn nhau. Hạ tầng đỡ cho việc huấn luyện mô hình, mô hình tạo ra giá trị cho dịch vụ, và số tiền thu được từ các dịch vụ ấy lại chảy ngược về hạ tầng. Nếu một tầng rung lắc, hai tầng còn lại cũng rung lắc theo.

Dĩ nhiên, bức tranh này cũng có phần tối. Theo báo cáo hằng tuần về Trung Quốc của Hana Securities, vào đầu tháng 3 năm 2026, các nhà phát triển chủ chốt lần lượt rời khỏi bộ phận mô hình ngôn ngữ lớn của Alibaba, và giá cổ phiếu đã giảm vào ngày tin tức được công bố. Dù vốn và hạ tầng sâu đến đâu, khi những người thực sự xây mô hình rời đi, khoảng cách giữa kế hoạch và thực thi sẽ mở ra. Sự việc này cho thấy cuộc chiến giành nhân tài là một mắt xích yếu trong toàn bộ khu vực Big Tech của Trung Quốc, và Alibaba cũng không phải ngoại lệ.

Hạ tầng có thể mua bằng tiền, mô hình có thể lớn lên nhờ dữ liệu, nhưng những người nối hai thứ đó với nhau thì khó giữ lại chỉ bằng tiền hoặc dữ liệu. Việc một mắt xích mỏng nằm ẩn bên trong một cấu trúc dày chính là lý do đầu tiên khiến câu chuyện về năm công ty này không nên được đọc như một câu chuyện chiến thắng mà thôi.

Công ty thứ hai là Tencent (, Tencent). Mở báo cáo thường niên năm 2025 của Tencent, ta thấy một đoạn trong đó chủ tịch Ma Huateng viết rằng trí tuệ nhân tạo đã cải thiện độ chính xác của quảng cáo, làm sâu hơn cảm giác nhập vai trong trò chơi, và thúc đẩy tăng trưởng doanh thu ở dịch vụ đám mây. Tổng doanh thu của Tencent trong năm 2025 là 751,8 tỷ nhân dân tệ, tăng 14 phần trăm so với năm trước. Đằng sau con số đó là chiến lược riêng của Tencent. Thay

vì phô diễn một mô hình mới trước thế giới, Tencent chọn đẩy mô hình vào chính các dịch vụ của mình, vốn đã được hàng trăm triệu người sử dụng.

Bộ não ấy mang tên Hunyuan (, Hunyuan). Việc chủ tịch giải thích trí tuệ nhân tạo trong báo cáo kết quả kinh doanh thông qua hiệu quả của các mảng hiện có như quảng cáo, trò chơi và đám mây đã cho thấy tính cách của công ty này. Với Tencent, trước khi là một sản phẩm mới để bán, trí tuệ nhân tạo là công cụ giúp bán tốt hơn những sản phẩm mà công ty vốn đã bán.

Khoản đầu tư của Tencent cũng không nhỏ. Công ty đã rót những khoản tiền lớn vào hạ tầng trí tuệ nhân tạo, đồng thời duy trì chi tiêu vốn ở mức hai chữ số thấp tính theo tỷ lệ phần trăm doanh thu trong khi mua máy chủ và thiết bị tính toán. Công ty cũng dành riêng chi phí cho các sản phẩm trí tuệ nhân tạo mới, dùng cho hoạt động tiếp thị Yuanbao, dịch vụ chatbot của mình, tuyển dụng nhân tài và bảo đảm dữ liệu cho huấn luyện mô hình. Tuyên bố của Tencent rằng họ sẽ giảm mua lại cổ phiếu trong năm 2026 và chuyển số tiền đó sang đầu tư trí tuệ nhân tạo được hiểu như một tín hiệu rằng công ty xem cuộc đấu này là một chiến dịch dài kéo dài nhiều năm. Mua lại cổ phiếu là trả tiền cho cổ đông, còn đầu tư trí tuệ nhân tạo là đặt tiền vào tương lai. Khi cán cân nghiêng về phía tương lai, điều đó có nghĩa là công ty đặt năng lực cạnh tranh trong vài năm tới lên trước việc bảo vệ giá cổ phiếu hôm nay.

Tencent khác Alibaba ở một điểm căn bản: công ty không dựng lên một nền tảng mới. Tencent đã có một nền tảng khổng lồ trong WeChat. Công ty đã cắm Hunyuan vào khắp hệ sinh thái WeChat, từ mạng xã hội, tìm kiếm, nội dung và gợi ý sản phẩm cho đến dịch vụ khách hàng. Bên trong WeChat, Tencent đặt các dịch vụ tác nhân có thể tự xử lý nhiệm vụ và kết nối người dùng với vô số mini program. Các tác nhân này càng đưa người dùng vào mini program nhiều hơn, năng lượng càng chảy mạnh hơn trong hệ sinh thái WeChat, và năng lượng ấy lại chuyển thành doanh thu. Trong trò chơi, Tencent đã dùng AI tạo sinh để tăng tốc sản xuất nội dung, và bằng công nghệ đa phương thức, công ty đã xây dựng đồng đội AI cho các trò chơi cạnh tranh cùng các nhân vật không phải người chơi có cảm giác như thật trong những bối cảnh không cạnh tranh.

Trong quảng cáo, Tencent bổ sung các giải pháp tự động xử lý nhằm mục tiêu, đặt giá thầu và phân phối quảng cáo, giúp doanh thu quảng cáo vượt mức trung bình của ngành. Trong fintech, công ty tinh chỉnh chấm điểm tín dụng và phát hiện gian lận bằng các mô hình nhẹ, qua đó hạ tỷ lệ nợ xấu. AI của Tencent xuất hiện trước hết như hiệu suất tốt hơn trong các sản phẩm hiện có, rồi sau đó mới được bán như một sản phẩm mới. AI thấm vào báo cáo lợi nhuận bằng cách làm cho quảng cáo chính xác hơn, trò chơi cuốn hút hơn và các khoản vay ít rắc rối hơn.

Trong thị trường doanh nghiệp, Tencent cũng đang chuyển hướng sang các tác tử. Tencent Cloud vượt điểm hòa vốn vào năm 2024, và đến năm 2025 đã tạo ra lợi nhuận hoạt động sau điều chỉnh, nhờ nhu cầu AI từ các công ty. Ngoài những công cụ tạo biên bản cuộc họp hoặc các tính năng tóm tắt trao đổi dịch vụ khách hàng, Tencent đã phát hành các sản phẩm tác tử tự

chủ có thể tự thực hiện quy trình làm việc. Các tác tử này chạy trong sandbox bảo mật, và được thiết kế để người dùng có thể điều khiển nhiều tác tử từ một cửa sổ trò chuyện, đồng thời gọi các chức năng từ mạng lưới thương nhân và nhà phát triển rộng lớn của Tencent. Cụm từ sandbox bảo mật đáng được chú ý ở đây. Một tác tử tự xử lý công việc vừa tiện lợi vừa có rủi ro, vì vậy công ty cố gắng ngăn tai nạn bằng cách cho tác tử chạy trong các không gian tách biệt.

Xu hướng trong các văn bản quản trị của chính phủ về yêu cầu an toàn đối với tác tử AI không xa với xu hướng các công ty nhốt tác tử trong sandbox. Bản thân mô hình Hunyuan, đóng vai trò như bộ não, cũng đã có chỗ đứng trong các lĩnh vực đa phương thức như 3D, chuyển văn bản thành hình ảnh, và mô hình thể giới, dựa trên dữ liệu độc quyền của công ty. Tencent cho biết họ dự định thử nghiệm nội bộ các mô hình được cải tiến, rồi phát hành theo từng giai đoạn. Bức tranh của Tencent hiện ra như sau: họ biến AI thành tiền trên nền tảng lớn mà họ đã sở hữu, thay vì mua một nền tảng mới.

Công ty thứ ba là Baidu (, Baidu). Baidu có tính chất khác với hai công ty đầu tiên. Nếu Alibaba đặt hạ tầng lên trước và Tencent đặt ứng dụng lên trước, Baidu đặt trọng tâm vào công nghệ nền tảng nhằm xây lại cấu trúc của chính mô hình. Những tài liệu thể hiện rõ hướng đi này là bài đăng blog chính thức và báo cáo kỹ thuật về ERNIE 5.0, được công bố vào tháng 2 năm 2026.

Theo báo cáo kỹ thuật, ERNIE 5.0 là mô hình có 2,4 nghìn tỷ tham số, được thiết kế để hiểu và tạo văn bản, hình ảnh, video, và âm thanh trong một kiến trúc duy nhất. Lựa chọn của Baidu nằm ở việc rời bỏ phương pháp hiện có. Cho đến lúc đó, ngành này thường dùng late fusion, tức gắn thêm các thành phần xử lý hình ảnh vào một khung được xây dựng cho văn bản. Baidu bỏ phương pháp gắn thêm đó và chọn một kiến trúc dự đoán thống nhất, đặt mọi loại thông tin trong cùng một không gian token và huấn luyện lại từ đầu.

Cách giải thích là văn bản được xử lý nhanh bằng cách nhìn trước qua nhiều token cùng lúc, hình ảnh được xem như video một khung hình, còn âm thanh dự đoán phần tiếp theo ở cấp codec. Mục tiêu không phải là một sự ghép nối bề mặt được dán từ bên ngoài, mà là một sự hợp nhất ăn khớp với nhau ở tầng ý nghĩa. Ngôn ngữ kỹ thuật khá khó, nhưng mục tiêu thì rõ. Baidu không muốn tiếp nhận văn bản, hình ảnh và âm thanh qua những cửa sổ riêng. Họ muốn tiếp nhận và xử lý chúng qua một cửa sổ duy nhất, bằng cùng một ngôn ngữ.

2. Quy mô bốn nghìn tỷ đồng nghĩa với chi phí tính toán khổng lồ. Baidu cố vượt qua bức tường này bằng hai cách. Một là kiến trúc siêu thưa. Thay vì tách sẵn chuyên gia thị giác và chuyên gia văn bản từ đầu, họ xây một nhóm chuyên gia dùng chung, chỉ gọi những chuyên gia cần thiết ở từng thời điểm tùy theo tính chất của token. Ngay cả ở quy mô 2.4 nghìn tỷ tham số, số tham số thực sự được kích hoạt để xử lý một token vẫn được giữ dưới 3 phần trăm tổng số. Nó giống như xây một nhà kho lớn nhưng

mỗi lần chỉ bật đèn ở những kệ cần dùng. Cách còn lại là huấn luyện đàn hời. Một khi mô hình đã được huấn luyện, các mô hình cấp thấp hơn có thể được rút ra ngay tại chỗ cho những điều kiện phần cứng khác nhau, chẳng hạn giới hạn bộ nhớ hoặc độ trễ, nhờ đó bỏ được chi phí huấn luyện lại.

Cả hai phương pháp đều hướng tới cùng một mục tiêu. Baidu muốn giữ sức mạnh của một mô hình lớn trong khi kéo chi phí sử dụng xuống mức của một mô hình nhỏ. Điều này cũng đáng chú ý vì nó cho thấy vì sao vấn đề này là chuyện sống còn đối với Baidu. Dưới các biện pháp kiểm soát xuất khẩu của Hoa Kỳ, tài nguyên tính toán vừa đắt vừa khó kiếm. Công ty nào có thể tạo ra cùng mức hiệu năng với ít tính toán hơn sẽ trụ được lâu hơn trong các giới hạn đó. Đối mới kiến trúc của Baidu trước hết là một biện pháp sinh tồn, rồi mới là điều để tự hào.

Baidu tiếp tục đường hướng này với ERNIE 5.1 vào năm 2026. Theo thông báo chính thức, 5.1 kế thừa nền tảng tiền huấn luyện của 5.0, đồng thời giảm mạnh tổng số tham số và nén số tham số hoạt động xuống khoảng một nửa. Công ty nói điều này cho phép họ đạt hiệu năng cao với chi phí thấp so với chi phí tiền huấn luyện của các mô hình cạnh tranh. ERNIE 5.1 đặt trọng tâm vào sáng tác văn bản phù hợp với cảm hứng, cảm xúc và cách biểu đạt, cùng với sản xuất nội dung đọc được ý định sâu bên dưới của người dùng. Kết quả benchmark của nó cần được nhìn thận trọng. Báo cáo kỹ thuật đưa ra điểm số cao trong một số bài kiểm tra, nhưng các thước đo để đánh giá diffusion và hiệu năng vẫn chưa ổn định thành chuẩn chung của ngành.

Vì vậy, văn bản này chỉ xác nhận đến mức này: Baidu đã đặt cược vào đối mới cấu trúc trong công nghệ lõi và giảm chi phí. Văn bản sẽ không đưa ra những nhận định xếp hạng, chẳng hạn một mô hình đã vượt lên trước mô hình khác. Thận trọng với xếp hạng không phải là rụt rè. Đó là sự trung thực. Khi cây thước rung, giá trị đo được cũng rung theo. Về nhân tài, Baidu cũng đang tuyển thêm nhiều nhà nghiên cứu ở nước ngoài, mở rộng quy mô các chương trình tuyển dụng hằng năm, và cạnh tranh với các hãng công nghệ lớn toàn cầu để giành người. Xây lại cấu trúc của công nghệ lõi, rốt cuộc, là công việc do con người làm. Baidu, giống Alibaba, đối mặt với cùng một mắt xích mỏng manh mang tên nhân tài.

Công ty thứ tư là ByteDance (, ByteDance). Công ty này, vốn xây dựng một hệ sinh thái video toàn cầu thông qua TikTok, nay đứng ở vị trí có thể đưa dữ liệu từ hàng trăm triệu người tiêu dùng quay trở lại hệ thống, dẫn đầu là video và các dịch vụ xã hội vượt ra ngoài văn bản. Các báo cáo so sánh Mỹ-Trung cho thấy ByteDance đã lên kế hoạch chi hơn 150 tỷ nhân dân tệ cho vốn đầu tư trong năm 2025, trong đó một phần lớn dành cho các trung tâm dữ liệu ở nước ngoài và việc mua chip trí tuệ nhân tạo. Cách giải thích là ByteDance đang tăng mức phụ thuộc vào chip do Trung Quốc sản xuất và tích trữ tài nguyên tính toán để giảm bất định do các biện pháp kiểm soát xuất khẩu của Mỹ đối với bán dẫn tiên tiến tạo ra.

Dòng tiền này chảy vào hiệu năng của mô hình nền tảng riêng của ByteDance, Doubao. Doubao đã được nhúng vào các sản phẩm tiêu dùng được hàng trăm triệu người sử dụng, hỗ trợ hệ sinh thái xã hội và nội dung của công ty. Đây là nơi nguồn sức mạnh của ByteDance hiện ra rõ. Công ty vận hành một số dịch vụ tiêu dùng có lượng sử dụng lớn nhất thế giới, và dữ liệu video cùng dữ liệu xã hội từ các dịch vụ ấy nuôi các mô hình của họ. Dữ liệu làm mô hình lớn lên, và mô hình mạnh hơn lại kéo người dùng quay về dịch vụ. Vòng phản hồi đó chính là vốn liếng của công ty.

Tổ chức nghiên cứu của ByteDance là nhóm Seed. Tháng 2 năm 2026, nhóm này chính thức phát hành Seed 2.0, danh mục thể hệ tiếp theo của họ. Seed không dừng lại ở một chatbot đơn lẻ. Nhóm chia nghiên cứu nền tảng ra nhiều mảng: mô hình ngôn ngữ lớn và hạ tầng, thị giác, giọng nói, tương tác đa phương thức và mô hình thể giới, AI cho khoa học, robot, và AI có trách nhiệm. Nhóm nhìn nhận AI là đã vượt qua giai đoạn công cụ để bước vào thời đại của các tác tử tự xử lý nhiệm vụ. Nhóm đặt ra các mục tiêu trong đó AI trực tiếp tham gia nghiên cứu khoa học, viết phần mềm, và tạo ra giá trị trong nền kinh tế thực.

Việc sơ đồ tổ chức được chia nhỏ đến mức này cho thấy tham vọng của công ty. Đặt ngôn ngữ, âm thanh, video, robot và khoa học dưới cùng một mái nhà có thể được hiểu như một tín hiệu rằng ByteDance không định giữ AI chỉ làm bạn đối thoại bên trong màn hình. Công ty muốn đưa AI vươn ra thế giới vật lý.

Nỗ lực của ByteDance nhằm đạt kết quả đẳng cấp thế giới trong tạo video có thể được xem là dấu hiệu cho thấy hướng đi của công ty. Nhưng những mô tả về một mô hình tạo video cụ thể đang được lan truyền trên thị trường như một "vị vua mới" lại đến từ siêu dữ liệu video không có bản chép lời, nên cuốn sách này không xem chúng là bằng chứng sự thật. Điều có thể xác nhận là việc Seed 2.0 được phát hành chính thức và cấu trúc tổ chức đứng sau nó. Những tên bộ phận như robot học hay mô hình thể giới không lập tức có nghĩa là sản phẩm đã hoàn thiện. Vẫn còn khoảng cách giữa mục tiêu của một tổ chức nghiên cứu và kết quả trên thị trường. Lập trường của cuốn sách này là không nhảy qua khoảng cách đó để tuyên bố một kết quả trước khi bằng chứng đủ sức chống đỡ.

Công ty thứ năm là iFlytek. Nếu bốn công ty đầu hoạt động trên những sân chơi rộng của điện toán đám mây, video và tìm kiếm, iFlytek đào sâu vào thung lũng hẹp hơn và sâu hơn của giọng nói. Thung lũng đó nối trực tiếp với giáo dục, một ưu tiên chính sách quốc gia. Các tài liệu của iFlytek tại World Digital Education Conference cho thấy một công ty đang lấp đầy đời sống người tiêu dùng và nơi làm việc bằng các thiết bị vật lý như sổ tay AI, máy ghi âm, bút từ điển và máy phiên dịch. Công ty dùng công nghệ nhận dạng giọng nói của mình để tự động xử lý các công việc trí tuệ như cuộc họp, biên bản và phân tích tài liệu. Trong lĩnh vực dịch thuật, công ty cung cấp cùng lúc thiết bị cầm tay, thiết bị đeo và giải pháp đám mây, tiến vào một thị trường làm hạ thấp rào cản ngôn ngữ.

Trong khi các công ty khác cạnh tranh ở các lớp vô hình của phần mềm và điện toán đám mây, iFlytek bán những thiết bị con người có thể cầm trên tay và đặt mình bên cạnh tai và miệng của họ. Giọng nói là một giao diện xuất hiện trước màn hình. Nó trao cho AI một lối vào ngay cả ở những nơi việc đọc và viết còn khó khăn.

Tuyển phòng thủ của iFlytek không nằm ở doanh số thiết bị, mà ở sự bám sát các tiêu chuẩn giáo dục công lập cấp quốc gia. Theo báo cáo của Korea Information Society Development Institute về chiến lược AI Plus của Trung Quốc, Quốc vụ viện Trung Quốc, thông qua văn kiện Opinions on Deepening AI Plus Action, đã đưa các đối tác học tập thông minh và giáo dục phối hợp giữa con người với máy móc vào chính sách quốc gia dành cho ngành giáo dục. Ngôn ngữ của một văn kiện chính sách đi xuống thị trường công nghiệp. Khi nhà nước viết trong văn kiện rằng cần nuôi dưỡng các đối tác học tập thông minh, văn kiện ấy nhanh chóng trở thành nhu cầu khổng lồ của giáo dục công lập đặt trước các doanh nghiệp. Trên sân khấu của World Digital Education Conference, iFlytek biến chương trình nghị sự quốc gia này thành một hệ thống vận hành thực tế.

Các lãnh đạo công ty xuất hiện với tư cách diễn giả tại các phiên của hội nghị và nói rằng AI sẽ thiết kế lại cách đào tạo nhân tài. Bằng việc khởi động một sáng kiến AI cho phát triển giáo dục, họ tạo ra lý do để bước vào thị trường giáo dục công lập. Trong khu triển lãm, các dòng sản phẩm xuất hiện, nơi công nghệ mô hình lớn bước vào hoạt động giảng dạy trong lớp học. Ở đây có thể thấy rõ một vòng vận động: chính sách mở cửa, doanh nghiệp bước qua cánh cửa đó, rồi sản phẩm của họ thể hiện mục tiêu của chính sách bằng hình thái vật chất.

Năm sau, vị thế của công ty được nâng lên một bậc. iFlytek dẫn dắt việc công bố chính thức một đề xuất xây dựng hệ thống tiêu chuẩn giáo dục số quốc tế và một khung tham chiếu chung cho các mô hình lớn trong giáo dục. Việc công ty cùng National Institute of Education Sciences của Trung Quốc phát triển và công bố một gia sư thông minh cho giáo dục khoa học ở bậc tiểu học và trung học cho thấy rõ sự gần gũi này. Gia sư đó được thiết kế để vận hành thông qua một khung gọi là chuỗi tư duy của giảng dạy, nhằm gợi mở quá trình lập luận của học sinh.

Điều này có nghĩa là chính phủ đang làm việc với một công ty AI tư nhân để đưa vai trò của giáo viên trường công vào một hệ thống. Ở đây, hoạt động kinh doanh của công ty và chính sách của nhà nước không phải là hai thực thể xa lạ. Công ty lấp đầy hạ tầng vật chất của lớp học bằng sản phẩm của mình, trong khi chính phủ bao bọc những sản phẩm đó bằng ngôn ngữ của tiêu chuẩn quốc gia. Khi một công ty ngồi tại bàn nơi các tiêu chuẩn được xây dựng, những công ty đi sau phải vận hành trên nền các tiêu chuẩn ấy. Ai nắm tiêu chuẩn, người đó trở thành người gác cổng của thị trường.

Vị thế đã giành được trong giáo dục công lập dẫn thẳng sang thị trường tiêu dùng. iFlytek đưa nội dung phổ biến lên riêng các thiết bị học tập AI của mình, cố giữ cả giọng nói lẫn nội dung giáo dục trong thị trường giáo dục công lập do nhà nước dẫn dắt và thị trường dạy thêm

tư nhân, nơi phụ huynh sẵn sàng mở ví. Uy tín có được trong giáo dục công lập chuyển thành niềm tin đối với thiết bị giáo dục tư nhân, rồi dữ liệu thu được từ giáo dục tư nhân lại cải thiện sản phẩm giáo dục công lập. Hai thị trường nuôi lẫn nhau. Công bằng mà nói, ở đây cũng cần có một tiếng nói khác. Trong giới học thuật phương Tây, đã có lo ngại rằng AI đang đi nhanh hơn hệ thống học tập và có thể tạo ra những ảnh hưởng khó đoán đối với tư duy phản biện và sức sáng tạo của trẻ em.

Cùng một công nghệ ấy được một bên đưa vào lớp học không chút do dự, trong khi bên kia đo đếm ngưỡng cửa rất cẩn thận. Bên nào đúng không phải là điều mà các tài liệu hiện có có thể quyết định. Đây vẫn là một câu hỏi mở, chỉ có thể trả lời sau khi dữ liệu tích lũy nhiều hơn. Việc học sinh nhỏ tuổi học theo chuỗi suy nghĩ của một gia sư AI là con đường đào tạo nhân tài, hay là con đường xóa mất quá sớm cơ hội tự vật lộn và tự học của các em, không phải là điều văn bản này có thể trả lời thay họ.

Đặt năm công ty cạnh nhau, một mẫu hình chung hiện ra. Tất cả đều đang cố biến trí tuệ nhân tạo thành một khoản mục trên bảng doanh thu, xây hạ tầng, mô hình và dịch vụ như một cơ thể thống nhất, rồi di chuyển trong những đường ranh chính sách do chính phủ vẽ ra. Alibaba đặt lớp hạ tầng dày và vận hành cả bàn tay mở quyền truy cập lẫn bàn tay bán hàng. Tencent biến trí tuệ nhân tạo thành doanh thu trên một nền tảng lớn mà họ đã sở hữu. Baidu cố thiết kế lại chính kiến trúc mô hình để kéo chi phí xuống. ByteDance đưa dữ liệu video và mạng xã hội của hàng trăm triệu người dùng quay trở lại hệ thống của mình. iFlytek bắt tay với các tiêu chuẩn quốc gia trong thung lũng hẹp của giọng nói và giáo dục.

Đây là năm con đường khác nhau, nhưng tất cả đều chạy qua cùng một cánh đồng đã được các văn kiện của chính phủ đánh dấu sẵn. Khi những văn kiện như ý kiến của Quốc vụ viện về AI Plus trước tiên đặt hướng cho giáo dục, sản xuất và tiêu dùng, doanh nghiệp sẽ đẩy vốn và mô hình theo các hướng ấy. Chính sách tạo ra nhu cầu, vốn xây hạ tầng, và dữ liệu làm mô hình lớn lên. Bức tranh hiện ra là ba lực cùng quay trong một quốc gia.

Dù vậy, sẽ là sai lầm nếu đọc mô hình này như một câu chuyện chiến thắng trơn tru. Các con số doanh thu và đầu tư do doanh nghiệp công bố có sức nặng của hồ sơ công khai, nhưng những mục như hiệu năng chip tự phát triển, thứ hạng mô hình, và số lượng chính xác các mô hình phái sinh thường dựa trên các tiêu chuẩn thay đổi hoặc tuyên bố của chính công ty. Như các sự việc liên quan đến nhân sự rời đi cho thấy, ngay cả một cấu trúc được xây rất dày cũng vẫn có những mắt xích mỏng. Sức ép bên ngoài từ các biện pháp kiểm soát xuất khẩu của U.S. đè lên cả năm công ty, và giá cả cùng sự bất định của tài nguyên tính toán do các biện pháp đó tạo ra đang phủ bóng lên các trung tâm dữ liệu ở nước ngoài của Alibaba, kiến trúc tiết kiệm chi phí của Baidu, và kho dự trữ chip của ByteDance.

Vì vậy, full stack mà năm công ty này xây dựng chưa giống một pháo đài đã hoàn tất, mà giống một cấu trúc vẫn đang được dựng lên, trong đó chính sách nhà nước, vốn doanh nghiệp,

và dữ liệu người dùng liên tục truyền lực qua lại. Cấu trúc ấy sẽ cao đến đâu, và vết nứt sẽ xuất hiện ở đâu, chỉ có thể rõ hơn khi các hồ sơ công bố tiếp theo và các văn bản chính sách tiếp theo xuất hiện. Điều hiển có thể xác nhận chỉ là thế này: năm công ty đang đi về cùng một đích bằng những tuyến đường khác nhau, và tất cả các tuyến đường đó đều được vẽ trên cùng một bản đồ văn kiện nhà nước. Đọc Năm Công Ty Công Nghệ Lớn Qua Trụ Sở Doanh Nghiệp Và Địa Chỉ

Alibaba (, Alibaba) có một cấu trúc trong đó công ty holding niêm yết tại Cayman Islands chồng lên các công ty vận hành bên trong China. Văn phòng điều hành chính được dùng trong các hồ sơ đối ngoại là No. 969 Wenyi West Road, Yuhang District, Hangzhou, Zhejiang Province, China (969). Thông tin liên hệ công ty và hướng dẫn khuôn viên cũng ghi No. 960-1 Wenyi West Road, Yuhang District, Hangzhou, mã bưu chính 311121.

Khi tìm hiểu chiến lược AI của Alibaba, địa chỉ Hangzhou này không phải chi tiết trang trí. Điện toán đám mây, thương mại điện tử, logistics, và dữ liệu thanh toán đã được tích lũy trong nhiều năm trong cùng một vùng đô thị, rồi Tongyi Qianwen (, Qwen) và Aliyun (, Alibaba Cloud) được xây trên nền đó. Trụ sở của Alibaba không hẳn là phòng thí nghiệm của một công ty mô hình, mà là trung tâm vận hành nơi dữ liệu, điện toán đám mây, và thương mại gặp nhau.

Công ty holding niêm yết của Tencent (, Tencent) giao dịch tại Hong Kong, nhưng trụ sở của tập đoàn đặt tại Tencent Binhai Towers, số 33 Haitian 2nd Road, quận Nanshan, Shenzhen 518054 (33). Vị trí này, Shenzhen, nói lên phần nào tính cách của công ty.

Tencent là một công ty nền tảng vận hành cùng lúc trò chơi, mạng xã hội, fintech và dịch vụ đám mây, còn Shenzhen là một thành phố dày đặc các công ty phần cứng, viễn thông và sản xuất. Cách mô hình Hunyuan (, Hunyuan) đi vào WeChat (, WeChat), trò chơi, quảng cáo và dịch vụ đám mây giống với cấu trúc trụ sở đó. Trụ sở của Tencent không phải là một phòng thí nghiệm nghiên cứu thuần túy. Đó là phòng điều khiển cho các dịch vụ được hàng trăm triệu người sử dụng.

Văn phòng điều hành chính của Baidu (, Baidu) đặt tại Baidu Campus, số 10 Shangdi 10th Street, quận Haidian, Beijing 100085 (10). Khu Shangdi và Zhongguancun từ lâu đã là một trung tâm cũ của ngành tìm kiếm internet và phần mềm Trung Quốc. Vị trí này cũng giải thích vì sao Baidu nói về ERNIE (, ERNIE) và Apollo (, Apollo) cùng nhau. Dữ liệu ngôn ngữ của một công ty tìm kiếm, hệ thống quảng cáo, bản đồ và nghiên cứu lái xe tự hành của họ được nối với nhau trong nguồn nhân lực nghiên cứu của Beijing. Trụ sở của Baidu là nơi ký ức của công cụ tìm kiếm Trung Quốc được đưa vào các mô hình lớn và lái xe tự hành.

ByteDance (, ByteDance) là một công ty tư nhân chưa niêm yết, nên họ không liên tục công bố một địa chỉ trụ sở duy nhất theo cách một công ty niêm yết làm trong hồ sơ công bố. Phần giới thiệu chính thức của công ty nói rằng công ty được thành lập năm 2012 bởi một nhóm do Zhang Yiming () và Liang Rubo () dẫn dắt, và có văn phòng tại nhiều thành phố trên thế giới, gồm Beijing, Shanghai, Shenzhen và Hangzhou. Trong thông tin công ty công khai

và các cơ sở dữ liệu bên ngoài, số 48 Zhichun Road, quận Haidian, Beijing (48), thường được ghi là địa chỉ của các pháp nhân liên quan đến Beijing ByteDance.

Địa chỉ này cần được đọc một cách thận trọng. TikTok, Douyin (), Toutiao (), PICO và Volcano Engine () vận hành thông qua các pháp nhân và đơn vị kinh doanh khác nhau, nên một địa chỉ không thể giải thích toàn bộ công ty. Dù vậy, có thể thấy rõ rằng nhóm thuật toán và nhân tài nội dung tập trung quanh Haidian, Beijing, chính là điểm khởi đầu của AI tại ByteDance.

Địa chỉ liên hệ chính thức của iFLYTEK (, iFLYTEK) là No. 666 West Wangjiang Road, Hefei, Anhui, China (666). Hefei là thành phố của University of Science and Technology of China (), đồng thời là nơi đã kết nối nhận dạng giọng nói và công nghệ giáo dục với các dự án quốc gia. Trụ sở của iFLYTEK mang một bầu không khí khác với trụ sở của các công ty nền tảng ở Beijing hoặc Shenzhen. Công ty này từ lâu đã bám vào giọng nói, giáo dục, dịch vụ công, kỳ thi và dữ liệu học tập, và mô hình lớn Spark (, Spark) cũng xuất hiện từ nền tảng đó. Việc trụ sở đặt tại Hefei cho thấy công ty đã đi một con đường gần với giáo dục, giọng nói và khu vực công hơn là cuộc cạnh tranh giữa các nền tảng trung tâm.

Wu Yongming

Wu Yongming (, Eddie Wu) là một trong những nhà đồng sáng lập Alibaba và giữ chức CEO của Alibaba Group từ tháng 9 năm 2023. Tên của ông không được công chúng biết rộng rãi như Jack Ma (, nhưng bên trong Alibaba, ông là nhân vật đã làm việc với công nghệ và sản phẩm trong nhiều năm. Điều đáng nói là ông bước ra phía trước vào thời điểm Alibaba phải tự mô tả lại mình, không phải như một công ty thương mại điện tử, mà như một công ty hạ tầng cloud và AI. Alibaba trong thời gian dài mang hình ảnh của hệ sinh thái Taobao (, Tmall (và Alipay (, nhưng trong thời đại mô hình lớn, Alibaba Cloud và Qwen đã trở thành gương mặt tiếp theo của công ty. Việc Wu Yongming được bổ nhiệm đánh dấu trọng tâm của công ty chuyển từ nhà vận hành nền tảng sang nhà vận hành hạ tầng công nghệ.

Nhiệm vụ mà ông nhận lấy không hề nhỏ. Alibaba là một công ty thương mại điện tử mạnh ở Trung Quốc, nhưng trên thị trường đám mây, họ phải cạnh tranh với Huawei, Tencent và Baidu; còn trong lĩnh vực mô hình lớn, họ cũng phải theo kịp tốc độ của những công ty đang lên như DeepSeek và Moonshot. Wu Yongming chọn gắn đám mây và AI lại với nhau, đặt chúng làm lõi của tập đoàn. Trong các báo cáo kết quả kinh doanh của công ty, câu nói rằng doanh thu liên quan đến AI của Alibaba Cloud đang tăng nhanh xuất hiện lặp đi lặp lại, và dòng mô hình Qwen tạo ra một làn sóng tải xuống lớn trong hệ sinh thái mã nguồn mở. Lựa chọn này rất Alibaba. Nó không dừng lại ở việc xây dựng một mô hình rồi nhận tiếng vỗ tay. Nó nối mô hình đó với việc dùng đám mây và hệ sinh thái nhà phát triển. Cách quản trị của Wu Yongming gần với ngôn ngữ của một người điều hành kinh doanh nền tảng hơn là ngôn ngữ của một phòng thí nghiệm nghiên cứu.

Vị trí của Wu Yongming khó ở chỗ điểm mạnh và điểm yếu của Alibaba đến từ cùng một nơi. Công ty có một mảng kinh doanh đám mây khổng lồ và dữ liệu thương mại. Điều đó cho họ những kênh rộng để bán mô hình dưới dạng dịch vụ. Ở phía bên kia, tổ chức này lớn và đầy những lợi ích cạnh tranh nhau. Khi các báo cáo về việc nhân tài rời khỏi bộ phận mô hình lớn lan ra thị trường và giá cổ phiếu dao động, điểm yếu này hiện rõ. Cuộc cạnh tranh AI không thể thắng chỉ bằng tiền và máy chủ. Những nhà nghiên cứu và kỹ sư thực sự xây dựng mô hình phải tin công ty và ở lại lâu dài. Wu Yongming phải làm ba việc cùng lúc: đầu tư vào hạ tầng đám mây, theo đuổi chiến lược mô hình mở, và giữ nhân tài. Hạ tầng dày đặc của trụ sở Hàng Châu không thể lập tức biến thành lòng trung thành của giới nghiên cứu.

Ý nghĩa của ông trong chiến lược AI của Alibaba là sự cân bằng. Ông phải mở Qwen để kéo các nhà phát triển vào, đồng thời đưa họ quay lại với doanh thu đám mây. Ông phải đáp ứng các chuẩn quản lý trong Trung Quốc, đồng thời phát hành những mô hình mà các nhà phát triển toàn cầu có thể dùng. Alibaba không phải là một công ty phòng thí nghiệm nghiên cứu ở Bắc Kinh. Nó là một công ty vận hành ở Hàng Châu. Wu Yongming là người đang sắp xếp lại công ty vận hành này để nó có thể sống sót trong thời đại mô hình lớn. Nếu ông thành công, Alibaba sẽ trở thành một công ty hạ tầng AI, gắn dữ liệu thương mại điện tử và điện toán đám mây thành một cơ thể. Nếu ông thất bại, Qwen sẽ chỉ còn là tên của một mô hình tốt, trong khi đám mây và thương mại đi theo hai hướng riêng. Nhiệm kỳ của ông nằm ngay trên ngã rẽ đó.

Những lựa chọn của Wu Yongming cũng nằm trong cuộc chuyển giao thế hệ rộng hơn của Alibaba. Nếu thời đại Jack Ma là thời đại của thương mại điện tử và sức hút của nhà sáng lập, thì Alibaba hôm nay phải tự giải thích lại bằng ngôn ngữ của đám mây, mô hình và hệ sinh thái nhà phát triển toàn cầu. Wu Yongming là gương mặt của sự chuyển đổi ấy. Ông không giống một hình mẫu nhà sáng lập thường gửi đi những thông điệp chính trị mạnh mẽ trước công chúng, mà gần hơn với một nhà quản lý nội bộ đang sắp xếp lại một tổ chức phức tạp. Có lẽ đó chính là điều mà cuộc chuyển đổi AI đang đòi hỏi.

Cạnh tranh trong các mô hình lớn không phụ thuộc nhiều vào một thông báo riêng lẻ, mà phụ thuộc vào đầu tư hạ tầng bền bỉ, chính sách giá, phạm vi phát hành mô hình và cách quản lý đối tác. Mở Qwen để nhà phát triển trên khắp thế giới có thể dùng, đồng thời dẫn dòng sử dụng đó quay lại thành doanh thu cho Alibaba Cloud, là một việc tinh tế. Nếu giữ quá kín, hệ sinh thái sẽ không lớn lên. Nếu mở quá rộng, việc kiếm tiền sẽ trở nên khó khăn. Wu Yongming là người điều chỉnh độ rộng của cánh cửa đó.

Bên trong Alibaba, vai trò lãnh đạo của ông giống công việc dọn dẹp âm thầm hơn. Khi điện toán đám mây, thương mại, nghiên cứu mô hình và hoạt động ở nước ngoài đi với tốc độ khác nhau, người quản lý phải quyết định mảng kinh doanh nào nhận tiền và nhân sự trước. Wu Yongming nắm trong tay bảng phân bổ đó. Một chiến lược AI không tự bộc lộ trong khẩu hiệu, mà trong ngân sách.

Ma Huateng

Ma Huateng (, Pony Ma) là đồng sáng lập, chủ tịch kiêm giám đốc điều hành của Tencent. Ông là một nhà sáng lập hiếm hoi trong ngành internet Trung Quốc đã trụ lại được trong thời gian dài. Ông đã nhiều lần dịch chuyển trọng tâm của công ty, từ trình nhắn tin QQ sang WeChat, từ trò chơi sang fintech, rồi đến điện toán đám mây và nội dung. Chiến lược AI của Tencent nằm trên cùng đường chuyển động đó. Thay vì xem các mô hình lớn như những chiến tích nghiên cứu đứng riêng, Ma Huateng giải thích chúng như những công cụ được đặt vào các dịch vụ hiện có để thay đổi doanh thu, thời gian người dùng ở lại, hiệu quả quảng cáo và tốc độ sản xuất trò chơi. Phần trong báo cáo thường niên năm 2025 mô tả trí tuệ nhân tạo như một cách cải thiện hiệu suất trong quảng cáo, trò chơi và điện toán đám mây cho thấy rõ bản năng quản trị của ông. Với Tencent, AI là động cơ khiến bàn cờ hiện có vận hành chính xác hơn trước khi nó trở thành một biển hiệu mới.

Sức mạnh của Ma Huateng nằm ở các điểm chạm với người dùng. WeChat đã ăn sâu vào đời sống hằng ngày ở Trung Quốc. Thanh toán, trò chuyện, mini program, tin tức, tin nhắn công việc và dịch vụ công đều tụ lại trong một ứng dụng. Khi mô hình Hunyuan đi vào hệ sinh thái này, nó gặp được những dữ liệu mà một chatbot đứng riêng khó có thể có được. Người dùng có thể dùng các chức năng AI mà không cần mở một ứng dụng mới, còn công ty có thể nối luồng sử dụng đó với quảng cáo, thương mại, trò chơi và đám mây. Ma Huateng hiểu rõ lợi thế này. Vì vậy Tencent thích đặt AI phía sau các dịch vụ mà họ đã sở hữu hơn là đưa ra những tuyên bố nhằm làm cả thế giới kinh ngạc vì AI. Cách làm lặng lẽ này ít xuất hiện rục rờ trên tiêu đề báo chí, nhưng có thể tồn tại lâu hơn trên sổ sách kinh doanh.

Điểm yếu của ông cũng đến từ chính nơi đó. Vì hệ sinh thái WeChat quá lớn, các kết quả sai của mô hình, vấn đề quyền riêng tư, và những chuyện liên quan đến trò chơi cũng như nội dung dành cho trẻ vị thành niên có thể nhanh chóng trở thành vấn đề xã hội. Tencent cố gắng vận hành các agent bên trong những hộp cát bảo mật vì họ hiểu gánh nặng này. Ma Huateng không giống một người lạc quan tuyệt đối về công nghệ. Ông là một nhà quản lý biết rõ rủi ro của một công ty có khối lượng người dùng khổng lồ. Việc ông chọn đặt các mô hình lớn vào những khu vực có thể kiểm soát như nhắm mục tiêu quảng cáo, sản xuất trò chơi, dịch vụ khách hàng và công cụ đám mây, thay vì vội vàng đẩy chúng ra tuyến đầu, là một thứ chủ nghĩa thương mại thận trọng. AI của Tencent tiến lên bằng vô số lần cài vào nhỏ, chứ không bằng những tuyên bố táo bạo.

Vai trò của Ma Huateng trong AI Trung Quốc là "AI bên trong nền tảng". Trong khi DeepSeek làm rung chuyển giá của các mô hình mở và Alibaba đẩy hạ tầng đám mây ra tuyến đầu, Tencent nằm trong tay thói quen của người dùng. Những thói quen đó rất khó mua bằng tiền. Nhưng thói quen không kéo dài mãi mãi. Người dùng trẻ đang chuyển thời gian của họ sang các nền tảng video ngắn và nội dung như Douyin và Xiaohongshu (). Ma Huateng phải bảo

vệ chiều sâu của WeChat, đồng thời bảo đảm rằng ông không bỏ lỡ giao diện mới của kỷ nguyên AI. Ông sẽ đưa Hunyuan ra phía trước đến đâu, và sẽ giấu nó như một chức năng hậu trường đến mức nào, đó là câu hỏi trung tâm trong chiến lược AI của Tencent. Cách quản lý của Ma Huateng dựa ít hơn vào những tuyên bố hào nhoáng về mô hình, và nhiều hơn vào việc không đánh mất thói quen của người dùng.

Phong cách quản lý lặng lẽ của Ma Huateng vừa là sức mạnh vừa là điểm yếu trong cuộc đua AI. Tencent đôi khi trông có vẻ chậm. Họ có ít tuyên bố về mô hình khiến thế giới giật mình hơn Alibaba hay DeepSeek, và cái tên Hunyuan thường đứng phía sau WeChat hoặc các thương hiệu trò chơi của công ty. Nhưng Tencent biết người dùng dành thời gian ở đâu. Nhân vật trong trò chơi, gợi ý quảng cáo, thanh toán và dịch vụ khách hàng, cùng các kết nối Mini Program đều là những nơi mà các chức năng AI nhỏ biến thành tiền. Ma Huateng đã chọn một cách làm từ từ chiếm lấy những nơi đó.

Phương pháp này trông kém nổi bật hơn trên các bảng xếp hạng công nghệ, nhưng có thể tạo sức mạnh rõ rệt trên sổ sách của công ty. Tuy vậy, câu chuyện sẽ đổi khác nếu giao diện mới ra đời bên ngoài WeChat. Nếu người dùng bắt đầu dùng các AI agent như những ứng dụng riêng, hoặc dùng chúng ở cấp hệ điều hành, vòng khép kín của WeChat cũng có thể bắt đầu lung lay. Điều Ma Huateng phải ngăn chặn không chỉ là hiệu năng mô hình của các đối thủ. Ông phải ngăn màn hình đầu tiên của người dùng thay đổi.

Ông là một nhà quản lý đặt công nghệ một cách lặng lẽ vào nơi người dùng vốn đã ở lại, thay vì cố thuyết phục họ bằng cách đưa công nghệ lên trước. Phương pháp này là một mô hình của AI nền tảng Trung Quốc. Ngay cả khi tên mô hình không xuất hiện dày đặc, công ty vẫn thắng nếu giá quảng cáo tăng, thời gian sản xuất game giảm, và chi phí dịch vụ khách hàng đi xuống. Ma Huateng đang nhắm tới chiến thắng âm thầm đó.

Robin Li

Li Yanhong (, Robin Li) là đồng sáng lập kiêm giám đốc điều hành của Baidu. Ông bắt đầu với vai trò gương mặt đại diện của ngành công cụ tìm kiếm Trung Quốc. Tìm kiếm có mối liên hệ tự nhiên với các mô hình lớn. Tìm kiếm là công nghệ hiểu câu hỏi của người dùng và xếp hạng tài liệu, còn mô hình lớn trả lời câu hỏi và lắp ghép tài liệu theo cách mới. Vì lý do đó, quyết định của Baidu đưa ERNIE () ra tuyến trước không phải là một bước đuổi theo muện màng, mà là phần mở rộng của một trục công nghệ đã có từ lâu. Li Yanhong đã xây dựng quảng cáo tìm kiếm, đồ thị tri thức, xử lý ngôn ngữ tự nhiên, bản đồ, và xe tự lái bên trong công ty, rồi cố biến sự tích lũy ấy thành các dịch vụ sinh ra từ AI. Khuôn viên Baidu ở Shangdi, Bắc Kinh, là nơi ký ức của kỷ nguyên tìm kiếm Trung Quốc bước sang kỷ nguyên mô hình lớn.

Điểm nổi bật trong cách quản lý của ông là chi phí và thương mại hóa. Khi Baidu tung ra nhiều thế hệ trong họ ERNIE, công ty nói về hiệu năng mô hình, nhưng cũng nhấn mạnh chi phí

suy luận thấp hơn và một cấu trúc mà khách hàng doanh nghiệp có thể dùng được. Họ ERNIE 5.1, được giới thiệu vào năm 2026, đưa ra kích thước nhỏ hơn cùng với năng lực suy luận và năng lực xử lý văn bản. Robin Li hiểu rằng Baidu không thể thắng nếu chỉ để mô hình nằm lại như một thành tựu nghiên cứu khổng lồ. Thị trường quảng cáo tìm kiếm đã trưởng thành, còn thời gian của người dùng di động bị chia giữa ByteDance và Tencent. Để Baidu giành lại vị trí dẫn đầu, mô hình phải làm thay đổi kết quả tìm kiếm, tạo nhu cầu cho dịch vụ đám mây doanh nghiệp, và kết nối với Apollo tự lái. AI của Robin Li là câu tiếp theo của tìm kiếm.

Chương trình tự lái Apollo (), thứ ông đã thúc đẩy trong nhiều năm, cũng khó có thể bị gạt ra khỏi chiến lược AI của Baidu. Trong khi các công ty công nghệ lớn khác đặt trọng tâm vào chatbot và dịch vụ đám mây, Baidu dành nhiều năm thu thập dữ liệu từ đường phố. Tự lái dùng dữ liệu khác với tài liệu web. Nó cần bản đồ, lidar và camera, hồ sơ lái xe, cùng các phán đoán an toàn. Robin Li muốn nối dữ liệu này với các mô hình lớn. Nếu thành công, Baidu sẽ trở thành một công ty AI có bản sắc riêng, nối tìm kiếm với tự lái. Nếu thất bại, khoản đầu tư vào tự lái có thể vẫn là một chi phí kéo dài. Vì vậy chiến lược của ông luôn đòi hỏi tầm nhìn dài hạn. Dùng dòng tiền từ tìm kiếm để nuôi các mảng kinh doanh tương lai là cách ổn định, nhưng nếu thị trường không chờ đợi, nó có thể trông giống một chiến lược chậm chạp.

Robin Li đại diện cho một câu hỏi trong AI Trung Quốc: "Một công ty công nghệ cũ có thể trở lại không?" Alibaba và Tencent nắm hệ sinh thái nền tảng, còn ByteDance nắm thời gian con người dành cho video ngắn. Baidu từng nắm cổng vào câu hỏi và tri thức. Trong thời đại mô hình lớn, cổng vào ấy có thể lớn lên trở lại, hoặc chatbot và agent có thể đẩy ô tìm kiếm sang một bên. Robin Li phải bảo vệ ô tìm kiếm trong khi xây dựng một cỗ máy trả lời vượt ra ngoài nó. Nhiệm vụ mâu thuẫn ấy đang đặt trước mặt ông. Baidu tiếp tục cải thiện ERNIE không chỉ vì cuộc đua hiệu năng mô hình. Đó là công việc sinh tồn, nhằm biến hoạt động tìm kiếm cũ thành một cánh cửa mới mà không đóng cánh cửa cũ lại.

Khó khăn của Robin Li đến từ việc Baidu từng là lối vào internet Trung Quốc. Ô tìm kiếm từng là cánh cửa nơi con người gặp thông tin. Giờ đây, dòng đề xuất, video ngắn, chatbot và agent cùng chia sẻ cánh cửa ấy. Robin Li phải giành lại thời gian đã mất. Nếu ERNIE sắp xếp kết quả tìm kiếm tốt hơn và nổi quảng cáo, dịch vụ đám mây, cùng tự lái, Baidu có thể một lần nữa trở thành cổng vào tri thức. Nếu người dùng thay vào đó gửi câu hỏi của họ cho AI bên trong các ứng dụng khác và không còn đi qua ô tìm kiếm, quyền lực cũ của Baidu sẽ tiếp tục co lại.

Vì vậy, chiến lược AI của ông mang tính cấp bách. Robin Li là người đã mở ra thời đại tìm kiếm, và cũng là người phải chấp nhận thời đại sau tìm kiếm. Ông càng cố bảo vệ ô tìm kiếm, công ty càng trông cũ kỹ. Baidu càng tự mình đi ra khỏi ô tìm kiếm, họ càng ăn vào một phần hoạt động kinh doanh hiện có. Gánh nặng tự thay thế chính mình này là lý do quá trình chuyển đổi của Baidu khó đến vậy.

Sức mạnh còn lại trong tay ông vẫn mang màu sắc của tri thức. Baidu từ lâu đã quan sát điều con người thắc mắc. Nếu có thể nối dữ liệu câu hỏi ấy với hệ thống quảng cáo tìm kiếm, bản đồ và dữ liệu lái xe tự động, AI của Baidu sẽ có một chất liệu khác với các công ty khác. Nhiệm vụ của Robin Li là biến những tài sản cũ này thành nhiên liệu cho một mô hình mới, chứ không để chúng nằm lại như một di sản già nua.

Liang Rubo

Liang Rubo () là đồng sáng lập và giám đốc điều hành của ByteDance. Sau khi công ty thuật toán đề xuất do Zhang Yiming () sáng lập phát triển thành một nền tảng toàn cầu, Liang Rubo tiếp quản tổ chức khổng lồ đó. Muốn hiểu AI của ByteDance, phải nhìn vào ý nghĩa của sự kế nhiệm này. Zhang Yiming đã dùng thuật toán để định hình lại thời gian con người dành cho việc tiêu thụ nội dung. Liang Rubo phải vận hành công ty thuật toán ấy trong khi cùng lúc chống chịu áp lực chính trị, quy định toàn cầu, cạnh tranh nền tảng và đầu tư vào các mô hình lớn. TikTok và Douyin là những sản phẩm có một số hệ thống đề xuất mạnh nhất thế giới. Đằng sau các hệ thống đề xuất ấy là dữ liệu hành vi người dùng và các mô hình hiểu video. Liang Rubo chịu trách nhiệm đưa những tài sản đó vào AI tạo sinh.

Điểm mạnh của ByteDance là video và thời gian người dùng ở lại trên nền tảng. Đây không phải là một công ty chỉ nắm giữ một kho tài liệu web dạng chữ rất lớn. ByteDance đã tích lũy video ngắn, âm nhạc, bình luận, thời lượng xem, mẫu chỉnh sửa và phản ứng với quảng cáo. Khi AI tạo sinh tiến về video và các hệ thống đa phương thức, khối dữ liệu này có sức mạnh lớn hơn. Nhiệm vụ của Liang Rubo là biến dữ liệu đó thành Doubao (), Volcano Engine (), các mô hình tạo video và công cụ cho nhà phát triển. Ông là một nhà điều hành có sức mạnh phân phối hơn là một người xây mô hình. ByteDance có thể đặt một tính năng mới vào trong ứng dụng và thử nghiệm nó trên hàng trăm triệu người. Các công ty phòng thí nghiệm nghiên cứu khó theo kịp tốc độ đó. Liang Rubo là người quản lý tốc độ ấy.

Nhưng AI của ByteDance va mạnh vào những bức tường chính trị và quản lý. TikTok đã đứng ở trung tâm các tranh chấp về quyền riêng tư và an ninh tại Hoa Kỳ và châu Âu. Nếu các mô hình lớn được kết hợp với dữ liệu nền tảng này, tranh chấp sẽ còn lớn hơn. Liang Rubo phải cùng lúc xử lý quản lý trong Trung Quốc và sự nghi ngờ ở nước ngoài. Ông phải quyết định mô hình nào giữ trong Trung Quốc, dịch vụ nào đưa ra nước ngoài, và dữ liệu nào phải tách theo khu vực. Vận hành đi trước công nghệ. Ở điểm này, ông khác với một nhà sáng lập xuất thân từ nghiên cứu. Sức cạnh tranh AI của ByteDance đến không nhiều từ một bài nghiên cứu xuất sắc, mà từ khả năng giữ cho vô số thử nghiệm sản phẩm và phản ứng trước quy định chạy liên tục, không bị đứt đoạn.

Vai trò của Liang Rubo trong AI Trung Quốc là bước chuyển "từ đề xuất sang tạo sinh". Thuật toán đề xuất chọn video tiếp theo mà người dùng có khả năng thích. Mô hình tạo sinh tạo ra video mà người dùng muốn, viết nội dung quảng cáo, hoặc xử lý hỗ trợ khách hàng. Cái trước là

công nghệ lựa chọn, cái sau là công nghệ sản xuất. ByteDance là một trong số ít công ty có thể nối hai thứ này lại với nhau. Nếu Liang Rubo thành công, ByteDance sẽ chuyển từ một nền tảng hiển thị nội dung thành một nền tảng tạo ra nội dung. Nếu ông thất bại, lượng thời gian người dùng khổng lồ của ByteDance vẫn sẽ kiếm ra tiền, nhưng các công ty khác sẽ lấy thị trường công cụ sản xuất trong kỷ nguyên mô hình lớn. Cách quản lý của ông là một phép thử xem tốc độ của video ngắn có thể được biến thành một cuộc chuyển đổi công nghệ dài hơi hay không.

Điều khiến Liang Rubo khác với Zhang Yiming là ông đại diện cho thời điểm vận hành, chứ không phải thời điểm sáng lập. ByteDance đã trở thành một công ty toàn cầu khi được chuyển giao cho ông. Công ty đã quá lớn, tranh chấp thì nhiều, và các chính phủ trên khắp thế giới đang theo dõi sát. Dẫn dắt một cuộc chuyển đổi AI trong một tổ chức như vậy có thể phức tạp hơn việc tạo ra một công ty mới. Các động cơ đề xuất của Douyin và TikTok vốn đã mạnh. Câu hỏi là liệu những động cơ ấy có thể tiếp tục giữ vị trí trung tâm trong thời đại AI tạo sinh hay không.

Liang Rubo phải biến một thuật toán chọn nội dung thành một mô hình tạo ra nội dung. Với nhà quảng cáo, mô hình đó phải tạo được video và lời quảng cáo. Với người dùng, nó phải cung cấp công cụ sáng tạo. Với khách hàng doanh nghiệp, ByteDance phải bán năng lực mô hình thông qua Volcano Engine. Sức mạnh của ByteDance nằm ở tốc độ thử nghiệm. Điểm yếu của họ là sự thiếu tin cậy mà công ty phải đối mặt trên khắp thế giới. Liang Rubo là người vận hành phải xử lý cả hai việc ấy cùng lúc.

Phong cách lãnh đạo của ông ít kịch tính hơn huyền thoại về nhà sáng lập, nhưng ByteDance lúc này lại cần đúng thứ cảm giác vận hành ấy. Phát hành một mô hình dễ hơn nhiều so với việc vận hành nó mà không xảy ra sự cố bên trong các dịch vụ có hàng trăm triệu người dùng. Liang Rubo phải định vị lại một công ty được xây trên thuật toán đề xuất thành một công ty AI tạo sinh. Đó là công việc lặng lẽ, nhưng sức nặng của nó là thật.

Nếu ông thất bại, ByteDance vẫn sẽ là một nền tảng lớn, nhưng vị trí dẫn đầu trong các công cụ sản xuất bằng AI có thể rơi vào tay một công ty khác. Vì vậy, nhiệm vụ của Liang Rubo không phải là phòng thủ. Đó là giành lấy màn hình kế tiếp trước tiên.

Liu Qingfeng

Liu Qingfeng () là nhà sáng lập iFlytek và đã lãnh đạo công ty trong thời gian dài. Ông là một gương mặt biểu tượng của ngành nhận dạng giọng nói Trung Quốc. Khác với các công ty nền tảng ở Bắc Kinh và Thâm Quyển, iFlytek lớn lên tại Hợp Phì. Công ty chọn con đường gắn công nghệ giọng nói với giáo dục, dịch vụ công, kỳ thi, dịch thuật và thiết bị thông minh. Điểm mạnh của Liu Qingfeng là khả năng chuyển ngôn ngữ công nghệ thành ngôn ngữ của nhiệm vụ quốc gia. Nhận dạng giọng nói biến lời nói của con người thành văn bản. Ở một quốc gia như Trung Quốc, nơi có nhiều phương ngữ và khoảng cách vùng miền rộng, công nghệ này nổi

thẳng với các lĩnh vực công như giáo dục, hành chính, hỗ trợ tiếp cận cho người khuyết tật, hồ sơ tại tòa án và bệnh viện. Liu Qingfeng đã đặt mối nối này vào trung tâm hoạt động kinh doanh.

Mô hình lớn Xinghuo (, Spark) của iFlytek cũng đứng trên lịch sử đó. Công ty không đột ngột nhảy vào thị trường chatbot. Họ xây dựng mô hình của mình trên nhiều năm tích lũy dữ liệu giọng nói, nội dung giáo dục, công cụ thi cử và học tập, cùng các sản phẩm lớp học. Liu Qingfeng không trình bày mô hình lớn chỉ như dịch vụ trò chuyện đa dụng. Ông cố đưa chúng vào dòng công việc nơi giáo viên soạn bài, học sinh giải bài, phát âm được sửa và văn bản hành chính được xử lý. Việc iFlytek giới thiệu các mô hình lớn giáo dục và gia sư thông minh tại World Digital Education Conference cũng đi theo hướng ấy. Khi chính phủ nói về chuyển đổi AI trong giáo dục, công ty cho thấy điều đó có nghĩa gì thông qua sản phẩm vật lý và các tiêu chuẩn đề xuất. Liu Qingfeng xử lý tốt nhịp qua lại này.

Những rủi ro mà công ty của ông đối mặt cũng rất rõ. Giáo dục và dịch vụ giọng nói xử lý dữ liệu nhạy cảm. Hồ sơ học tập của một đứa trẻ, phát âm, kết quả kiểm tra, phương pháp lớp học của giáo viên và nội dung các cuộc họp của cơ quan công quyền đều có thể trở thành vật liệu cho một mô hình. Liu Qingfeng phải tìm câu trả lời giữa việc phổ biến công nghệ và bảo vệ dữ liệu. Công ty càng đi sâu vào giáo dục công, thị trường càng ổn định, nhưng trách nhiệm xã hội cũng lớn lên cùng lúc. Các câu hỏi nối nhau xuất hiện. Gia sư AI có thu hẹp quá mức cách suy nghĩ của một đứa trẻ không? Nó có vượt khỏi việc hỗ trợ phán đoán của giáo viên và bắt đầu thay thế phán đoán đó không? Bao nhiêu dữ liệu giọng nói được lưu trữ, và lưu trong bao lâu? Cách quản trị của Liu Qingfeng không thể tránh những câu hỏi này.

Trong ngành AI Trung Quốc, Liu Qingfeng cho thấy khả năng tồn tại của một "công ty chuyên môn dựa trên địa phương." Không phải công ty nào cũng cần xây dựng nền tảng ở Haidian của Bắc Kinh hay Nanshan của Thâm Quyển. Lịch sử của iFlytek cho thấy hệ sinh thái đại học và nghiên cứu của Hợp Phì, chính quyền địa phương và nhu cầu từ giáo dục công có thể kết hợp với nhau để nuôi lớn một công ty chuyên về AI giọng nói. Nếu Liu Qingfeng muốn sống sót trong cuộc đua mô hình lớn, ông không nên chỉ bám vào bảng xếp hạng mô hình đa dụng. Thứ công ty của ông nắm giữ từ lâu là thế giới thật của lớp học và giọng nói. Khi Spark trở thành một mô hình hiểu bối cảnh ấy tốt hơn, iFlytek có thể tồn tại ở một vị trí khác với Alibaba hay Tencent. Liu Qingfeng là nhân vật cho thấy AI Trung Quốc không chỉ là câu chuyện của các gã khổng lồ nền tảng.

Con đường của Liu Qingfeng cho thấy AI Trung Quốc không thể được giải thích chỉ bằng cạnh tranh nền tảng ở vùng thủ đô. Hefei không quen thuộc với công chúng như Beijing hay Shenzhen, nhưng đây là thành phố nơi University of Science and Technology of China, các viện nghiên cứu và chính sách công nghiệp của chính quyền địa phương đã cùng hội tụ. Trên mảnh đất ấy, iFlytek đã đào một cái giếng duy nhất trong thời gian dài: nhận dạng giọng nói. Sự bền

bỉ đó đang có ý nghĩa trở lại trong thời đại mô hình lớn. Khi mọi mô hình chỉ cạnh tranh bằng khả năng tạo câu, iFlytek nắm trong tay âm thanh lời nói và dữ liệu từ lĩnh vực giáo dục.

Liu Qingfeng đang cố đưa tài sản này vào mô hình Spark. Phép thử của ông không phải là thắng trên bảng xếp hạng mô hình đa dụng. Đó là chứng minh liệu AI giọng nói có thật sự giảm bớt công việc trong lớp học và phòng họp, tòa án và bệnh viện, cũng như các môi trường dịch vụ công hay không. Con đường này có thể trông hẹp, nhưng trên thực tế có thể rất rộng. Nếu mô hình trở thành thứ mà các chuyên gia sẵn sàng trả tiền để dùng trong những môi trường chuyên môn, iFlytek không phải đối đầu trực diện với các gã khổng lồ nền tảng.

Công nghệ giọng nói mà ông đã bám giữ suốt thời gian dài đang có một thân thể mới trong thời đại mô hình lớn. Con người vẫn đặt câu hỏi bằng lời nói và học bằng lời nói. Trong lớp học và cơ quan công quyền, bệnh viện và phòng họp, giọng nói thường tự nhiên hơn bàn phím. Liu Qingfeng hiểu lối vào tự nhiên này. Sức cạnh tranh của iFlytek đến từ quãng thời gian họ đã dành để bảo vệ lối vào đó.

Nếu Liu Qingfeng giữ con đường này, bản đồ AI Trung Quốc sẽ rộng hơn. Không chỉ các công ty mô hình của Beijing và các công ty phần cứng của Shenzhen, mà cả công ty giọng nói và giáo dục ở Hefei cũng có thể trở thành một trục của thời đại mô hình lớn.

Chương 7. SenseTime và mô hình công ty niêm yết dành cho các doanh nghiệp thị giác máy tính và trí tuệ không gian

Tài báo cáo thường niên năm 2025 của SenseTime từ trang công bố thông tin của Hong Kong Exchanges and Clearing Limited (HKEX), ngay sau bìa sẽ thấy một bảng xếp các số liệu trong năm năm. Đó là phần Tóm tắt tài chính năm năm. Năm 2021 là RMB 4,7 tỷ, năm 2022 là RMB 3,8 tỷ, năm 2023 là RMB 3,4 tỷ. Dòng đầu tiên đập vào mắt là doanh thu giảm đều qua từng năm. Rồi hướng đi đổi chiều vào năm 2024, và cột năm 2025 ghi con số vượt RMB 5 tỷ. Chính xác là RMB 5,01464 tỷ, tăng 32,9% so với RMB 3,77210 tỷ của năm trước. Trong tài liệu công bố kết quả kinh doanh, công ty mô tả con số này là mức cao kỷ lục trong lịch sử doanh nghiệp.

Nếu nhìn bảng này một lúc, một cái tên quen thuộc với độc giả Hàn Quốc bắt đầu hiện ra khác đi. SenseTime (, SenseTime) từ lâu được biết đến với nhận diện khuôn mặt và phân tích video. Đó là một công ty AI thị giác, với công nghệ được đưa vào hệ thống kiểm soát ra vào bằng camera đô thị của Trung Quốc và chức năng chỉnh ảnh trên điện thoại thông minh. Trong chương này, tôi sẽ đọc các tài liệu để xem vì sao công ty ấy mở trang đầu trong hồ sơ công bố của một công ty niêm yết bằng sự phục hồi doanh thu, sự phục hồi ấy đến từ đâu, và tư cách công ty niêm yết đòi hỏi gì ở họ. Công bố thông tin là lời một công ty để lại cho thế giới dưới chính tên mình, nên ngôn ngữ quảng bá và các con số đã kiểm toán nằm trong cùng một tài liệu. Tách hai thứ ấy ra để đọc là một nửa công việc của chương này.

Các chương trước bắt đầu từ văn kiện nhà nước. Chương này bắt đầu từ tài liệu doanh nghiệp. Điều đó không có nghĩa câu chuyện không liên quan gì đến nhà nước. Khi đọc hồ sơ công bố của các công ty AI Trung Quốc, cái bóng của chính sách thẩm qua từng trang tài liệu. Những lựa chọn của doanh nghiệp, như mua loại bán dẫn nào, xây trung tâm dữ liệu ở đâu, và đưa ứng dụng nào ra phía trước, đều được quyết định trong các điều kiện bên ngoài như kiểm soát xuất khẩu và nỗ lực tự chủ về năng lực tính toán của Trung Quốc. Ở chương trước, tôi đã bàn về cách National Data Administration biến các bộ dữ liệu thành ngôn ngữ của chính sách công nghiệp, và cách Cyberspace Administration of China, hay CAC, rà soát những mô hình nào có thể đi vào các dịch vụ hướng tới công chúng.

Để thấy các đường chính sách ấy rơi xuống báo cáo kết quả kinh doanh của một công ty như thế nào, mở báo cáo tài chính do chính công ty nộp sau khi được kiểm toán viên rà soát sẽ nhanh hơn đọc một văn kiện nhà nước. Lần theo các con số của một công ty, SenseTime, ta có thể thấy nhà nước, doanh nghiệp và dữ liệu lồng vào nhau trong cùng một tài liệu.

Trước hết là bản chất của sự phục hồi. Trọng tâm của lần công bố này không nằm ở việc doanh thu tăng. Trọng tâm nằm ở việc khoản lỗ đã thu hẹp. Lợi nhuận gộp tăng từ RMB 1.61969 tỷ trong năm 2024 lên RMB 2.05642 tỷ trong năm 2025. Khoản lỗ ròng, gánh nặng kéo

dài của công ty, giảm từ RMB 4.30660 tỷ trong năm 2024 xuống RMB 1.78200 tỷ trong năm 2025, chưa bằng một nửa mức của năm trước. Nếu nhìn rộng hơn, mức giảm này còn nổi bật hơn.

Khoản lỗ trước thuế thu nhập là RMB 17.1 tỷ trong năm 2021. Năm đó, SenseTime là một công ty lỗ hơn RMB 10 tỷ. Đến năm 2025, con số này đã giảm xuống RMB 1.65960 tỷ. Trong bốn năm, quy mô khoản lỗ đã co lại gần bằng một phần mười. Tuy vậy, mức giảm này có cả các yếu tố kế toán sẽ xuất hiện ở phần sau, nên nếu đọc thẳng nó như một mức cải thiện tương ứng của hoạt động kinh doanh cốt lõi thì sẽ bỏ lỡ bức tranh thật.

Con số mà công ty nhấn mạnh trong bản công bố là một chỉ tiêu đã điều chỉnh. Công ty cho biết lỗ ròng đã điều chỉnh đã thu hẹp trong bốn kỳ nửa năm liên tiếp, và EBITDA của nửa sau năm 2025 lần đầu tiên chuyển sang dương kể từ khi niêm yết, đạt RMB 380 triệu. EBITDA đã điều chỉnh cả năm cũng thu hẹp từ RMB -3.117 tỷ trong năm 2024 xuống RMB -645.4 triệu trong năm 2025. Cần nói rõ một điểm ở đây. EBITDA là chỉ tiêu lợi nhuận trước lãi vay, thuế, khấu hao tài sản hữu hình và khấu hao tài sản vô hình.

Với một công ty xây dựng trung tâm dữ liệu, khấu hao không phải là khoản nhỏ. Giá trị của nhà xưởng và thiết bị được mua bằng số tiền lớn sẽ được ghi giảm thành chi phí qua từng năm, còn EBITDA là con số tạm thời đặt khoản chi phí đó sang một bên. Vì vậy, EBITDA dương trong nửa cuối năm không có nghĩa là cả công ty đã có lợi nhuận ròng cho cả năm. Xét toàn bộ năm 2025, SenseTime vẫn là một công ty lỗ ròng ở mức khoảng RMB 1.7 tỷ. Đó là lý do công ty chỉ dùng cụm từ 'chuyển sang có lãi' trong hai điều kiện: nửa cuối năm và EBITDA.

Khi đọc một bản công bố thông tin, người đọc phải nhìn vào những điều kiện mà công ty gắn kèm khi chọn các con số có lợi cho mình. Ngay cả khi tiêu đề thông cáo báo chí mở đầu bằng 'doanh thu cao kỷ lục' và 'chuyển sang có lãi', bảng bên dưới vẫn cho thấy khoản lỗ ròng cả năm. Bản công bố thông tin của một công ty niêm yết đặt điều công ty muốn nói và những con số công ty bắt buộc phải báo cáo trong cùng một tài liệu. Đưa ngôn từ có lợi lên phía trước và đặt các số liệu bắt buộc trong phần ghi chú không phải là lỗi. Đó là một phần quy tắc của thể loại công bố thông tin. Khi người đọc biết các quy tắc ấy và đi theo tài liệu xuống tận phần ghi chú, câu mà công ty nhấn mạnh và con số mà công ty không thể che giấu sẽ kiểm tra lẫn nhau.

Vậy tăng trưởng doanh thu đến từ đâu? Phần Management Discussion and Analysis của báo cáo nói rằng mức tăng doanh thu năm 2025 đến từ sự mở rộng liên tục của Generative AI. Nếu nhìn bằng con số, bức tranh là thế này. Doanh thu từ mảng Generative AI tăng 51.0% so với cùng kỳ năm trước, lên RMB 3.6295 billion, và riêng mảng này chiếm 72.4% tổng doanh thu của tập đoàn.

Tầm biến lâu năm của công ty, vị trí số 1 trên thị trường computer vision của Trung Quốc trong chín năm liên tiếp, vẫn xuất hiện trong báo cáo. Nhưng nơi đem tiền về đã chuyển ra phía sau tầm biến ấy. Một công ty facial recognition đã đổi cơ cấu doanh thu thành cơ cấu của một

công ty Generative AI. Câu đó là lõi của sự chuyển dịch kinh doanh được mô tả trong bản công bố thông tin này.

Khi gần ba phần tư doanh thu đến từ một mảng, việc đó có hai mặt. Nó có nghĩa nguồn tăng trưởng đã rõ, và cũng có nghĩa nếu mảng ấy rung lắc, cả công ty sẽ rung lắc theo. Khi mảng vision từng chống đỡ hơn một nửa công ty, quy định về facial recognition hoặc thay đổi trong ngân sách giám sát đô thị là rủi ro của công ty. Giờ đây, cạnh tranh và giá cả trong thị trường Generative AI đã ngồi vào vị trí đó. Rủi ro không biến mất. Nó chỉ chuyển sang một ghế mới.

Mọi cuộc chuyển đổi đều có một nền móng vật chất. Doanh thu từ AI tạo sinh không tăng chỉ nhờ phần mềm. Bên dưới nó là trung tâm dữ liệu. Trong phần Báo cáo của Hội đồng quản trị của báo cáo, Shanghai SenseTime Technology Development, đơn vị vận hành Shanghai Qianlun và Shanghai Lingang Artificial Intelligence Data Center (AIDC), ghi nhận doanh thu năm 2025 là RMB 1.8537 tỷ. Con số này chiếm 37.0 phần trăm tổng doanh thu của tập đoàn. Nói cách khác, hơn một phần ba doanh thu của SenseTime đến từ các trung tâm dữ liệu do chính công ty xây dựng và vận hành.

Các ghi chú về chính sách kế toán cho biết doanh thu từ AIDC và dịch vụ đám mây này được tính phí theo hình thức Subscription hoặc Consumption. Khách hàng truy cập hạ tầng và phần mềm trong suốt thời hạn hợp đồng và dùng tài nguyên khi cần, nên công ty ghi nhận doanh thu Over time, theo tỷ lệ thời gian. Đây là dấu hiệu cho thấy tính chất của hoạt động kinh doanh đang thay đổi, từ một công ty bán phần mềm sang một công ty cho thuê tài nguyên tính toán.

Ở đây, phần công bố đặt cạnh nhau những con số không thật khớp với cách công ty tự mô tả về mình. Công ty nói rằng mình đang chuyển sang dịch vụ đám mây dựa trên thuê bao, nhưng phương pháp ghi nhận doanh thu năm 2025 cho thấy doanh thu được ghi nhận một lần tại một Point in time cụ thể vẫn chiếm đa số, ở mức RMB 4.04422 tỷ. Doanh thu được ghi nhận Over time là RMB 970.41 triệu. Khoản này đang tăng, nhưng vẫn chưa bằng một nửa. Cách ghi nhận doanh thu quan trọng vì nó là thước đo mức độ ổn định của doanh thu.

Doanh thu được ghi nhận một lần có thể phình lên trong năm một hợp đồng lớn được ký xong, rồi biến mất vào năm sau. Doanh thu được ghi nhận theo thời gian đi vào đều đặn mỗi năm, miễn là khách hàng còn giữ hợp đồng. Việc chuyển sang mô hình thuê bao là một hướng đi, chưa phải trạng thái đã hoàn tất. Phần công bố đặt nơi công ty muốn đến và nơi công ty đang đứng vào cùng một bảng. Một nửa công việc khi đọc tài liệu của một công ty niêm yết là đo khoảng cách giữa tương lai mà công ty phác ra ở những trang đầu và hiện tại được ghi lại trong các ghi chú kế toán.

Khoản lỗ thu hẹp không chỉ đến từ doanh thu. Công ty đã cắt giảm chi phí. Trong khi doanh thu tăng 33 percent, chi phí nghiên cứu và phát triển (R&D) giảm 8.6 percent, từ RMB 4.13188 billion xuống RMB 3.77535 billion. Chi phí bán hàng giảm từ RMB 654.69 million xuống RMB 569.20 million, còn chi phí hành chính giảm từ RMB 1.46359 billion xuống RMB 1.22575 billion.

Khi doanh thu tăng, chi phí bán hàng thường cũng tăng theo. Công ty này mở rộng doanh thu trong khi cắt giảm cả ba nhóm chi phí. Mức giảm đó thể hiện trực tiếp trong dòng tiền. Tiền thuần dùng cho hoạt động kinh doanh giảm từ RMB 3.92672 billion năm 2024 xuống RMB 311.42 million năm 2025. Lượng tiền công ty chi ra trong quá trình vận hành kinh doanh đã giảm gần mười lần.

Dù vậy, vẫn còn một câu hỏi chưa có lời đáp: công ty có thể tiếp tục tăng doanh thu trong khi cắt chi R&D hay không. Doanh thu năm nay đến từ các mô hình đã được phát triển từ trước, còn doanh thu năm sau sẽ chịu tác động từ phần chi phí phát triển đã bị giảm trong năm nay. Với một công ty AI, chi R&D là hạt giống mua lấy doanh thu tương lai. Cắt hạt giống đó để bù lỗ năm nay có thể ăn vào tăng trưởng của năm sau. Bản công bố chỉ ghi nhận sự cải thiện của năm nay. Nó không nói mức cải thiện đó có thể phải trả giá thế nào vào năm sau. Khi con số đẹp hơn xuất hiện mà không đi kèm rủi ro của phương pháp tạo ra nó, người đọc phải tự lấp khoảng trống đó khi đọc hồ sơ.

Công ty cũng cắt chi phí bằng cách đưa một mảng kinh doanh ra khỏi tập đoàn. Trong phần phân tích doanh thu theo bộ phận của báo cáo, một lý do được nêu cho sự sụt giảm doanh thu ở một bộ phận cụ thể là việc Deconsolidation of Autonomous Driving business vào August 2025. Lái xe tự hành tiêu tốn tiền mặt vì nhu cầu nghiên cứu và phát triển của nó không dừng lại. Khi mảng kinh doanh đó được đưa ra khỏi báo cáo tài chính hợp nhất của tập đoàn, các chi phí từng tạo ra lỗ cũng rời khỏi sổ sách của tập đoàn. Một phần con số cho thấy khoản lỗ giảm một nửa đến từ sự tách bạch kế toán này. Chỉ từ câu ngắn đó, khó có thể nói công ty thật sự đã đóng cửa mảng kinh doanh này hay chỉ giảm tỷ lệ sở hữu đủ để loại nó khỏi phạm vi hợp nhất.

Đây là điểm người đọc cần thận trọng. Sẽ là nói quá nếu coi việc lỗ ròng giảm từ 4,3 tỷ nhân dân tệ xuống 1,7 tỷ nhân dân tệ là bằng chứng cho thấy mảng kinh doanh cốt lõi của công ty đã cải thiện đúng bằng mức đó. Phần công bố tóm tắt không cho phép chúng ta chia tách chính xác bao nhiêu phần cải thiện đến từ hoạt động cốt lõi, bao nhiêu phần đến từ điều chỉnh kế toán khi tách một mảng kinh doanh thua lỗ ra khỏi sổ sách. Điều bản công bố cho thấy rõ là công ty đã chuyển vốn sang những lĩnh vực gần hơn với khả năng sinh lời và loại những mảng hút tiền mặt ra khỏi tài khoản của mình. Hướng đi ấy là rõ ràng, nhưng gán cho từng yếu tố một tỷ trọng số học cố định thì vượt quá những gì tài liệu này cho phép.

Đây là cấu trúc cơ bản của sự chuyển hướng kinh doanh được mô tả trong công bố doanh thu. Tuy vậy, công ty đặt một câu chuyện công nghệ bên cạnh câu chuyện tài chính. Thư của chủ tịch nói rằng AI tạo sinh đang chuyển từ thị giác sang không gian. Nếu đi theo chính ngôn ngữ của công ty về lý do của sự chuyển hướng này, ta sẽ thấy vì sao một công ty niêm yết nay đưa thuật ngữ Spatial Intelligence lên phía trước. Nếu báo cáo tài chính ghi lại những gì đã xảy ra, thì câu chuyện kỹ thuật trong bức thư chứa câu chuyện tiếp theo mà công ty muốn bán cho nhà đầu tư. Đọc cùng nhau, ngày hôm qua và ngày mai của công ty hiện lên trên cùng một màn

hình.

Logic của bức thư bắt đầu bằng việc xem xét cấu trúc của các mô hình đa phương thức hiện có. Công ty nói cách tiếp cận được dùng rộng rãi là một trung tâm encoder-connector-LLM: một cấu trúc ghép nối xử lý thông tin thị giác riêng rồi gắn nó vào một mô hình ngôn ngữ lớn được xây cho văn bản. Công ty viết rằng cấu trúc phân đoạn này có giới hạn rõ ràng trong việc vượt qua trần của trí thông minh. Nó có thể nhìn một hình ảnh và trả lời bằng lời, nhưng còn thiếu hụt trong việc hiểu không gian ba chiều và đưa ra phán đoán bên trong không gian đó. Chẩn đoán này không chỉ riêng SenseTime mới đưa ra.

Các báo cáo kỹ thuật từ những công ty khác cũng chỉ ra cùng một giới hạn: mô hình nhận nhiều phương thức đầu vào, nhưng đầu ra vẫn lấy văn bản làm trung tâm. Họ giải thích rằng phương pháp late-fusion, tức thêm một bộ giải mã chỉ dành cho thị giác ở phía sau, cản trở sự tích hợp sâu giữa các phương thức và buộc phải đánh đổi giữa năng lực tạo sinh và khả năng hiểu văn bản. Khi nhiều công ty mô tả cùng một nút thắt bằng cùng một ngôn ngữ, sự đối hướng này có thể được đọc như một bức tường kỹ thuật mà cả ngành cùng va phải, chứ không phải khẩu hiệu tiếp thị của một công ty.

Câu trả lời mà SenseTime đề xuất xuất hiện trong các tài liệu công bố của công ty dưới hai tên gọi. Một là SenseNova-SI, mô hình trí tuệ không gian được phát hành vào tháng 11 năm 2025. Tên còn lại là NEO, một kiến trúc đa phương thức gốc được phát hành dưới dạng mã nguồn mở vào tháng 12 năm 2025. Công ty cho biết NEO đã xây dựng lại thiết kế từ các nguyên lý nền tảng, với native embeddings, native RoPE và native multi-head attention, nhằm kết hợp thông tin văn bản và hình ảnh ở tầng lõi. Công ty cũng viết rằng cách này đã tích hợp về căn bản trí tuệ logic và trí tuệ không gian.

Công ty cũng viết rằng kiến trúc này đạt hiệu năng hàng đầu với một phần mười dữ liệu huấn luyện và năng lực tính toán mà các mô hình cạnh tranh cần đến. Việc nó được phát hành dưới dạng mã nguồn mở gắn với xu hướng mã nguồn mở trong AI Trung Quốc đã được bàn ở các chương trước. Với một công ty có nguồn lực hạn chế, mở kiến trúc để thu hút một hệ sinh thái nhà phát triển là lựa chọn mở ra một con đường khác với các mô hình quy mô lớn đóng.

Những câu này là tuyên bố của công ty trong tài liệu công bố, không phải sự thật đã được kiểm chứng độc lập. Báo cáo cũng nói rằng SenseNova-SI vượt Google Gemini-3-Pro trong một đánh giá tổng hợp trên nhiều chỉ số trí tuệ không gian. Thứ hạng benchmark có thể thay đổi tùy theo hạng mục nào được đo trong điều kiện nào, và các chỉ số đo trí tuệ không gian vẫn chưa ổn định thành chuẩn chung của ngành. Vì vậy, cách đọc an toàn hơn là coi đây là sự kiện rằng 'SenseTime đã công bố điều này trong báo cáo của chính mình.' Ranh giới mà văn bản này cố giữ là dùng lời kể của chính công ty làm bằng chứng cho hoạt động doanh nghiệp, nhưng không dùng nó làm căn cứ để phán xét hiệu năng tuyệt đối của mô hình.

Điều tương tự cũng áp dụng cho con số 'một phần mười.' Một so sánh không nói mô hình cạnh tranh nào được dùng làm mốc, hoặc tác vụ nào được đo, chỉ cho thấy hướng mà công ty muốn nhấn mạnh. Nó không chứng minh hiệu năng ngang bằng. Một tuyên bố xếp hạng chắc chắn rằng các mô hình Trung Quốc đã bắt kịp hoặc vượt các mô hình hàng đầu của Hoa Kỳ là quá sớm ở giai đoạn này, khi ngay cả các chỉ số đo diffusion cũng chưa được chuẩn hóa.

Dù vậy, việc công ty chọn câu chuyện về hiệu quả thay vì cuộc đua quy mô, rồi đặt câu chuyện đó ngay phần mở đầu tài liệu niêm yết, tự thân đã là sự kiện mà chương này xem xét. Cuộc đua quy mô kiểu Mỹ có nghĩa là đổ thêm dữ liệu và thêm GPU. Nhưng các công ty Trung Quốc phải đối mặt với các biện pháp kiểm soát xuất khẩu, khiến họ không tiếp cận được những chất bán dẫn mới nhất. Với một công ty buộc phải tạo ra hiệu năng trong điều kiện thiếu tài nguyên, câu chuyện 'một phần mười' vừa là chiến lược tài chính, vừa phản ánh vị thế chính trị của họ.

Nếu công ty có thể mua bao nhiêu chất bán dẫn tiên tiến tùy ý, họ sẽ ít có lý do hơn để đặt hiệu quả lên đầu câu đầu tiên. Một biện pháp cấp chính phủ, tức kiểm soát xuất khẩu của Hoa Kỳ, đã làm thay đổi cả câu chuyện công nghệ mà các công ty Trung Quốc chọn đưa ra trước công chúng. Đây là một cảnh cho thấy chính phủ và doanh nghiệp gắn vào nhau xuyên biên giới. Sự gắn kết đó cũng chạm tới bức tranh lớn hơn trong so sánh của Samjong KPMG về năng lực cạnh tranh AI của Hoa Kỳ và Trung Quốc, cùng các tài liệu phân tích ngành của tổ chức này.

Có những lý do thị trường khiến trí tuệ không gian được đưa lên phía trước. Công ty cho biết SenseNova-SI được xây dựng cho các ứng dụng như trí tuệ hiện thân và lái xe tự hành. Khi AI chỉ nằm trong trình duyệt web, một con mắt có thể đọc hình ảnh là đủ. Nhưng khi robot đi trong nhà máy và nhặt đồ vật, hoặc khi xe chạy trên đường, nó phải hiểu độ sâu và bối cảnh không gian. Điểm đáng chú ý là việc loại mảng kinh doanh lái xe tự hành khỏi phạm vi hợp nhất, đã bàn ở phần trước, và câu chuyện về trí tuệ không gian này lại xuất hiện trong cùng một báo cáo. Công ty đang lùi khỏi việc trực tiếp vận hành mảng kinh doanh lái xe tự hành, trong khi trình bày mô hình trí tuệ không gian hỗ trợ lái xe tự hành như một hướng đi cốt lõi.

Cách sắp xếp này là đưa khoản lỗ của một mảng dịch vụ vật lý ra khỏi sổ sách, trong khi bán một mô hình hiểu được thế giới vật lý. Thay vì trực tiếp sản xuất hoặc vận hành xe, công ty đang chuyển sang vị trí cung cấp trí tuệ khiến xe có thể di chuyển. Công ty cũng phát tín hiệu rằng trong quý II năm 2026, họ sẽ phát hành mô hình kiến trúc NEO thế hệ thứ hai để các agent có thể xử lý các nhiệm vụ trong kịch bản phức tạp hiệu quả hơn. Dự báo đó có chuyển thành doanh thu hay không là câu hỏi dành cho các công bố của năm tài chính tiếp theo. Khoảng cách giữa định hướng và kết quả là dấu hỏi luôn phải giữ lại khi đọc về một công ty niêm yết.

Bây giờ, góc nhìn mở rộng từ SenseTime như một công ty riêng lẻ sang thực tế tài chính rộng hơn của các công ty AI Trung Quốc, như được hé lộ qua các báo cáo công bố của doanh nghiệp niêm yết. Bức tranh ở đây không chỉ có một màu. Ngay cả trong cùng một ngành, sức bền tài chính cũng khác nhau rất rõ.

Ở một đầu là một công ty nền tảng khổng lồ, tạo ra dòng tiền từ hoạt động kinh doanh cốt lõi. Công bố năm 2025 của Tencent cho thấy tổng doanh thu tăng 14% so với năm trước, lên RMB 751.766 tỷ; lợi nhuận gộp tăng 21%, lên RMB 422.593 tỷ; biên lợi nhuận gộp tăng từ 47% lên 51%. Tencent cũng chi mạnh cho AI. Chỉ riêng năm 2025, công ty đã chi RMB 18 tỷ, khoảng \$2.5 tỷ, cho New AI Products; khoản tiền đó được dùng để tuyển nhân tài, huấn luyện mô hình, huấn luyện mô hình nền tảng Hunyuan của riêng Tencent, suy luận chatbot và marketing.

Chi phí bán hàng và marketing tăng 15%, lên RMB 41.727 tỷ, trong khi chi phí quản lý chung và hành chính tăng 21%, lên RMB 136.127 tỷ. Dù vậy, dòng tiền tự do năm 2025 của Tencent vẫn tăng 18%, lên RMB 182.6 tỷ, và riêng việc mua lại cổ phiếu đã tiêu tốn khoảng RMB 73.3 tỷ. Dù hóa đơn AI có lớn đến đâu, dòng tiền từ hoạt động kinh doanh cốt lõi vẫn lớn hơn nhiều, nên công ty có thể cùng lúc chịu được cuộc cạnh tranh hạ tầng và trả lợi ích cho cổ đông. Những cỗ máy tạo tiền cũ như game, quảng cáo và thanh toán bù đắp cho khoản lỗ từ đầu tư AI.

Ở đầu kia là SenseTime. SenseTime không có một cỗ máy tạo tiền khổng lồ như game hay quảng cáo. Vì vậy, cách công ty tiến tới lợi nhuận không đến từ cùng kiểu dư địa mà Tencent có, mà đến từ một cuộc tái cấu trúc cấp bách nhằm cắt chi phí và rút khỏi một số mảng kinh doanh. Và tiền để mở rộng trung tâm dữ liệu phải được huy động trực tiếp từ thị trường. Ngày 23 tháng 7 năm 2025, SenseTime ký Subscription and Placing Agreement với các bên thứ ba độc lập, gồm Infini Capital Management, và phát hành 1,666,667,000 cổ phiếu Class B mới với giá HK\$1.50 mỗi cổ phiếu.

Số tiền thu ròng huy động theo cách này vào khoảng 3,146 tỷ HKD. Công ty cho biết sẽ phân bổ số tiền đó như sau: 30 phần trăm để mở rộng hạ tầng AIDC và dùng chip sản xuất tại Trung Quốc, 30 phần trăm để thương mại hóa AI tạo sinh và các mô hình lớn đa phương thức, 20 phần trăm cho đổi mới kinh doanh, và 20 phần trăm cho vốn lưu động chung.

Bảng phân bổ đó cô đọng vị thế của công ty. Vì không thể xây trung tâm dữ liệu bằng tiền mặt của chính mình, công ty phát hành cổ phiếu để huy động vốn, rồi cam kết hơn một nửa số tiền đó cho hạ tầng và thương mại hóa mô hình, trên thực tế là mua trước doanh thu của năm sau. Việc một công ty đã công bố có lãi EBITDA trong nửa cuối năm nhưng cũng phát hành cổ phiếu mới trong cùng năm để huy động cả vốn lưu động cho thấy một lần nữa các điều kiện đứng sau chữ lợi nhuận hẹp đến mức nào. Phát hành cổ phiếu mới làm loãng tỷ lệ sở hữu của các cổ đông hiện hữu.

Việc công ty phải ra thị trường gọi vốn dù phải trả cái giá đó có nghĩa là tiền mặt do hoạt động kinh doanh cốt lõi tạo ra không đủ để duy trì cuộc đua hạ tầng. Trong cùng năm Tencent dùng tiền mặt dư thừa để mua lại cổ phiếu của chính mình và trả tiền lại cho cổ đông, SenseTime phát hành cổ phiếu mới và lấy tiền từ cổ đông. Hai công ty trong cùng ngành, cùng quốc gia, cùng niêm yết trên thị trường Hong Kong, lại đi theo hai hướng trái ngược về vốn.

Điểm yếu của công ty cũng hiện ra ở phía khách hàng. Trong ghi chú của báo cáo có tiêu đề "Information about major customers", một khách hàng riêng lẻ, Client A, chiếm 19,0 phần trăm tổng doanh thu của tập đoàn trong năm 2025. Năm 2024, Client B chiếm 12,2 phần trăm. Việc một số ít khách hàng lớn kiểm soát một phần năm doanh thu có thể có nghĩa là công ty đã có được người mua ổn định. Nó cũng có thể có nghĩa là nếu một khách hàng như vậy rời đi, doanh thu sẽ rung lắc mạnh.

Việc khách hàng lớn nhất theo tỷ trọng doanh thu chuyển từ Client B sang Client A khi năm thay đổi cũng là điểm đáng theo dõi. Nếu những khách hàng lớn nâng đỡ doanh thu thay đổi hằng năm, điều đó có thể cho thấy nền móng bên dưới khoản doanh thu ấy vẫn chưa thật vững. Hồ sơ niêm yết chỉ ghi nhận mặt hai lưỡi này bằng các con số, nhưng không công bố danh tính khách hàng. Bên mua đang mua sức mạnh tính toán trung tâm dữ liệu và các mô hình AI tạo sinh ở quy mô lớn là một nền tảng lớn trong Trung Quốc hay một đơn vị nhà nước, đó vẫn là câu hỏi nằm ngoài tài liệu.

Bên ngoài hai công ty này là các doanh nghiệp mới nổi chưa niêm yết hoặc chỉ vừa lên sàn. Báo cáo của Samjong KPMG so sánh năng lực cạnh tranh AI của Hoa Kỳ và Trung Quốc, cùng với các tài liệu phân tích ngành, mô tả những công ty này như đang đặt số phận vào IPO và vốn đầu tư bên ngoài trong khi thiếu dòng tiền vận hành. Vì số liệu về số tiền gọi vốn và định giá của các công ty mới nổi này thường khác nhau tùy bên công bố và thời điểm công bố, chương này sẽ không trình bày bất kỳ con số cụ thể nào như một sự thật đã chốt.

Nếu chỉ viết những gì có thể xác nhận, bản đồ vốn của các công ty AI Trung Quốc được chia thành một số ít gã khổng lồ có thể trụ lại bằng tiền mặt từ hoạt động kinh doanh cốt lõi và nhiều công ty phải vượt qua các giai đoạn thua lỗ bằng các đợt tiếp vốn từ thị trường vốn. SenseTime nằm đâu đó giữa hai nhóm này. Vì là công ty niêm yết, SenseTime có nghĩa vụ công bố thông tin, nên thực tế tài chính của họ xuất hiện trong tài liệu. Tuy vậy, tiền để mở rộng hạ tầng vẫn phải được huy động từ thị trường.

Đối với những công ty này, tư cách niêm yết có hai ý nghĩa. Một mặt, đó là nguồn vốn. Như trường hợp SenseTime phát hành cổ phiếu mới trên thị trường chứng khoán Hong Kong để trả chi phí trung tâm dữ liệu, niêm yết thường trở thành kênh giúp một công ty AI có khả năng tạo tiền mặt hạn chế tiếp tục cạnh tranh về hạ tầng. Mặt khác, đó là kỷ luật. Một công ty niêm yết phải công bố doanh thu và thua lỗ mỗi nửa năm, số liệu theo từng mảng kinh doanh và tỷ trọng của các khách hàng lớn. Ngay cả khi một công ty chọn các chỉ số có lợi và đưa chúng lên tiêu đề thông cáo báo chí, bảng số liệu bên cạnh vẫn ghi lại lỗ ròng, dòng tiền ra và mức độ tập trung khách hàng.

Khác với một kỳ lân chưa niêm yết, vốn chỉ có thể đưa các tỷ lệ tăng trưởng ra phía trước trong tài liệu gửi nhà đầu tư, một công ty niêm yết phải đặt chính những lời tuyên bố của mình bên cạnh các con số đã được kiểm toán. Nhờ nghĩa vụ công bố đó, chương này có thể đọc thực

tế tài chính của các công ty AI Trung Quốc qua SenseTime. Khác với tin tức chỉ xoay quanh những điểm benchmark hào nhoáng, hồ sơ công bố còn ghi lại chi phí và thua lỗ phát sinh để tạo ra những điểm số ấy. Niêm yết là cửa sổ để huy động tiền, đồng thời là bức tường kính khiến công ty không thể giấu mình.

Nhìn theo trục chính của cuốn sách này, tức sự đan cài giữa chính phủ, doanh nghiệp và dữ liệu, công bố của SenseTime lặng lẽ cho thấy một sự thật. Đó là dòng nói rằng 30 percent số vốn huy động được sẽ được phân bổ cho việc đưa chip do Trung Quốc sản xuất vào sử dụng. Điều này có nghĩa là các chất bán dẫn mà công ty dùng khi xây dựng trung tâm dữ liệu vừa là một lựa chọn tài chính của doanh nghiệp, vừa là vấn đề gắn với chính trị bên ngoài, kiểm soát xuất khẩu và chính sách nội bộ về tự lực điện toán của Trung Quốc. Những quyết định chuyển doanh thu sang trí tuệ không gian, đưa mảng lái xe tự động ra khỏi sổ sách, phát hành cổ phiếu mới để xây trung tâm dữ liệu, và đưa chip do Trung Quốc sản xuất vào các trung tâm dữ liệu đó không thể chỉ được giải thích bằng các phép tính bên trong công ty.

Nếu các biện pháp kiểm soát bán dẫn của Hoa Kỳ không tồn tại, bảng phân bổ vốn huy động của công ty có lẽ đã có hình dạng khác, và sức nặng của một câu chuyện công nghệ xoay quanh hiệu quả cũng đã khác. Một hồ sơ công bố của doanh nghiệp phản chiếu cấu trúc kiểm soát giữa các quốc gia và chính sách tự lực. Nếu chương trước cho thấy các văn kiện chính phủ ràng buộc tập dữ liệu và mô hình bằng ngôn ngữ của chính sách công nghiệp, thì chương này cho thấy sức nặng của chính sách ấy rơi xuống một dòng trong bảng phân bổ vốn của một công ty niêm yết.

Doanh thu hằng năm 5 billion yuan. Lợi nhuận EBITDA 380 million yuan trong nửa cuối năm. Khoản lỗ ròng hằng năm vẫn ở mức 1.7 billion yuan. Việc ba con số này xuất hiện cùng nhau trong cùng một hồ sơ công bố nói chính xác hơn bất kỳ câu nào khác về vị trí hiện nay của một công ty Trung Quốc trong lĩnh vực thị giác và trí tuệ không gian. Công ty công bố mức doanh thu cao nhất của mình, lần đầu tiên đạt lợi nhuận trong nửa năm, vẫn lỗ nếu tính cả năm, và phát hành cổ phiếu trên thị trường để mở rộng các trung tâm dữ liệu.

Liệu đợt phục hồi này có tiếp tục sang năm sau hay không, liệu câu chuyện mới về trí tuệ không gian có trở thành doanh thu thật hay không, và việc giảm chi tiêu cho nghiên cứu và phát triển sẽ buộc tăng trưởng của năm sau trả giá đến mức nào, tất cả vẫn còn là câu hỏi bỏ ngỏ. Câu trả lời sẽ hiện ra trong một bảng tóm tắt tài chính duy nhất mà công ty sẽ lại nộp cho Sở Giao dịch Chứng khoán Hong Kong vào năm sau. Đọc SenseTime qua trụ sở pháp lý và các địa chỉ của công ty

SenseTime (, SenseTime) có cấu trúc của SenseTime Group Inc., một công ty niêm yết tại Hong Kong. Trong báo cáo thường niên năm 2025, văn phòng đăng ký được ghi là địa chỉ Vistra Cayman tại Grand Pavilion, Hibiscus Way, 802 West Bay Road, Grand Cayman, Cayman Islands. Trụ sở chính và địa điểm kinh doanh chính tại Trung Quốc của công ty nằm ở số 1900 đường

Hongmei, quận Xuhui, Thượng Hải 200233 (1900).

Trụ sở chính và địa điểm kinh doanh chính tại Hong Kong của công ty nằm ở tầng 2 của Harbour View 1, số 12 Science Park East Avenue, Hong Kong Science Park, Sha Tin, Hong Kong (1212). Cách sắp đặt này cho thấy rõ tính chất của SenseTime. Pháp nhân dùng cấu trúc niêm yết ngoài khơi, trong khi các trụ sở chính của nghiên cứu và kinh doanh lại đặt ở Thượng Hải và Hong Kong. Nghiên cứu thị giác máy tính, khách hàng chính phủ và doanh nghiệp, cùng quyền tiếp cận thị trường vốn được phân bổ qua ba địa chỉ này.

Địa chỉ đường Hongmei ở Thượng Hải quan trọng hơn khi đọc hoạt động kinh doanh thực của công ty. SenseTime niêm yết tại Hong Kong, nhưng sức nặng của doanh thu và hạ tầng lại dựa nhiều vào hoạt động ở Trung Quốc đại lục. Trung tâm dữ liệu Shanghai Lingang, SenseCore (, SenseCore), các mô hình trí tuệ không gian và dịch vụ AI cho doanh nghiệp đã tạo ra những con số trong chương này. Địa chỉ Hong Kong Science Park giống như một cửa sổ cho hợp tác nghiên cứu quốc tế và quản trị công ty niêm yết, còn đăng ký tại Cayman là lớp vỏ doanh nghiệp thường thấy ở các công ty công nghệ Trung Quốc niêm yết tại Hong Kong. Khi tách ba địa chỉ này ra, SenseTime đứng trên một tam giác: niêm yết tại Hong Kong, vận hành ở Thượng Hải, và gắn với một mạng lưới nghiên cứu toàn cầu.

Tang Xiao'ou

Tang Xiao'ou () là người sáng lập SenseTime và là một nhân vật mang tính biểu tượng trong nghiên cứu thị giác máy tính của Trung Quốc. Ông làm giáo sư nhiều năm tại Chinese University of Hong Kong (), và được nhớ đến như người đã đưa nghiên cứu về nhận dạng khuôn mặt và hiểu hình ảnh đi vào công nghiệp. Cũng cần nói thêm rằng tên SenseTime bắt nguồn từ Shang Tang (), tên của một triều đại cổ ở Trung Quốc. Ngay từ đầu, công ty đã xuất hiện với cả một cái tên Trung Quốc và ngôn ngữ kỹ thuật của thị trường bài báo khoa học toàn cầu. Tang Xiao'ou đứng ở trung tâm của sự kết hợp ấy. Các bài báo, phòng thí nghiệm đại học, nghiên cứu sinh tiến sĩ, và vốn công nghệ từ Hong Kong cùng Trung Quốc đại lục đều nổi lại quanh ông. SenseTime nhấn mạnh thành quả nghiên cứu và đào tạo nhân tài từ những ngày đầu vì người sáng lập của công ty vốn là một giáo sư.

Công ty do ông xây dựng đã trở thành biểu tượng cho thời kỳ hoàng kim đầu tiên của AI Trung Quốc. Khi deep learning tạo ra những bước tiến lớn trong nhận dạng hình ảnh, nhận dạng khuôn mặt, giám sát đô thị, camera điện thoại thông minh, xác minh danh tính trong tài chính, và công nghệ cảm nhận cho xe tự lái cùng lúc bước vào thị trường. SenseTime đã đi rất đúng con sóng đó. Hạ tầng đô thị khổng lồ của Trung Quốc, nhu cầu an toàn công cộng, hệ sinh thái sản xuất điện thoại thông minh, và khách hàng là các công ty internet đều cần công nghệ thị giác máy tính. Các thuật toán bước ra từ phòng thí nghiệm của Tang Xiao'ou không nằm lại trong bài báo nghiên cứu. Chúng đi xuống camera, máy chủ, và hệ thống điều khiển đô thị. Đây là câu chuyện thành công của một nhà khoa học, đồng thời cũng là gương mặt nhạy cảm của

quá trình công nghiệp hóa AI kiểu Trung Quốc. Công nghệ nhận ra khuôn mặt mang theo cả sự tiện lợi lẫn vấn đề giám sát.

Di sản của Tang Xiao'ou có hai lớp. Một là nhân tài. SenseTime đã đào tạo nhiều nhà nghiên cứu, và một số người trong đó tỏa ra các startup AI khác cùng phòng thí nghiệm nghiên cứu của các công ty lớn. Những công ty do cựu nhân sự SenseTime sáng lập cũng xuất hiện, trong đó có MiniMax. Lớp còn lại là sức nặng của một công ty niêm yết. SenseTime bước vào thị trường Hong Kong và phải công bố số liệu cho nhà đầu tư. Một công ty bắt đầu từ phòng thí nghiệm đã trở thành một doanh nghiệp phải báo cáo doanh thu, lỗ, mức độ tập trung khách hàng, và chi phí nghiên cứu phát triển. Sự thay đổi này đem đến cho văn hóa lấy nghiên cứu làm trung tâm của Tang Xiao'ou cả điểm mạnh lẫn gánh nặng. Để một công ty mạnh về nghiên cứu cũng trở nên mạnh trên sổ sách, doanh thu lập lại quan trọng hơn số lượng bài báo. Áp lực đó cũng giải thích vì sao về sau SenseTime bắt đầu nhấn mạnh trí tuệ không gian, mô hình tạo sinh, và doanh thu từ trung tâm dữ liệu.

Ông qua đời vào năm 2023, nhưng cái bóng của ông vẫn còn phủ lên hướng đi của công ty. Ngay cả khi SenseTime nói về trí tuệ không gian, điểm xuất phát của họ vẫn là hiểu biết thị giác. Khả năng đọc video, không gian 3D, robot và thế giới xe tự lái là phần kéo dài của nghiên cứu thị giác máy tính. Điều Tang Xiao'ou để lại không phải là một sản phẩm nhận diện khuôn mặt cụ thể, mà là một cách nhìn dữ liệu thị giác như lối vào của trí tuệ. Cách nhìn đó không dễ biến mất, ngay cả trong thời đại các mô hình ngôn ngữ lớn. Nếu các mô hình văn bản đọc thế giới qua chữ, SenseTime vẫn cố đọc thế giới qua màn hình và không gian. Tên tuổi Tang Xiao'ou vẫn là điểm khởi đầu của hướng đi ấy.

Khi nói về Tang Xiao'ou, không thể tránh khỏi cái bóng của đạo đức. Thị giác máy tính là một lực đẩy đã làm ngành AI của Trung Quốc lớn lên, nhưng nó cũng kéo theo tranh luận về nhận diện khuôn mặt và giám sát đô thị. Những công nghệ do nghiên cứu của ông và công ty của ông tạo ra đã được dùng trong các lĩnh vực tiện ích như camera điện thoại thông minh, xe tự lái và kiểm tra công nghiệp, đồng thời cũng tiến sát đến an ninh công cộng và giám sát. Nếu xóa đi lưỡi dao hai mặt này, lịch sử của SenseTime cũng trở nên mờ nhạt. Ý nghĩa thật sự của Tang Xiao'ou không nằm nhiều ở chiến thắng thuần túy của công nghệ, mà nằm ở việc cho thấy những trách nhiệm nào xuất hiện khi nghiên cứu đi xuống các vùng nhạy cảm của xã hội.

Câu hỏi đó tiếp tục xuất hiện khi SenseTime trong thời đại mô hình lớn nói về trí tuệ không gian. Một mô hình hiểu không gian có thể giúp robot và ô tô, nhưng nó cũng có thể trở thành công cụ đọc thành phố và con người ở mức gần hơn. Vì thế, di sản của Tang Xiao'ou vừa mang tính kỹ thuật, vừa mang tính chính trị. SenseTime đang mang cả hai khuôn mặt của di sản ấy khi bước sang chương tiếp theo.

Tên tuổi ông cũng còn lại trong văn hóa nghiên cứu của SenseTime. Ngay cả sau khi công ty gặp khó khăn, thói quen tiếp tục đưa các bài báo và kết quả nghiên cứu lên phía trước vẫn

mang dấu ấn của nhà sáng lập. Thị trường đòi sổ sách, nhưng SenseTime vẫn cố giải thích chính mình bằng nghiên cứu. Sự bền bỉ này sẽ trở thành nguồn sức mạnh cho công ty hay trở thành gánh nặng, câu trả lời sẽ đến từ lần chuyển tiếp công nghệ kế tiếp.

Xu Li

Xu Li () là đồng sáng lập SenseTime, hiện giữ chức chủ tịch kiêm tổng giám đốc điều hành của công ty. Các tài liệu quản trị doanh nghiệp chính thức của công ty cho biết ông có nền tảng nghiên cứu về thị giác máy tính và chịu trách nhiệm về chiến lược, phát triển kinh doanh và hoạt động hằng ngày của tập đoàn. Nếu Tang Xiao'ou là biểu tượng của phòng thí nghiệm và thời kỳ thành lập công ty, thì Xu Li là người đang cố biến công ty đã niêm yết sau IPO thành một doanh nghiệp vận hành thực sự. SenseTime lớn lên nhờ nhận dạng khuôn mặt và trí tuệ thị giác đô thị, nhưng cùng lúc công ty phải đối mặt với quản lý nhà nước, thay đổi thị trường, lệnh trừng phạt của Hoa Kỳ và làn sóng chuyển sang mô hình lớn. Nhiệm vụ của Xu Li là phải quyết định lại công ty sẽ kiếm tiền bằng cách nào giữa biến động đó.

Tại buổi trình bày kết quả kinh doanh năm 2025 của SenseTime, Xu Li mô tả sự tích hợp sâu giữa ngôn ngữ và thị giác như một con đường mở rộng ranh giới của trí tuệ nhân tạo. Nhận xét ấy cô đọng sự đổi hướng của công ty. SenseTime nay muốn vượt ra khỏi nhãn hẹp là một công ty nhận dạng khuôn mặt và tự mô tả mình là doanh nghiệp trong các lĩnh vực mô hình lớn đa phương thức, trí tuệ không gian, tác nhân doanh nghiệp và hạ tầng trung tâm dữ liệu. Với SenseNova (), kiến trúc NEO và các mô hình trí tuệ không gian ở tuyến đầu, Xu Li đang viết 'chương tiếp theo của một công ty thị giác.'

Sự chuyển đổi này có thể trông tự nhiên, nhưng không dễ. Trước đây, khách hàng muốn những camera có thể nhận ra người và vật thật tốt. Khách hàng mới muốn mô hình, sức mạnh tính toán, tự động hóa quy trình làm việc, an toàn dữ liệu và chi phí thấp hơn. Công ty phải bán sản phẩm và hạ tầng cùng một lúc.

Cách Xu Li điều hành công ty đang bị thử thách ngay trong các con số kế toán. Năm 2025, SenseTime tăng doanh thu và đưa EBITDA điều chỉnh sang mức dương trong nửa cuối năm, nhưng cả năm vẫn lỗ ròng. Những con số này cho thấy lần ranh mà Xu Li đang đi trên đó. Nếu ông cắt chi tiêu cho nghiên cứu và phát triển, khoản lỗ sẽ thu hẹp, nhưng lần chuyển đổi kỹ thuật tiếp theo có thể yếu đi. Nếu ông tăng đầu tư vào trung tâm dữ liệu và mô hình, sản phẩm tương lai có thể mạnh hơn, nhưng khoản lỗ hiện tại sẽ phình ra. Tổng giám đốc của một công ty niêm yết phải giải thích cả hai điều đó với nhà đầu tư. Xu Li phải kể một câu chuyện công nghệ, đồng thời giải thích việc phát hành cổ phiếu mới, mức độ tập trung khách hàng, mua sắm chip sản xuất tại Trung Quốc, và cơ cấu doanh thu của trung tâm dữ liệu. Đó là vị trí mà một nhà quản lý xuất thân từ nghiên cứu cũng phải học ngôn ngữ của thị trường vốn.

Cách Xu Li phát triển AI của SenseTime là một bước chuyển 'từ thị giác sang không gian.' Công nghệ nhận diện khuôn mặt đọc một cảnh. Trí tuệ không gian cố đọc thế giới, thời gian, và chuyển động mà cảnh đó nằm trong đó. Robot, lái xe tự động, nhà máy thông minh, và tạo nội dung đều cần kiểu hiểu không gian này. Bước đi của SenseTime theo hướng đó không vứt bỏ di sản của người sáng lập. Nó mở rộng di sản ấy. Dù vậy, các con số đã công bố sẽ quyết định thành bại. Xu Li phải bảo vệ niềm tự hào công nghệ của công ty, đồng thời xây dựng một cấu trúc trong đó trung tâm dữ liệu và AI cho doanh nghiệp biến thành tiền thật. Đây cũng là lý do SenseTime lại hiện ra khác đi ở cuối chương này. Tên Xu Li nằm đúng nơi chuyển đổi kỹ thuật và sống còn tài chính xuất hiện trong cùng một bảng.

Thời gian còn lại của Shi Yi không giống những năm đầu của SenseTime. Một công ty trẻ có thể thuyết phục nhà đầu tư bằng ước mơ và tốc độ tăng trưởng. Sau khi niêm yết, công ty phải trả lời bằng con số. Năm 2025, SenseTime nói về đà phục hồi doanh số và EBITDA dương trong nửa cuối năm, nhưng cũng công bố khoản lỗ ròng cả năm và một đợt phát hành cổ phiếu mới. Shi Yi phải chịu trách nhiệm cho cả hai câu trong cùng một báo cáo. Những gì ông nói về trí tuệ không gian và mô hình đa phương thức không chỉ là nỗ lực chạy theo một xu hướng công nghệ.

Để công ty trông giống một công ty tăng trưởng trở lại, công ty phải cho nhà đầu tư thấy thị trường kế tiếp sau nhận diện khuôn mặt. Doanh thu trung tâm dữ liệu là một câu trả lời. Nhưng doanh thu hạ tầng kéo theo chi tiêu vốn và chi phí điện. Shi Yi phải giữ ba vai trò sát với chính mình: nhà nghiên cứu, người phụ trách bán hàng, và giám đốc tài chính. Lần công bố thông tin tiếp theo của SenseTime sẽ cho thấy bước ngoặt của ông chỉ là lời nói, hay là một thay đổi trong cấu trúc.

Để Shi Yi thành công, ngôn ngữ của SenseTime phải trở thành một khoản mục trong ngân sách của khách hàng. Trí tuệ không gian, NEO và đa phương thức là những thuật ngữ rõ ràng với kỹ sư, nhưng các nhà quản lý mua sắm cần chúng được chuyển thành chi phí và hiệu quả. Việc quản trị của Shi Yi phụ thuộc vào sự chuyển đổi đó. Công ty không cần câu: "Chúng tôi đã xây dựng một mô hình tốt." Công ty cần câu: "Khách hàng trả tiền cho chúng tôi hằng tháng."

Chương 8. Huawei Ascend, Quỹ Lớn và Đông Số Tây Toán trong nỗ lực tự chủ phần cứng

Vào mùa xuân năm 2026, nếu mở tài liệu bài phát biểu chính tại một sự kiện học thuật tổ chức ở Shanghai, một thuật ngữ lạ sẽ đập vào mắt trước tiên: "Tao Scaling." Đó là cái tên mới mà Huawei đưa ra vào thời điểm Moore's Law đã đi đến giới hạn. Bài trình bày nói rằng thay vì khắc transistor nhỏ hơn, Huawei sẽ giảm điện trở và độ trễ trên các đường mà tín hiệu đi qua bên trong chip. Họ đưa ra một phương pháp gọi là logic folding, tức xếp các mạch lên cao thành nhiều lớp thay vì trải chúng sang ngang. Rồi họ gắn vào đó một năm mục tiêu. Huawei nói rằng đến năm 2031, họ sẽ nâng mật độ của chip di động và chip AI Ascend lên mức tương đương 1.4 nanometer mà không cần thiết bị lithography cực tím cực xa, chỉ dùng thiết bị cực tím sâu đời cũ hơn.

Muốn hiểu bài trình bày này, trước hết phải nhìn vào tính chất của sân khấu nơi nó xuất hiện. Một bài phát biểu chính không phải là bàn phản biện học thuật dành cho các bài báo khoa học. Đó là nơi một công ty công bố hướng đi của mình. Những con số được viết ở đó có thể là kết quả đã đo được, cũng có thể là mục tiêu mà công ty định đạt tới. Khi hai loại con số này đứng cạnh nhau trên cùng một slide, người đọc thường nhằm mục tiêu thành kết quả. Vì vậy chương này sẽ thận trọng với từng con số khi bàn về lộ trình Ascend. Trước bức tường cấm vận của Hoa Kỳ, các con số Huawei đưa ra cho thấy sự bền bỉ của các kỹ sư Trung Quốc, đồng thời cũng cho thấy nỗ lực kỹ thuật nhằm đi vòng qua bức tường ấy vẫn chưa hoàn tất.

Vì vậy chương này bắt đầu từ các con số trong lộ trình Ascend, rồi kéo một đường thẳng từ những con số đó đến nguồn vốn nhà nước và lưới điện đang chống đỡ chúng. Khi đã biết cách đọc các con số trước, sau này ta có thể nhìn Big Fund và dự án East Data, West Computing bằng cùng một con mắt.

Khi đọc giả Hàn Quốc đọc câu này, từ đầu tiên cần nắm không phải là con số, mà là nhãn "equivalent." Tương đương mức 1.4 nanometer không có nghĩa là bề rộng đường mạch thật sự đã được làm mỏng xuống 1.4 nanometer. Cách nói đó gần với việc Huawei, bằng cách xếp chip thành nhiều lớp, chỉ đưa mật độ trên một đơn vị diện tích lên mức ấy. Khi bán dẫn được xếp cao lên, nhiệt sinh ra bên trong không dễ thoát ra ngoài, và mức tiêu thụ điện cũng tăng. Vì vậy nếu đọc bài trình bày này như bằng chứng rằng chip của Huawei đạt cùng hiệu năng trên mỗi watt như các chip 1.4 nanometer tiên tiến nhất của Hoa Kỳ, tấm bản đồ sẽ bị vẽ sai.

Đạt mật độ tương đương và đạt hiệu năng tương đương trên mỗi watt là hai bài toán khác nhau, và bài thuyết trình chỉ nói đến vấn đề thứ nhất, còn vấn đề thứ hai vẫn để ngỏ. Cách thận trọng hơn là nhớ rằng đây vẫn là một mục tiêu do một công ty đưa ra trên sân khấu công khai, chứ không phải một con số đã được kiểm chứng độc lập.

Cũng trong bài thuyết trình đó có câu rằng kết quả này đã được "xác minh bằng 381 chip sản xuất hàng loạt." Trong ngành bán dẫn, sản xuất hàng loạt thường chỉ trạng thái mà hàng trăm nghìn đến hàng chục triệu chip rời dây chuyền với tỷ lệ đạt ổn định, rồi chảy vào thị trường điện thoại thông minh hoặc các trung tâm dữ liệu lớn. Con số 381 trông không giống sản xuất thương mại ở quy mô đó, mà giống một đợt chạy thử nhỏ nhằm tìm lỗi thiết kế và đo tỷ lệ đạt hơn. Nó có thể là bằng chứng cho thấy công nghệ xếp chồng dọc hoạt động trong môi trường phòng thí nghiệm, nhưng đó không phải quy mô thích hợp để chứng minh rằng Huawei đã nắm được bài toán kinh tế nhằm đáp ứng nhu cầu của các công ty trên khắp thế giới mua hàng chục nghìn chip mỗi lần. Con số càng lấp lánh, càng phải kiểm tra trước nó được dùng trên sân khấu nào và ở đó nó có nghĩa gì.

Đây là cái bẫy người đọc cứ gặp lại khi đọc lộ trình của Huawei (, Huawei). Các chỉ số được công bố thường đặt chồng hai thứ lên nhau. Một là nỗ lực kỹ thuật nhằm đi vòng qua các giới hạn vật lý. Hai là cách diễn đạt nỗ lực ấy như thể nó đã hoàn tất. Thứ nhất là thật. Thứ hai có thể phồng lên. Chỉ khi đọc tách riêng hai lớp này, lộ trình mới không nghiêng sang phóng đại hoặc phủ nhận. Nếu theo tiếp câu chuyện Ascend thêm một đoạn, hai lớp ấy sẽ hiện rõ hơn.

Báo cáo của Samjong KPMG về năng lực cạnh tranh AI giữa Mỹ và Trung Quốc đưa ra một mô tả khá bình tĩnh về dòng chip Ascend. Ascend 910 đời đầu, ra mắt năm 2019, được sản xuất trên quy trình 7 nanometer của TSMC tại Đài Loan. Sau đó, các lệnh trừng phạt của Mỹ khiến việc sản xuất chuyển sang SMIC, một xưởng đúc chip của Trung Quốc, và con chip trở thành 910B. Ở thời điểm đó, nó đã đi qua một giai đoạn khó khăn khi hiệu năng thực tế còn lùi lại một chút. Các lệnh trừng phạt không lập tức nâng hiệu năng lên. Chúng từng đẩy nó lùi lại, và riêng sự thật ấy đã đáng để ghi nhớ.

Tự chủ không được vẽ như một đường thẳng đi lên trơn tru, mà như một đường cong có cả bước lùi. Và bước lùi này không chỉ là vấn đề của riêng Huawei. Đó là cú giảm mà cả ngành bán dẫn Trung Quốc cảm nhận, một ngành từng phụ thuộc vào Đài Loan, Hà Lan và Nhật Bản trong các quy trình tiên tiến. Cách nhà nước cố gắng bù vào khoảng hụt đó trở thành điểm khởi đầu cho câu chuyện sau này về Big Fund.

Ascend 910C, được công bố vào tháng 5 năm 2025, được mô tả là dùng quy trình 7 nanomet đã cải tiến của SMIC và đạt khoảng 60% hiệu năng của H100 của Nvidia. Nếu tách riêng con số 60% ấy, ta có thể tưởng như cuộc đua bắt gần kết thúc. Nhưng câu chuyện chỉ đầy đủ khi đọc cả chú thích trong báo cáo. 910C không phải là một bước nhảy được tạo ra bên trong một con chip duy nhất. Nó được xây bằng cách nối hai chip 910B vào một gói. Nó đạt 60% bằng cách tăng gấp đôi diện tích silicon. Khi diện tích và số lượng chip tăng, điện năng tiêu thụ cũng tăng theo, và các trung tâm dữ liệu vốn phải giành giật từng watt điện sẽ gánh thêm bấy nhiêu chi phí vận hành.

Vì vậy, cuốn sách này không nói rằng 910C ngang hàng với H100. Nó chỉ giữ lại đúng điều báo cáo nêu: hai chip được buộc lại với nhau để đi đến mức đầu đó quanh 60%. Nếu xóa đi sự thật rằng cùng một công việc đòi hỏi nhiều silicon hơn và nhiều điện hơn, con số 60% sẽ trông nhẹ hơn thực chất của nó. Nhìn ngược lại, điều đó cũng có nghĩa là một quốc gia có dư địa chi nhiều điện hơn và nhiều diện tích hơn có thể đáp ứng nhu cầu tính toán trước mắt theo cách này. Giới hạn và khả năng nằm trong cùng một câu.

Lùi xa hơn trong lộ trình, những con số ở xa hơn xuất hiện. Ascend 910D, dự kiến cho nửa cuối năm 2025, nhắm tới 320 teraflops thông qua đóng gói 3D. Ascend 920, mẫu chưa ấn định ngày phát hành, đưa ra băng thông bộ nhớ 4 terabyte mỗi giây và 900 teraflops, đồng thời ghi rõ quy trình SMIC 6 nanomet. Con số 6 nanomet ấy có sức nặng vì thiết bị quang khắc. Vì Hoa Kỳ, cùng các đồng minh như Hà Lan và Nhật Bản, đã chặn việc nhập khẩu thiết bị cực tím cực sâu, Huawei và SMIC phải vẽ độ rộng đường này bằng thiết bị cực tím sâu cũ hơn, với độ phân giải thấp hơn. Phương pháp là multiple patterning, in ánh sáng qua nhiều lớp phủ lặp lại, nhưng cách này làm tăng mạnh số bước quy trình, tăng lỗi và đẩy chi phí của từng con chip lên cao.

Ngay cả khi con số 900 teraflops là đúng, đó cũng chỉ là hiệu năng đỉnh của một con chip được chế tạo trong điều kiện lý tưởng. Liệu con chip ấy có thể được đưa ra thị trường với số lượng lớn và giá hợp lý hay không lại là một câu hỏi hoàn toàn khác. Dòng cao nhất trong bảng hiệu năng và bảng tỷ lệ thành phẩm của nhà máy không nằm trên cùng một tờ giấy. Mặt trước càng rực rỡ, ta càng phải nhìn kỹ sự im lặng ở phía sau. Vào thời điểm Hoa Kỳ đưa cùng lúc tới 140 công ty Trung Quốc vào danh sách cấm xuất khẩu bán dẫn tiên tiến, sự im lặng ấy không chỉ là vấn đề kỹ thuật. Nó còn là vấn đề địa chính trị.

Còn một điểm nữa. Nếu không đọc riêng huấn luyện và suy luận, ta sẽ hiểu sai toàn bộ hệ sinh thái này. Báo cáo nói mô hình R1 của DeepSeek đã chạy suy luận thành công trên Ascend 910C. Việc một mô hình đại diện cho Trung Quốc chạy trên chip do Trung Quốc chế tạo mang sức nặng biểu tượng. Nhưng chính báo cáo đó cũng ghi rằng quá trình huấn luyện ban đầu của R1 vẫn diễn ra trên Nvidia H100. Suy luận, tức việc chạy một mô hình đã hoàn tất và trả lời người dùng, có thể được xử lý ở một mức nào đó bằng năng lực tính toán của một chip đơn lẻ. Huấn luyện một mô hình khổng lồ thì khác. Hàng chục nghìn chip phải được buộc lại với nhau như một mạng thần kinh duy nhất và liên tục khớp các giá trị không ngừng. Trong môi trường ấy, tốc độ liên lạc giữa các chip quan trọng hơn hiệu năng của bất kỳ chip riêng lẻ nào.

Việc DeepSeek chỉ dùng Ascend cho suy luận nên được hiểu là dấu hiệu cho thấy năng lực tính toán cụm của Huawei vẫn chưa thể thay thế Nvidia trong việc huấn luyện một mô hình từ đầu. Ascend đã trở thành biểu tượng của sự thay thế bằng sản phẩm do Trung Quốc chế tạo, nhưng những giới hạn vật lý nằm dưới biểu tượng ấy vẫn tập trung ở ba nơi: nhiệt, tỷ lệ thành phẩm và nút nghẽn trong huấn luyện theo cụm. Đây là những giới hạn đã được xác nhận,

không phải lời tuyên bố rằng Trung Quốc sẽ không vượt qua được chúng trong tương lai. Điều đó có nghĩa là tài liệu hiện có chỉ cho ta biết đến mức này.

Các công ty AI của Trung Quốc vẫn tiếp tục tung ra các mô hình lớn bất chấp những giới hạn phần cứng rõ ràng này vì có một cấu trúc bên ngoài con chip đang đỡ cho họ. Nhà nước đi đầu trong việc cung cấp tiền và điện. Sự tương phản trở nên sắc nét khi đặt cạnh dòng vốn ở Hoa Kỳ. Chỉ riêng trong quý đầu năm 2025, các công ty AI của Hoa Kỳ đã thu hút 60,9 tỷ USD đầu tư nhân, trong khi đầu tư tư nhân vào AI của Trung Quốc đã thu hẹp xuống khoảng 5,2 tỷ USD tính đến năm 2024. Khi vốn tư nhân cạn đi, các công ty thường cắt giảm nghiên cứu hoặc đóng cửa. Đó là logic bình thường của thị trường.

Tuy nhiên, ở Trung Quốc, logic đó không vận hành nguyên dạng. Nhà nước trực tiếp bước vào nơi khoảng trống cần được lấp đầy. Nếu Hoa Kỳ thúc đẩy scaling law bằng hàng trăm tỷ đô la vốn tư nhân, thì Trung Quốc chịu cùng một sức ép ấy bằng hai trụ cột: vốn nhà nước và việc huy động hạ tầng trên quy mô lãnh thổ của mình. Vấn đề không phải bên nào đúng. Điều quan trọng là khi nguồn vốn thay đổi, cách chia sẻ rủi ro cũng thay đổi theo.

Ở trung tâm của chuyện này là cái gọi là Big Fund. Tháng 5 năm 2024, chính phủ Trung Quốc khởi động giai đoạn ba của National Integrated Circuit Industry Investment Fund với 19 nhà đầu tư, gồm Ministry of Finance và một số tổ chức tài chính. Quy mô của quỹ là 344 tỷ nhân dân tệ, khoảng 47,5 tỷ đô la Mỹ, và thời hạn hoạt động được đặt dài, đến năm 2039. Con số lớn này cần được đối chiếu với nguồn gốc, nhưng hướng đi thì rõ. Dòng tiền không chảy vào các dịch vụ AI đã hoàn thiện, mà chảy vào phần nền bên dưới chúng: foundries, thiết bị bán dẫn, mạch tích hợp tiên tiến và vật liệu.

Bức tranh hiện ra là nhà nước dùng tiền công để chống lưng cho những công nghệ có thể thay thế thiết bị lithography bị Hoa Kỳ chặn, cùng các chip AI do Trung Quốc sản xuất. Ở nơi thị trường đã đánh giá rủi ro quá cao rồi lùi lại, nhà nước bước vào với một lịch trình gần hai mươi năm và gieo vốn xuống. Điều đáng nhìn ở đây là quỹ này nhắm vào thiết bị và vật liệu nền tảng, không phải phần mềm hay dịch vụ. Đó là lựa chọn tài trợ cho bộ rễ của AI thay vì bề mặt của nó, vì chính bộ rễ ấy là nơi các lệnh trừng phạt của Hoa Kỳ ép mạnh nhất.

Không chỉ có quỹ này. Tháng 1 năm 2025, Bank of China nói rằng họ sẽ cung cấp hơn 1 nghìn tỷ nhân dân tệ tài chính cho lĩnh vực AI trong năm năm tới. Tháng 3, National Development and Reform Commission công bố một quỹ 1 nghìn tỷ nhân dân tệ, sẽ cung cấp vốn trong hai mươi năm cho các hệ sinh thái startup tiên tiến như AI, công nghệ lượng tử và pin hydro. Hai con số 1 nghìn tỷ nhân dân tệ này là những con số lớn, vẫn cần kiểm tra lại với văn bản gốc, nhưng hướng đi thì rõ: tiền nhà nước đang vươn vào cả hạ tầng lẫn startup cùng lúc. Điểm cần theo dõi ở đây là chính National Development and Reform Commission đang công bố lịch trình cho vốn startup.

Cơ quan này đặt ra hướng đi rộng cho cả nước. Nếu phạm vi của nó kéo dài từ thiết bị bán dẫn đến các công ty khởi nghiệp AI, điều đó có nghĩa là tự chủ phần cứng không được xem là việc riêng của từng doanh nghiệp, mà là một mục trong kế hoạch quốc gia. Trình tự hiện ra ở đây là: các cơ quan chính phủ vạch hướng đi, các định chế tài chính thuộc sở hữu nhà nước cung cấp tiền, rồi các công ty xây dựng mô hình trên nền đó. Chính trình tự này khiến việc nhìn AI Trung Quốc chỉ qua bảng xếp hạng doanh nghiệp sẽ bỏ sót một nửa bức tranh.

Số tiền này thực sự hỗ trợ điều gì có thể được suy ra từ tình trạng tài chính của các công ty mới nổi. Những doanh nghiệp thường được gọi là các AI Tigers, như Zhipu, do các nhà nghiên cứu từ Tsinghua University sáng lập, cùng với Moonshot, Baichuan và MiniMax, vẫn tiếp tục huấn luyện các mô hình lớn trong khi gánh những khoản lỗ nặng. MiniMax được cho là đã ghi nhận khoản lỗ ròng 1,98 tỷ yuan trong năm 2025. Trong một thị trường thông thường, những khoản lỗ như vậy sẽ khiến nhà đầu tư rút lui và buộc công ty phải thu hẹp. Nhưng khi vốn đầu tư mạo hiểm tư nhân đã cạn, có lý do mạnh mẽ xem khả năng chịu đựng các khoản lỗ như thế là một lá chắn tài chính do vốn nhà nước tạo ra. Vốn nhà nước đang chống đỡ giai đoạn mà thị trường chưa đem lại lợi nhuận.

Lá chắn này giúp các công ty sống sót lâu hơn, nhưng nó cũng là sợi dây ràng buộc khiến họ khó đi quá xa khỏi hướng đi của nhà nước. Nguồn vốn cũng là nguồn của kế hoạch. Cấu trúc này cũng phải được đọc từ phía ngược lại. Một cấu trúc trong đó vốn nhà nước bù lỗ làm tăng tỷ lệ sống sót của doanh nghiệp, nhưng làm yếu đi sức sàng lọc riêng của thị trường và mang theo rủi ro giữ những công ty không có kết quả thực chất tồn tại quá lâu. Phía sau cuộc đua mô hình hào nhoáng nhìn thấy từ bên ngoài là một cấu trúc tài chính trong đó kế hoạch đi trước lãi lỗ.

Ngay cả khi đã có tiền, một mô hình cũng không thể vận hành nếu thiếu điện và các tòa nhà cần cho tính toán. Vì vậy, Trung Quốc bắt đầu xem lãnh thổ rộng lớn của mình như một bài toán bố trí. Cái tên xuất hiện ở đây là Eastern Data, Western Computing, hay Dongshu Xisuan. Theo nghĩa đen, nó có nghĩa là đưa dữ liệu từ phía đông sang phía tây để tính toán. Dữ liệu tuôn ra từ bờ biển phía đông, nơi công nghiệp tập trung, trong khi điện và đất cần để xử lý dữ liệu lại tập trung ở vùng nội địa phía tây. Chính sách này nhằm thẳng vào sự mất cân bằng do địa lý tạo ra. Nó xem lãnh thổ khổng lồ của đất nước như một bảng mạch lớn, và việc tổ chức lại không gian ở quy mô này hiếm khi xảy ra nếu thiếu bàn tay của kế hoạch nhà nước.

Chính sách này vận hành bằng cách vẽ lại bản đồ các trung tâm dữ liệu. Ủy ban Cải cách và Phát triển Quốc gia đã chia cả nước thành tám trung tâm dữ liệu cấp quốc gia. Bốn trung tâm ở miền đông và miền trung là nơi dữ liệu được tạo ra và nhu cầu tập trung: Jing-Jin-Ji quanh Bắc Kinh, Đồng bằng sông Dương Tử bao gồm Thượng Hải, Khu vực Vịnh Lớn Quảng Đông-Hồng Kông-Macao nối các thành phố ở Quảng Đông với Hồng Kông và Macao, và Chengyu, lấy Trùng Khánh và Thành Đô làm trung tâm. Bốn trung tâm ở miền tây xử lý tính toán và lưu trữ: Quý Châu, Nội Mông, Cam Túc và Ninh Hạ. Bản thiết kế sau đó bổ sung mười cụm công nghiệp, lấy

Đồng bằng sông Dương Tử và Chengyu làm trọng tâm.

Thiết kế này nhằm phân tán hạ tầng công nghệ thông tin, vốn đã tập trung ở một số đô thị lớn, ra khắp lục địa và mở rộng khả năng tiếp cận tài nguyên tính toán giữa các vùng. Theo hướng đó, Cục Dữ liệu Quốc gia đã đặt việc xây dựng một mạng tài nguyên tính toán nối miền đông với miền tây, tức mạng suanli, thành nhiệm vụ cho những năm tới. Khi một cơ quan phụ trách dữ liệu nói về một mạng tính toán giống như lưới điện, trung tâm dữ liệu không còn được xem là phòng máy chủ của từng công ty riêng lẻ. Chúng trở thành các nút trên sơ đồ dây dẫn của nhà nước. Đặt tên cơ quan ấy cạnh ngôn ngữ của chính sách, ta thấy cách Trung Quốc sắp đặt vấn đề: dữ liệu, tính toán và điện được buộc vào cùng một câu.

Đằng sau quyết định chuyển trọng tâm của các trung tâm dữ liệu sang các sa mạc và cao nguyên miền tây là logic năng lượng. Việc huấn luyện và chạy suy luận trên các mô hình lớn liên tục tiêu thụ lượng nhiệt do hàng chục nghìn GPU phát ra và lượng điện cần để làm mát số nhiệt ấy. Một báo cáo rà soát của Hoa Kỳ chỉ ra lợi thế lạng lẽ của Trung Quốc ở đây. Bằng cách dùng nguồn điện mặt trời và điện gió dồi dào, chi phí thấp ở miền tây cùng khí hậu mát mẻ, Trung Quốc có thể giữ chỉ số hiệu quả sử dụng điện của các trung tâm dữ liệu ở mức thấp.

Báo cáo đó còn nói rằng lợi thế năng lượng này là một thứ vũ khí không dễ nổi bật khi so với Hoa Kỳ. Trong khi các trung tâm dữ liệu của Mỹ bị kìm lại bởi tình trạng thiếu điện, Trung Quốc đang mở ra con đường biến lượng điện dư thừa ở miền tây thành năng lực tính toán. Hoa Kỳ dẫn trước về bán dẫn, còn Trung Quốc có dư địa về điện. Phần này cho thấy hình ảnh hai quốc gia đang đối mặt với những nút thắt khác nhau.

Phân tích của Hana Securities về Two Sessions năm 2026 cũng chỉ về cùng một hướng. Báo cáo ghi nhận rằng chính phủ đã nêu các cụm tính toán thông minh siêu lớn và hạ tầng mới dựa trên phối hợp điện lực - tính toán là nhiệm vụ của năm đó. Điều này cho thấy ý định xem cơ sở tính toán không phải là thiết bị công nghệ thông tin, mà là hạ tầng xương sống quốc gia gắn liền với mạng lưới phát điện và truyền tải điện. Việc hạ tầng tính toán được nêu như một nhiệm vụ quốc gia tại Two Sessions cho thấy chính sách này nằm trên lịch trình chính trị cao nhất, chứ không nằm trong danh mục dự án của riêng một bộ nào.

Vì một startup AI riêng lẻ khó có thể tự xây dựng một cụm gồm hàng chục nghìn GPU, nhà nước xây các trung tâm tính toán khổng lồ và cung cấp chúng như một dạng hàng hóa công để doanh nghiệp tư nhân có thể dùng với chi phí thấp. Những quan sát cho rằng trục của kinh tế Trung Quốc năm 2026 vẫn nghiêng về công nghệ cũng trùng với dòng chảy này. Chính phủ nêu hướng đi, vốn nhà nước cung cấp tiền, và đất đai cung cấp điện. Ba tầng này tạo thành đúng cấu trúc đã thấy ở chương trước: doanh nghiệp chạy theo những đường kẻ do văn kiện chính phủ vạch ra.

Trên nền tảng đó, thị trường mở rộng nhanh. Theo dữ liệu nghiên cứu thị trường được trích trong một báo cáo của Korea Institute for International Economic Policy, thị trường dịch vụ

trung tâm dữ liệu của Trung Quốc năm 2022 tăng 12,7 phần trăm so với năm trước, đạt 129,35 tỷ yuan. Thị trường này được dự báo sẽ tăng với tốc độ trung bình hằng năm 18,9 phần trăm trong 5 năm tiếp theo và đạt 307,5 tỷ yuan vào năm 2027. Chính sách về bản đồ, rồi thị trường dựng các tòa nhà lên trên đó. Ở cấp công ty, hạ tầng cũng trở thành một chiến lược sinh tồn. Báo cáo năm 2025 của SenseTime cho thấy quy mô tính toán vận hành của SenseCore, lấy Shanghai Lingang Artificial Intelligence Data Center làm trung tâm, đã đạt 40.400 petaflops trên cơ sở FP16.

SenseTime nói rằng công ty đã theo kịp lời kêu gọi của nhà nước về điều phối năng lực tính toán và, với sự hỗ trợ từ công ty pin CATL, đã xây dựng một hệ thống tích hợp liên kết quản lý năng lực tính toán, vận hành trung tâm dữ liệu và hệ thống lưu trữ năng lượng. Phương pháp này dùng các mô hình lớn để đọc tải của lưới điện và giảm chi phí năng lượng. Điều đó cho thấy thay đổi không chỉ giới hạn ở việc chuyển công suất về phía tây. Những điều chỉnh nhỏ hơn cũng đang diễn ra bên trong các trung tâm dữ liệu hiện có ở miền đông Trung Quốc, nơi cách dùng điện đang được sắp xếp lại. Bản đồ chính sách rộng của nhà nước và các điều chỉnh tinh tại điểm đặt của doanh nghiệp đang đi cùng một hướng.

Các công bố của SenseTime cũng cho thấy rõ hạ tầng này làm thay đổi sổ sách của một công ty như thế nào. Công ty này từng là một doanh nghiệp thua lỗ, chi rất mạnh cho nghiên cứu thị giác máy tính và lái xe tự động, ghi nhận khoản lỗ trước thuế 17,1 tỷ nhân dân tệ vào năm 2021. Sau đó, bản chất của công ty thay đổi, lấy trung tâm dữ liệu Lingang, bắt đầu vận hành năm 2022, làm điểm tựa. Tổng doanh thu năm 2025 tăng 33 phần trăm lên 5,01464 tỷ nhân dân tệ, trong đó 1,85370 tỷ nhân dân tệ, tức 37,0 phần trăm, đến từ các công ty con trung tâm dữ liệu này. Đây là kết quả của một mô hình kinh doanh trong đó khách hàng thuê sức mạnh tính toán và mô hình theo thuê bao hoặc theo mức giá dùng bao nhiêu trả bấy nhiêu.

Với lực đẩy đó phía sau, SenseTime ghi nhận EBITDA điều chỉnh 380 triệu nhân dân tệ trong nửa cuối năm 2025 và lần đầu tiên có lãi kể từ khi niêm yết. Hạ tầng đã vượt ra ngoài vai trò công cụ nghiên cứu và trở thành nguồn doanh thu có thể nuôi chính công ty. Đó cũng là khoảnh khắc khi vốn do Big Fund gieo xuống và điện năng do dự án East Data, West Computing đặt nền gặp nhau trong báo cáo kết quả kinh doanh của một công ty, rồi hiện ra thành một dòng lợi nhuận duy nhất. Nghịch lý là một công ty phần mềm cuối cùng kiếm được tiền bằng cách bán hạ tầng, và điều đó cho thấy tài nguyên tính toán đã trở thành nguồn tiền mặt vững chắc đến mức nào trong ngành AI của Trung Quốc.

Chi tiêu của Big Tech cũng làm mạng lưới này dày thêm. Trong khi các công ty Big Tech của Hoa Kỳ đang đổ hơn 350 tỷ USD vào hạ tầng AI trong năm 2025, đầu tư của các công ty đám mây Trung Quốc vẫn dưới 40 tỷ USD, khiến họ tụt lại về quy mô tuyệt đối. Dù vậy, chi tiêu của từng công ty riêng lẻ đang tăng mạnh. Tencent đưa 10,7 tỷ USD, tương đương 12 phần trăm doanh thu, vào hạ tầng trong năm 2024. Alibaba nói họ sẽ đầu tư 380 tỷ nhân dân tệ, khoảng

52,6 tỷ USD, vào hạ tầng đám mây và AI trong ba năm tới. ByteDance đặt chi tiêu vốn cho năm 2025 ở mức 150 tỷ nhân dân tệ, khoảng 20,6 tỷ USD. Mỗi con số lớn này là một khoản được công bố và cần được đối chiếu với nguồn gốc, nhưng tất cả đều chỉ về cùng một hướng.

Để hướng dòng tiền này sang các mạng tính toán do Trung Quốc chế tạo thay vì Nvidia, Huawei đảm nhiệm nguồn cung phần cứng. Cùng với các chip Ascend, Huawei giới thiệu các cụm tính toán lớn cho trung tâm dữ liệu, gồm Atlas 950 SuperPod và TaiShan 950 SuperPod, tại MWC 2026, nhằm lấp khoảng trống do các sản phẩm Nvidia để lại. Chiến lược của Huawei hiện ra ở đây. Nếu không thể vượt Nvidia bằng hiệu năng của một con chip riêng lẻ, họ sẽ cạnh tranh ở cấp hệ thống bằng cách buộc nhiều chip lại với nhau. Không nên quên các giới hạn vật lý của Ascend đã được bàn trước đó. Mức hiệu năng 60 phần trăm có được khi ghép hai chip, tỷ lệ thành phẩm bị kéo xuống bởi multiple patterning, và các nút nghẽn của huấn luyện theo cụm đều đi vào các SuperPod này.

Khi một hệ thống tăng về quy mô, gánh nặng về điện năng và nhiệt lượng cũng tăng theo, vì vậy cái tên SuperPod không xóa được các giới hạn vật lý. Đây là lúc lợi thế điện rẻ ở miền tây trở nên quan trọng. Ngay cả khi một con chip riêng lẻ chưa đạt hiệu suất tốt tính theo mỗi watt, phép tính của toàn bộ hệ thống sẽ thay đổi khi điện vừa rẻ vừa dồi dào. Bằng cách đó, Big Fund và Eastern Data, Western Computing hoàn thiện một cấu trúc chống đỡ các điểm yếu của Ascend từ bên cạnh.

Nhưng thiết kế ở quy mô lục địa này vẫn có khoảng cách giữa tham vọng và thực tế. Một báo cáo của Hoa Kỳ phơi bày khoảng cách đó bằng các con số. Báo cáo ghi nhận rằng dù chính phủ hậu thuẫn toàn diện cho Eastern Data, Western Computing, mức sử dụng thực tế tại các trung tâm dữ liệu miền tây chỉ còn ở mức 20 đến 30 phần trăm. Nói thẳng ra, phần lớn các cơ sở được xây ở miền tây bằng lượng vốn khổng lồ đang bị bỏ trống. Mức sử dụng thấp này xuất phát từ bản chất của tính toán AI. Như báo cáo của Hana Securities đã chỉ ra, nhu cầu suy luận có khả năng tăng nhanh hơn nhu cầu huấn luyện mô hình. Khi một mô hình lớn được huấn luyện lần đầu, dữ liệu được đẩy vào cùng một lúc và hệ thống chạy trong thời gian dài, nên khoảng cách tới miền tây không gây nhiều rắc rối.

Nhưng trong suy luận, nơi người dùng đặt câu hỏi theo thời gian thực và cần câu trả lời ngay, độ trễ do gửi dữ liệu qua hàng nghìn kilômét giữa người dùng ở miền đông và trung tâm dữ liệu ở miền tây sẽ bào mòn chất lượng dịch vụ. Vì vậy các ngành cần xử lý theo thời gian thực, như lái xe tự hành và tài chính, vẫn chọn các trung tâm dữ liệu đặt ở miền đông, trong khi những cơ sở khổng lồ ở miền tây nhận bằng điểm chỉ 20 đến 30 phần trăm mức sử dụng, với nhiều khoảng trống còn lại. Bản đồ chính sách và dòng chảy dữ liệu thực tế vẫn chưa chồng khít lên nhau. Những trung tâm dữ liệu trống này lặng lẽ cho thấy bản đồ do nhà nước vẽ ra không phải lúc nào cũng đi trước tốc độ của ngành công nghiệp ngoài thực địa.

Giới chức Trung Quốc hiểu rõ vấn đề này. Cùng báo cáo đó mô tả các nỗ lực xây dựng dịch vụ đám mây để bán phần công suất chưa được dùng ở miền tây, còn báo cáo của Hana Securities kết luận rằng một số khu vực vẫn thiếu nguồn cung năng lực tính toán và cần phân bổ vùng tốt hơn. Công suất chưa dùng ở miền tây và công suất thiếu hụt ở miền đông vẫn chưa được kết nối trơn tru. Đó là vị trí hiện nay của Eastern Data, Western Computing. Trên giấy, bức tranh ghép lại rất gọn: dữ liệu thu thập ở Jing-Jin-Ji đi qua mạng lưới năng lực tính toán, tới các cụm Ascend chạy bằng điện gió giá rẻ ở Nội Mông và Ninh Hạ, rồi được rèn thành trí tuệ.

Thị trường cũng đã sẵn sàng tăng lên 307,5 tỷ nhân dân tệ vào năm 2027, và những công ty như SenseTime đã cho thấy cách tồn tại bằng việc gắn 40.000 petaflops hạ tầng với hệ thống lưu trữ pin. Nhưng trước nhu cầu suy luận thời gian thực, thứ phải di chuyển gần như với tốc độ ánh sáng, khoảng cách xuyên lục địa vẫn chưa được xóa bỏ hoàn toàn.

Nhìn theo cách đó, Huawei Ascend, Big Fund và Eastern Data, Western Computing không phải là ba chính sách riêng rẽ đứng tách nhau. Chúng khóa chặt với nhau thành một kết cấu chống đỡ. Vốn nhà nước chịu được cơn hạn của tài trợ mạo hiểm và bảo đảm ngân sách nghiên cứu cho các kỳ lân. Năng lượng rẻ ở miền tây cắt giảm chi phí vận hành trung tâm dữ liệu. Chip Huawei và chi tiêu của các hãng công nghệ lớn đặt nền cho một mạng lưới tính toán Trung Quốc nhằm thay thế Nvidia. Và như trường hợp SenseTime cho thấy, mô hình bán lại hạ tầng đó có thể đưa sổ sách của một công ty quay lại trạng thái có lãi.

Mạng lưới này không phải là sản phẩm từ tài năng của một công ty duy nhất. Nó là kết quả của nhiều lớp vốn và năng lượng, phần cứng và hạ tầng đám mây xếp chồng lên nhau. Khoảng cách giữa cơ quan nhà nước và doanh nghiệp càng hẹp, hệ thống di chuyển càng nhanh, và chính khoảng cách hẹp ấy cũng quyết định quy mô rủi ro. Khi vốn di chuyển theo kế hoạch, ngay cả những khoản đầu tư thất bại cũng có thể sống sót rất lâu. Khi bản đồ đi trước thực địa, các cơ sở trống vẫn nằm lại, như những trung tâm dữ liệu ở miền tây. Tốc độ và rủi ro là hai mặt của cùng một kết cấu.

Cùng lúc đó, mỗi khớp nối trong kết cấu chống đỡ này đều chứa những con số chưa được chứng minh và những nút thắt chưa được tháo gỡ. Mật độ được mô tả là tương đương cấp 1,4 nanometer, 381 lượt sản xuất thử, và lời hứa đạt 900 teraflops trên quy trình 6 nanometer multi-patterning đều là các mục tiêu nhằm che lấp sự thiếu vắng thiết bị quang khắc bằng xếp chồng dọc và đóng gói vật lý. Mức tăng hiệu năng 60 phần trăm nhờ ghép hai chip với nhau, cùng tỷ lệ vận hành 20 đến 30 phần trăm của các trung tâm dữ liệu ở miền tây, cho thấy từ tự lực vẫn chưa trọn vẹn. Nó vẫn đang diễn ra. Vậy chúng ta nên đọc bao nhiêu phần của sự tự lực phần cứng này như sự thật, và từ đâu phải bắt đầu đọc nó như một mục tiêu?

Chỉ khi không nhìn vào dòng đầu của bảng hiệu năng đã công bố, mà nhìn vào tỷ lệ thành phẩm của nhà máy, mức vận hành của trung tâm dữ liệu và nút thắt truyền thông tách biệt huấn luyện khỏi suy luận, chúng ta mới có thể đọc bản đồ phần cứng của Trung Quốc đúng như

nó vốn có. Trên bản đồ ấy, nhà nước và doanh nghiệp ngồi gần nhau đến mức nào, và khoảng cách đó sẽ rộng ra hay thu hẹp lại trong vài năm tới, sẽ quyết định kết quả của phép thử tự chủ này. Huawei và tự chủ phần cứng được đọc qua trụ sở doanh nghiệp và địa chỉ

Tên công ty chính thức của Huawei là Huawei Investment & Holding Co., Ltd. (). Trong thông tin chính thức của công ty, trụ sở được ghi là Huawei Base, Bantian, Longgang District, Shenzhen, Guangdong Province, China 518129 ().

Thông tin liên hệ cũng chỉ về cùng căn cứ trụ sở Huawei ấy. Địa chỉ này có ý nghĩa khi hiểu địa lý của quá trình tự chủ phần cứng AI của Trung Quốc. Shenzhen là thành phố nơi thiết bị viễn thông, điện thoại thông minh, máy chủ và chuỗi cung ứng sản xuất tụ lại trong cùng một nơi. Ascend (), Kunpeng () và các cụm Atlas không phải là tên mô hình sinh ra trong phòng thí nghiệm. Chúng là các dòng sản phẩm được xây trên lớp tích lũy lâu dài của hệ sinh thái sản xuất và viễn thông Shenzhen.

Cấu trúc doanh nghiệp của Huawei cũng khác với các công ty Big Tech khác. Hồ sơ công ty chính thức của Huawei nói rằng Huawei là công ty chưa niêm yết, có cấu trúc sở hữu cổ phần của nhân viên, và nhà sáng lập Ren Zhengfei () nắm dưới 1 phần trăm cổ phần. Cấu trúc này cho phép công ty thúc đẩy nghiên cứu và phát triển dài hạn với một mức tự do nhất định trước sức ép của thị trường vốn. Đồng thời, nó ở rất xa kiểu công bố minh bạch của công ty niêm yết mà nhà đầu tư bên ngoài thường thấy. Điểm này phải được đọc cùng với lộ trình chip AI của Huawei. Công ty này không phải là một tổ chức phải xoa dịu giá cổ phiếu theo từng quý như công ty niêm yết. Đây là một tổ chức gánh cùng lúc kỳ vọng về hạ tầng viễn thông quốc gia và tự chủ bán dẫn. Việc địa chỉ trụ sở của Huawei là một căn cứ rộng lớn ở Bantian, Shenzhen, cho thấy Huawei vẫn là một công ty công nghệ sản xuất, tức là “làm ra sản phẩm và kiểm soát chuỗi cung ứng.”

Ren Zhengfei

Ren Zhengfei () là người sáng lập Huawei. Trong số các nhà sáng lập công ty công nghệ Trung Quốc, ông có một xuất thân rất khác thường. Từng là sĩ quan kỹ thuật trong Quân Giải phóng Nhân dân, ông lập ra một công ty thiết bị viễn thông và biến Huawei, từ một doanh nghiệp nhỏ bán tổng đài điện thoại, thành một thế lực toàn cầu trong thiết bị viễn thông. Ngôn ngữ quản trị của ông luôn mang cảm giác sinh tồn, chiến đấu và chịu đựng qua một mùa đông dài. Cảm giác ấy trở nên rõ hơn sau khi Huawei đối mặt với các lệnh trừng phạt của Hoa Kỳ. Chuỗi cung ứng nội bộ dày hơn của công ty, trải từ điện thoại thông minh, thiết bị viễn thông, máy chủ, thiết kế chip, hệ điều hành đến dịch vụ đám mây, không phải là sự đa dạng hóa ngẫu nhiên. Từ lâu, Ren Zhengfei đã gieo vào bên trong tổ chức ý nghĩ rằng sự phụ thuộc vào nhà cung cấp bên ngoài có thể đe dọa sự sống còn của công ty.

Chiến lược phần cứng AI của Huawei cũng xuất phát từ văn hóa này. Những tuyên bố rằng chip Ascend vượt Nvidia chỉ trong một đòn là phóng đại. Nhưng nói rằng Huawei đang cố xây dựng từng bước một hệ sinh thái có thể thay thế được, mỗi lần một con chip, một máy chủ và một cụm máy, thì không hề phóng đại. Phương pháp của Ren Zhengfei không phải là chờ một sản phẩm hoàn hảo. Nó gần với việc đưa một sản phẩm chưa hoàn hảo ra hiện trường, sửa nó ngay tại đó, rồi sống sót.

Trong những năm làm thiết bị viễn thông, Huawei mở thị trường bằng giá thấp và hỗ trợ tại chỗ bền bỉ. Chiến lược chip AI của họ cũng có tính cách tương tự. Nếu hiệu năng chỉ đạt 60 phần trăm, họ nối nhiều chip lại với nhau. Nếu hiệu suất điện năng thấp, họ vẽ lại cách bố trí trung tâm dữ liệu và nguồn cung năng lượng. Nếu nguồn cung bị chặn, họ tiếp tục tiến lên với thiết kế của chính mình và các quy trình sản xuất của Trung Quốc. Cách quản trị sinh tồn của Ren Zhengfei thừa nhận khoảng cách công nghệ, nhưng vẫn tìm chỗ để xoay xở bên trong chính khoảng cách đó.

Sức mạnh của ông là sự kiên nhẫn ở cấp tổ chức. Tổng giám đốc của một công ty niêm yết không thể né tránh giá cổ phiếu và các buổi trao đổi với nhà đầu tư. Huawei có thể giữ một mặt trận dài vì công ty vẫn chưa niêm yết và có cơ cấu cổ phần do nhân viên nắm giữ. Ngay cả khi các lệnh trừng phạt của Hoa Kỳ làm chấn động nặng nề mảng điện thoại thông minh, công ty vẫn giữ đội ngũ nghiên cứu và phát triển, không bỏ các trụ cột nội bộ như HarmonyOS (), Kirin (), và Ascend. Ren Zhengfei biến sức chịu đựng này thành văn hóa tổ chức. Phần cứng AI không được quyết định chỉ trong vài quý. Nó là một chuỗi dài chạy qua thiết bị quang khắc, phần mềm thiết kế, đóng gói, bộ nhớ, mạng, làm mát, và điện năng. Ren Zhengfei xem toàn bộ chuỗi đó như vấn đề sống còn của quốc gia.

Điều đó không có nghĩa cách quản trị của ông có mọi câu trả lời. Sự kiên nhẫn kiểu quân đội và việc ra quyết định tập trung cho phép huy động nhanh, nhưng cũng mang rủi ro nghe tin hiệu thị trường quá muộn. Hệ sinh thái của Nvidia mạnh không chỉ nhờ hiệu năng chip, mà còn vì CUDA, thói quen của nhà phát triển, công cụ phần mềm, và các đối tác đám mây cùng dịch chuyển với nhau. Huawei dưới thời Ren Zhengfei mạnh về phần cứng và chuỗi cung ứng, nhưng gặp giới hạn trong việc để một hệ sinh thái nhà phát triển lan rộng tự do trên toàn cầu. Sự thiếu tin cậy do các lệnh trừng phạt của Hoa Kỳ tạo ra cũng rất đáng kể. Vì vậy, cuộc chiến của Ren Zhengfei là cuộc chiến xây chip, đồng thời là cuộc chiến thuyết phục một hệ sinh thái. Tự lực của Huawei không kết thúc bên trong nhà máy. Nó chỉ trở thành tự lực trên sổ sách khi nhà phát triển và khách hàng thật sự chọn Ascend.

Sự hiện diện của Ren Zhengfei càng mạnh hơn khi Huawei chịu sức ép. Các lệnh trừng phạt càng khắc nghiệt, câu chuyện sinh tồn của người sáng lập càng trở thành chất keo giữ tổ chức lại với nhau. Nhưng khi chất keo quá mạnh, một tổ chức có thể chậm hấp thụ những thay đổi bên ngoài bức tường của mình. Những rào cản Huawei phải vượt qua trong chip AI sẽ không

được hạ thấp chỉ bằng lòng yêu nước. Các vấn đề lạnh lùng vẫn còn đó: tỷ lệ thành phẩm của quy trình sản xuất, hệ sinh thái phần mềm, thói quen của nhà phát triển, hiệu suất điện năng, và nguồn cung bộ nhớ.

Sức mạnh của Ren Zhengfei là khả năng giúp công ty chịu đựng những vấn đề này trong thời gian dài. Điểm yếu của ông là sức chịu đựng ấy đôi khi có thể làm chậm phán quyết nghiệt ngã của thị trường. Dù vậy, khó có thể nói về tự lực của Trung Quốc trong phần cứng AI mà không nhắc tên ông. Ông đã biến Huawei thành một tổ chức “bị chặn thì tìm đường vòng, thiếu thứ gì thì tự xây, thiếu nguồn lực thì bám trụ.” Tương lai của Ascend nhiều khả năng sẽ được quyết định trên nền văn hóa này.

Bên trong Huawei, ông không chỉ là người lựa chọn công nghệ. Ông quản lý cảm xúc của tổ chức. Một công ty bị trừng phạt sẽ cùng lúc cảm thấy sợ hãi và giận dữ. Ren Zhengfei đã biến những cảm xúc đó thành sự bền bỉ trong nghiên cứu và phát triển. Dự đoán rằng Huawei sẽ không dễ lùi bước khỏi chip AI, dù trước mắt chưa thể thắng rõ ràng, xuất phát từ bản năng tổ chức này.

Xu Zhijun

Xu Zhijun (徐志军, Eric Xu) là một lãnh đạo nòng cốt, người đã nhiều năm tham gia cơ chế chủ tịch luân phiên của Huawei. Trên sân khấu Huawei Connect năm 2025, ông giải thích chi tiết lộ trình Ascend và các kế hoạch cho Atlas 950 và 960 SuperPoD cùng supercluster. Nếu Ren Zhengfei đại diện cho triết lý sinh tồn của Huawei, thì Xu Zhijun là người chuyển triết lý đó thành ngôn ngữ của sản phẩm hạ tầng AI. Ông không chỉ nói về thông số của một con chip riêng lẻ, mà nói về năng lực tính toán ở cấp hệ thống, cụm máy, mạng, điện năng và làm mát. Cách nói này có ý nghĩa. Nếu Trung Quốc chưa thể bắt kịp Nvidia ngay trong các quy trình tiên tiến, Huawei phải thu hẹp khoảng cách ở cấp hệ thống bằng cách kết nối nhiều chip và máy chủ với nhau. Xu Zhijun giải thích con đường đó trên sân khấu công khai.

Chiến lược Ascend mà ông trình bày vừa thực tế vừa mạnh tay. Nó thực tế vì thừa nhận khoảng cách tuyệt đối về hiệu năng của từng con chip và chọn con đường tính toán theo cụm. Nó mạnh tay vì đặt quy mô của cụm đó ở mức rất lớn. SuperPoD và supercluster tránh phép so sánh giữa một GPU Nvidia và một chip Huawei, thay vào đó khiến người ta nhìn cả trung tâm dữ liệu như một sản phẩm. Xu Zhijun dẫn dắt sự thay đổi khung nhìn này. Khi đơn vị cạnh tranh chuyển từ chip sang hệ thống, kinh nghiệm của Huawei trong thiết bị viễn thông, mạng và thiết kế máy chủ trở thành sức mạnh thật sự. Điều này cũng khớp với các chính sách của Trung Quốc về điện năng và trung tâm dữ liệu. Nếu không có dự án Eastern Data, Western Computing () và kế hoạch mạng lưới tính toán quốc gia, chiến lược cụm lớn này sẽ có ít nền tảng hơn nhiều để đứng vững.

Nhiệm vụ của Xu Zhijun là thu hẹp khoảng cách giữa những con số được công bố trên sân khấu và hiệu năng ngoài thực tế. Các con số trong lộ trình gửi tín hiệu tới thị trường. Nhưng khi đi vào triển khai, khách hàng hỏi những câu khác: việc huấn luyện mô hình có hoàn tất ổn định hay không, công cụ phát triển có khó dùng hay không, các kỹ sư đã quen với CUDA có thể chuyển sang nhanh đến đâu, và họ có kiểm soát được chi phí điện cùng nhiệt lượng hay không. Xu Zhijun phải trả lời những câu hỏi ấy bằng sản phẩm. Nếu Atlas SuperPod của Huawei không muốn chỉ dừng lại ở một màn công bố bóng bẩy trên sân khấu, các khách hàng như DeepSeek, Alibaba, Tencent và SenseTime phải tạo ra những trường hợp thật, cho thấy họ đã dùng nó cho huấn luyện và suy luận, đồng thời giảm được chi phí. Một chiến lược hệ thống chỉ có được niềm tin khi những trường hợp như vậy tích lũy dần.

Vai trò của ông trong AI Trung Quốc là chuyển sự tự chủ phần cứng thành những điều có thể dùng được trong thực tế. Văn kiện của chính phủ nói về tự chủ bán dẫn, mạng lưới tính toán, trung tâm dữ liệu, dự án Eastern Data, Western Computing và Big Fund, tức National Integrated Circuit Industry Investment Fund (). Khách hàng doanh nghiệp hỏi về giá máy chủ, thông lượng mô hình, độ ổn định và khả năng tương thích phần mềm. Xu Zhijun phải nối hai thứ ngôn ngữ này lại với nhau. Ông phải kéo định hướng rộng của nhà nước xuống thành lộ trình sản phẩm, đồng thời dùng hạ tầng và nguồn vốn quốc gia để bù cho các giới hạn của sản phẩm.

Nếu mối nối này thành công, Huawei sẽ trở thành nền tảng phần cứng cho AI Trung Quốc. Nếu thất bại, Ascend sẽ vẫn là biểu tượng của tự chủ, nhưng sẽ không bao giờ lấp đầy hẻm khoảng trống Nvidia để lại trong huấn luyện quy mô lớn ngoài thực tế. Những phát biểu trên sân khấu của Xu Zhijun đáng chú ý không phải vì lời lẽ lớn lao, mà vì phía sau chúng là cả một chuỗi nút thắt thật trong hạ tầng AI của Trung Quốc.

Cách Xu Zhijun giải thích chiến lược cho thấy rất rõ Huawei muốn cạnh tranh ở đâu. Ông không chỉ chỉ vào các con số của một con chip riêng lẻ. Ông nói về cụm máy, SuperPods, siêu cụm, mạng lưới và toàn bộ hệ thống. Cách tiếp cận này hợp với thực tế của Trung Quốc. Khi quyền tiếp cận công nghệ tiến trình tiên tiến bị chặn, tuyên bố rằng một con chip có thể vượt Nvidia không mấy thuyết phục. Thực tế hơn là nối nhiều chip lại với nhau, dựa vào kinh nghiệm của Huawei trong thiết bị viễn thông, nung vào trung tâm dữ liệu và lưới điện quốc gia, rồi đưa ra hiệu năng ở cấp độ hệ thống.

Xu Zhijun đóng gói thực tế đó thành một lộ trình sản phẩm. Gọi là đóng gói không phải để chê bai. Thị trường cần một ngôn ngữ có thể làm cho những nút thắt kỹ thuật phức tạp trở nên rõ ràng ngay từ cái nhìn đầu tiên. Xu Zhijun cung cấp ngôn ngữ đó. Điều còn thiếu lúc này là bằng chứng từ thực địa. Khi các công ty mô hình lớn của Trung Quốc xây dựng được những trường hợp cho thấy họ có thể vận hành huấn luyện và suy luận một cách đáng tin cậy trên các cụm Ascend, hạ chi phí và giảm ma sát cho nhà phát triển, lời nói của ông sẽ trở thành một

chiến lược.

Điều ông phải chứng minh từ đây không phải là quy mô của các tuyên bố, mà là độ ổn định trong vận hành. Một cụm lớn bật lên một lần thật ấn tượng là chưa đủ. Nó phải huấn luyện mô hình trong nhiều tháng, chịu được sự cố và trở nên quen thuộc với các nhà phát triển. Chỉ khi lộ trình của Xu Zhijun vượt qua bài kiểm tra vận hành dài hơi đó, nó mới trở thành bằng chứng cụ thể cho năng lực tự chủ phần cứng AI của Trung Quốc.

Chương 9. AI+ trong nền kinh tế thực: Chính sách bước vào sản xuất, giáo dục, tiêu dùng và robot hình người

Tháng 12 năm 2025, tên của bảy bộ xuất hiện cạnh nhau ở phần đầu của một văn bản. Tựa đề của văn bản là "Kế hoạch triển khai Kế hoạch hành động đặc biệt về Trí tuệ nhân tạo + Sản xuất." Việc không phải một bộ, mà là bảy bộ cùng đặt tên trên một văn bản đã cho thấy ngay từ đầu kế hoạch này có sức nặng đến mức nào. Một thông báo dự án từ một bộ thường chỉ mang theo một dòng ngân sách và tên của cục chịu trách nhiệm. Một văn bản do bảy bộ ký nghĩa là nhà nước muốn ấn định hướng đi của chính sách công nghiệp ở cấp quốc gia.

Cách trình bày, với Ministry of Industry and Information Technology giữ vai trò dẫn dắt và nhiều bộ khác đứng sau hỗ trợ, gần như là một tín hiệu rằng Trung Quốc muốn nâng trí tuệ nhân tạo từ một dự án kỹ thuật thuộc một bộ thành một nhiệm vụ quốc gia do nhiều bộ cùng gánh vác. Khi mở bản tóm tắt và bản dịch tiếng Hàn của văn bản này do Korea Institute for Advancement of Technology chuẩn bị, vài trang đầu bắt đầu bằng thứ ngôn ngữ chính sách quen thuộc.

Sự kết hợp sâu giữa nền kinh tế thực và trí tuệ nhân tạo, lực lượng sản xuất chất lượng mới, vai trò dẫn đầu trong hệ sinh thái sản xuất toàn cầu. Đến đó, đây vẫn là những khẩu hiệu có thể thấy trong chính sách công nghiệp của hầu như bất kỳ quốc gia nào. Nhưng ngay khi văn bản đi xuống tầng nhà máy, chất liệu của ngôn ngữ thay đổi. Những lời nói về chatbot trau chuốt câu chữ biến mất. Thay vào đó là các câu nói về việc dự báo nhiệt độ kim loại nóng chảy trong lò theo thời gian thực, đo lực căng của một bó sợi trên dây chuyền dệt ở cấp micromet, và điều chỉnh các quy trình gấn, lắp ráp, kiểm tra linh kiện điện tử trong mili giây. Việc một văn bản của chính phủ dùng kiểu ngôn ngữ này tự nó đã là một tín hiệu. Nó có nghĩa là Trung Quốc muốn coi trí tuệ nhân tạo không phải như một trợ lý văn phòng, mà như bàn tay điều khiển máy móc trên tầng nhà máy.

Chương này bắt đầu từ cảnh đó trong văn bản. Nó lần theo, theo thứ tự, nhiệm vụ mà chính phủ muốn giao cho trí tuệ nhân tạo, cách nhiệm vụ đó biến thành thị trường cho doanh nghiệp, cách các văn bản từ những bộ khác trong giáo dục, tiêu dùng và quản trị nhân sự cùng chỉ về một hướng, và các robot hình người cùng các mô hình ngành dọc nay đã đi đến đâu ở cuối con đường ấy. Chương này đọc các con số và chỉ dẫn chứa trong một văn bản, lần theo cách những con số ấy chuyển thành doanh thu của một số doanh nghiệp cụ thể, rồi cuối cùng nhìn vào các chi phí và rủi ro bị thiếu trong bức tranh đó. Mục tiêu của chương này là nhìn chính sách, doanh nghiệp và dữ liệu từ cùng một tầm mắt, để thấy chúng ăn khớp ở đâu và lệch nhau ở đâu.

Hãy bắt đầu từ những con số được nêu trong kế hoạch hành động về sản xuất. Văn kiện này không dừng ở những tuyên bố mơ hồ. Nó ấn định các mục tiêu định lượng phải đạt được vào

năm 2027. Mục tiêu là xây dựng ba hoặc bốn mô hình ngôn ngữ lớn công nghiệp bao phủ toàn bộ lĩnh vực sản xuất, đưa ra 1.000 tác nhân AI công nghiệp có thể tự phán đoán và điều khiển máy móc tại các cơ sở sản xuất, tạo 100 bộ dữ liệu công nghiệp chất lượng cao để nuôi máy học, và phát triển 500 kịch bản ứng dụng tiêu biểu để dùng trong nền kinh tế thực.

Đi kèm với đó là các kế hoạch nuôi dưỡng hai hoặc ba doanh nghiệp dẫn đầu đẳng cấp thế giới, hỗ trợ các doanh nghiệp nhỏ và vừa chuyên môn hóa, tinh gọn, khác biệt và đổi mới, đồng thời chọn 1.000 doanh nghiệp chuẩn để các công ty khác noi theo. Khi nhìn từng con số một, bản chất của chính sách hiện ra rõ hơn. Ba hoặc bốn mô hình cho thấy một số ít mô hình đại diện cấp quốc gia. 1.000 tác nhân và 100 bộ dữ liệu cho thấy lớp ứng dụng sẽ được trải dày bên dưới chúng, xuyên qua nhiều ngành khác nhau.

Trên thực tế, chính phủ đã vẽ sẵn một cấu trúc hai tầng vào văn kiện: phía trên là một số ít mô hình nền tảng, phía dưới là một khối lớn các tác nhân công nghiệp. Năm gắn với mục tiêu, 2027, cũng đáng để chú ý. Đây không phải là một lý tưởng đặt ra cho mười năm sau. Nó đặt hạn chót để có kết quả trong khoảng hai năm, và một hạn chót gần như vậy tạo sức ép lên doanh nghiệp. Mục tiêu 1.000 doanh nghiệp chuẩn cũng nổi bật. Ý tưởng ở đây không phải là đưa một hoặc hai công ty xuất sắc lên làm biểu tượng, mà là đặt xuống hàng nghìn mô hình để các công ty khác có thể đi theo từng bước.

Để hỗ trợ mục tiêu này, văn kiện chia công việc thành bảy nhiệm vụ. Nhiệm vụ đầu tiên là xây nền móng. Văn kiện kêu gọi chip huấn luyện tiên tiến và hệ điều hành đám mây điện toán thông minh, đồng thời yêu cầu các công ty lập cơ chế giám đốc dữ liệu trong chính tổ chức của mình. Việc một văn kiện của chính phủ đi xa đến mức yêu cầu doanh nghiệp cảm chuyên gia dữ liệu vào sơ đồ tổ chức cho thấy bản chất của kế hoạch này. Nhiệm vụ thứ hai là nâng cấp thông minh, trải rộng trên toàn bộ quy trình, từ nghiên cứu và phát triển đến thử nghiệm thí điểm, sản xuất, tiếp thị và vận hành. Nhiệm vụ thứ ba là đổi mới sản phẩm, bao gồm robot phẫu thuật, giao diện não-máy tính, xe tự hành, thiết bị đeo và dây chuyền sản xuất cho humanoid có trí tuệ hiện thân.

Nhiệm vụ thứ tư là nuôi dưỡng các chủ thể công nghiệp bằng cách phát triển các kỳ lân và các doanh nghiệp chuyên sâu, có sức cạnh tranh. Nhiệm vụ thứ năm là mở rộng hệ sinh thái bằng việc cung cấp tiêu chuẩn và nhân tài. Nhiệm vụ thứ sáu rất đáng chú ý. Văn kiện yêu cầu xây dựng cơ chế quản trị an toàn để phòng trước rủi ro ảo giác của trí tuệ nhân tạo, đồng thời phân loại và xếp cấp dữ liệu công nghiệp. Nói cách khác, cùng một văn kiện đưa trí tuệ nhân tạo vào nền kinh tế vật chất cũng ghi rõ cách nghi ngờ và giám sát chính trí tuệ nhân tạo đó. Nhiệm vụ thứ bảy là hợp tác quốc tế, với kế hoạch dẫn dắt chương trình nghị sự về trí tuệ nhân tạo trong BRICS, Tổ chức Hợp tác Thượng Hải, ASEAN, diễn đàn G20 và Hội nghị Trí tuệ Nhân tạo Thế giới.

Văn kiện đặt hai hướng đi cạnh nhau: tổ chức công nghiệp trong nước, và đứng vào vị trí xây dựng quy tắc ở nước ngoài. Khi bảy nhiệm vụ này được nối lại, chúng tạo thành một chuỗi bắt

đầu từ nền tảng như chip và dữ liệu, đi qua quy trình và sản phẩm, tiến lên doanh nghiệp, tiêu chuẩn và nhân tài, rồi kết thúc ở an toàn và sân khấu quốc tế. Cấu trúc của văn kiện nằm ở chỗ nó buộc chung công việc tạo ra trí tuệ nhân tạo, giám sát nó, và đưa nó ra bên ngoài vào trong một kế hoạch duy nhất.

Mục tiêu của cấu trúc này không phải là hội thoại, mà là các hiện tượng vật lý. Kế hoạch thực hiện có các hướng dẫn chuyển đổi riêng cho năm ngành, và mỗi ngành đòi hỏi một kiểu trí tuệ nhân tạo khác nhau. Trong ngành nguyên vật liệu, văn kiện yêu cầu các tác nhân phát hiện trạng thái vận hành của lò cao luyện thép theo thời gian thực và tối ưu hóa các biến số quy trình ngay tại hiện trường. Trong hóa dầu, văn kiện yêu cầu các mô hình kết hợp dữ liệu thăm dò với cơ chế quy trình hóa học để giám sát an toàn và đường ống tại các khu vực làm việc của mỏ dầu. Trong phát triển vật liệu mới, ngôn ngữ trở nên tham vọng hơn. Kế hoạch là xây dựng một khung tính toán đa tỷ lệ, phân tích đồng thời hiện tượng, dữ liệu và hệ thống trên các không gian và thời gian khác nhau đối với hợp kim, gốm và vật liệu polymer, đồng thời phát triển năng lực thiết kế ngược từ thành phần, cấu trúc và hiệu năng mong muốn.

Ý tưởng là đặt hiệu năng mong muốn lên trước, rồi đi xuống để tìm các tổ hợp phân tử phù hợp với hiệu năng đó. Trong chế tạo thiết bị, kế hoạch đưa các hệ thống điều khiển số thế hệ mới vào máy công cụ và robot công nghiệp, để cảm biến thời gian thực, học tự chủ, ra quyết định thông minh và kiểm tra trực tuyến vận hành như một dòng chảy thống nhất. Trong đóng tàu, kế hoạch làm cho các quy trình trở nên thông minh từ thiết kế đến cắt, hàn, sơn và logistics, đồng thời cải thiện hiệu suất năng lượng trong hàng hải. Trong hàng không vũ trụ, kế hoạch nhằm đến việc xây dựng các nền tảng mô phỏng kết nối dữ liệu cơ động với các mô hình động lực học chất lưu, tinh chỉnh đường nét thân máy bay và thiết kế mặt cắt cánh, rồi rút ngắn chu kỳ thử nghiệm vật lý. Cùng một cụm từ, trí tuệ nhân tạo, mang một gương mặt khác trong từng ngành.

Khi cuộc thảo luận đi xuống hàng tiêu dùng và ngành thông tin điện tử, các đơn vị đo độ chính xác trở nên nổi bật. Trong dệt may và trang phục, doanh nghiệp được yêu cầu phân tích dữ liệu người tiêu dùng để đọc xu hướng thời trang, rồi kết nối các công cụ vật lý mô phỏng chân thực hệ thống vật chất với mô hình sinh tạo ba chiều để cung cấp dịch vụ thử đồ ảo. Trên dây chuyền sản xuất, mục tiêu là sản xuất tự chủ, có thể giám sát lực căng của các bó sợi ở cấp micromet và tự sửa lỗi. Trong ngành thông tin điện tử, văn kiện yêu cầu phá bỏ các bức tường dữ liệu giữa công cụ tự động hóa thiết kế điện tử dùng trong thiết kế linh kiện và hệ thống quản lý vòng đời sản phẩm, dùng công nghệ biên để tái cấu hình dây chuyền sản xuất theo thời gian thực, đồng thời điều chỉnh việc gắn, lắp ráp và kiểm tra linh kiện điện tử tiêu dùng ở cấp mili giây.

Trong dịch vụ phần mềm và công nghệ thông tin, kế hoạch yêu cầu một chuỗi công cụ nối thiết kế, phát triển, kiểm thử và bảo trì thành một dòng công việc thống nhất. Kế hoạch yêu cầu

doanh nghiệp dùng dữ liệu tổng hợp và kiểm thử đối kháng để trải nghiệm trước các điều kiện cực đoan như lưu lượng truy cập tăng vọt, đồng thời thiết lập quy trình làm sạch nhằm lọc bỏ xung đột giấy phép và lỗ hổng trong mã nguồn mở.

Một văn kiện đưa ra chỉ dẫn cho năm ngành khác nhau bằng năm ngôn ngữ khác nhau, nhưng các ngôn ngữ ấy có chung một điểm: chúng kéo trí tuệ nhân tạo xuống tận tầng điều khiển vật lý của xưởng sản xuất. Trong khi phần lớn AI phương Tây vẫn nằm ở giá trị gia tăng trong phần mềm, như dịch thuật hoặc tạo video, văn kiện này nhắm vào nền móng của chuỗi giá trị. Việc các đơn vị như micromet và mili giây xuất hiện trong một văn kiện của chính phủ cho thấy chính sách này đang nhắm tới cấp độ nào.

Dù một chính sách được thiết kế cẩn thận đến đâu, nó vẫn chỉ nằm trên giấy nếu các doanh nghiệp tuyến đầu không hấp thụ được. Vì vậy, kế hoạch triển khai còn nêu rõ các bước mà doanh nghiệp nên làm theo. Trước hết, kế hoạch yêu cầu họ chẩn đoán mức độ chuyển đổi thông minh của chính mình thông qua quản lý dữ liệu và các mô hình trưởng thành về sản xuất thông minh, tăng số lượng cảm biến và thiết bị đo, rồi biến các trung tâm dữ liệu hiện có thành trung tâm tính toán thông minh. Sau đó, kế hoạch hướng dẫn họ đưa vào cấu trúc sinh tạo tăng cường bằng truy xuất, kết nối với cơ sở dữ liệu công nghiệp tại hiện trường hoặc nền tảng tin tức để nâng hiệu suất, xử lý tri thức thường dùng thông qua cấu trúc đó, và xây dựng thư viện prompt tập hợp các chỉ dẫn lặp lại được dùng trong toàn công ty.

Kế hoạch này cũng bao gồm một giai đoạn tinh chỉnh: các doanh nghiệp sẽ tiếp tục huấn luyện các mô hình đã được tiền huấn luyện bằng một lượng nhỏ dữ liệu lỗi trên dây chuyền sản xuất, để mô hình có thể nhận ra những đặc điểm rất nhỏ; đồng thời dùng dữ liệu chuỗi thời gian để điều chỉnh mô hình cho việc lập lịch sản xuất và chẩn đoán hỏng hóc thiết bị.

Tùy theo tính chất công việc, các mô hình đã hoàn thiện sẽ được gom về một nơi dưới dạng microservices hoặc middleware giao diện lập trình ứng dụng, hoặc được phân tán trên cloud, edge và các thiết bị đầu cuối. Một văn kiện của chính phủ đã đi xuống tới mức như một sổ tay thực hành cho doanh nghiệp. Chuỗi bước này, từ chẩn đoán đến củng cố hạ tầng, rồi qua retrieval-augmented generation và tinh chỉnh để đi tới triển khai, cho thấy chính sách công nghiệp đã được viết như một bản đồ quy trình, chứ không phải một khẩu hiệu. Lý do dòng chảy từ chẩn đoán đến triển khai này quan trọng là chính phủ không nêu tên mô hình nào doanh nghiệp bắt buộc phải dùng. Chính phủ nêu quy trình mà họ phải đi theo. Doanh nghiệp có thể chọn mô hình, nhưng nhà nước sẽ chuẩn hóa cách mô hình đó được gắn vào quy trình riêng của doanh nghiệp.

Đến đây, khoảng cách giữa chính sách và doanh nghiệp thu hẹp lại. Khi nhà nước đặt các mục tiêu cấp trên là một trăm bộ dữ liệu và một nghìn agent, các nhà cung cấp hạ tầng như Alibaba Cloud và Huawei cung cấp năng lực tính toán và mô hình ở bên dưới, còn từng nhà sản xuất ở tầng dưới nữa bổ sung retrieval-augmented generation và tinh chỉnh để khớp với quy trình sản

xuất riêng của mình. Đây là lý do các phân tích gắn với U.S.-China Economic and Security Review Commission thường đọc chiến lược trí tuệ nhân tạo mở của Trung Quốc cùng với tham vọng thống trị công nghiệp.

Mô hình càng được phát hành rộng rãi, các ứng dụng công nghiệp chạy trên đó càng nhiều. Khi những ứng dụng ấy lớn lên, chúng tạo ra một vòng phản hồi, trong đó dữ liệu hiện trường tích tụ và có thể được dùng để sửa lại mô hình một lần nữa. Cách phân tích này không xem việc mở các mô hình open-source chỉ như thiện chí vị tha. Nó đọc việc đó theo logic công nghiệp: một khi mô hình được phát hành, các doanh nghiệp trong và ngoài nước xây dựng ứng dụng trên mô hình ấy, và lớp ứng dụng đó đến lượt mình làm dày thêm nền móng cho các ngành công nghiệp của chính quốc gia đó.

Đây là nhận xét rằng mở và kiểm soát không đối lập nhau, mà trở thành hai mặt trước sau của cùng một chiến lược. Khi đặt nhận xét này lên cấu trúc hai tầng của kế hoạch hành động về sản xuất, bức tranh trở nên rõ hơn. Các mục tiêu của chính phủ tạo ra tầng trên, các mô hình mở lấp đầy tầng giữa, còn dữ liệu từ các doanh nghiệp tại hiện trường làm chắc tầng dưới trước khi lại lưu chuyển lên trên. Chính sách của chính phủ, mô hình của các công ty hạ tầng, và dữ liệu của các doanh nghiệp tuyến đầu khớp với nhau thành một vòng lặp. Một khi vòng lặp này bắt đầu quay, công ty phát hành mô hình sẽ mở rộng ảnh hưởng trên khắp các ngành nơi mô hình của họ được cài đặt, còn các ngành đó nhanh chóng thử nghiệm ứng dụng trên một mô hình mà họ có được với chi phí thấp. Cấu trúc này, trong đó mở trở thành con đường mở rộng thị trường, giải thích bằng ngôn ngữ chính sách công nghiệp vì sao các công ty Trung Quốc vội vã phát hành mô hình của mình.

Nếu sản xuất là câu chuyện về phía cung, thì giáo dục và tiêu dùng là những văn kiện mở cánh cửa ở phía cầu. Trong kế hoạch hành động 'Artificial Intelligence + Education' và các văn kiện chiến lược Artificial Intelligence Plus do State Council cùng các bộ liên quan ban hành, chính phủ ấn định việc đưa trí tuệ nhân tạo vào trường học như một mục tiêu quốc gia. Một câu tiêu biểu nói rằng họ sẽ phát triển các bạn đồng hành học tập thông minh và các mô hình giáo dục hợp tác giữa con người và máy móc có giáo viên tham gia. Khi nhìn vào cách một câu đó vận hành trên thị trường, sự ăn khớp giữa chính sách và doanh nghiệp lại xuất hiện. Nếu các bạn đồng hành học tập thông minh trở thành một phần tiêu chuẩn của giáo dục công, những công ty đáp ứng tiêu chuẩn đó trước sẽ có được kênh bán hàng trong cả lớp học lẫn gia đình.

Từ đây hiện ra hình ảnh của một công ty như iFlytek, vốn hoạt động trong cả nhận dạng giọng nói và giáo dục, cung cấp bảng đen thông minh cho trường học và bán thiết bị học tập cá nhân cho phụ huynh. Khi doanh nghiệp ngồi cùng nhà nước tại nơi các tiêu chuẩn được đặt ra, những tiêu chuẩn đó sớm trở thành đường vào thị trường cho chính các doanh nghiệp ấy. Thay vì phân bổ trực tiếp hỗ trợ tài chính cho học sinh, nhà nước tạo ra nhu cầu học tập thông qua chính sách. Các hộ gia đình sau đó trả chi phí để đáp ứng nhu cầu ấy, còn doanh nghiệp thu

doanh thu. Ngay cả khi chính sách giáo dục được viết bằng ngôn ngữ phúc lợi, nó vẫn tạo ra kết quả công nghiệp.

Có một điểm cần thận trọng ở đây. Phía sau chính sách giáo dục là nỗi lo xã hội rằng trí tuệ nhân tạo đe dọa việc làm của con người. Chính phủ thường xử lý nỗi lo này không phải chủ yếu bằng chi tiêu công, mà bằng cách chuyển nó thành tự phát triển dưới tên gọi học tập suốt đời. Thay vì mở rộng trợ cấp thất nghiệp, chính phủ hướng dẫn công dân học kỹ năng mới. Vào khoảng khắc đó, nhu cầu học tập trở thành nhu cầu thị trường trong edtech. Khi gánh nặng học tập chuyển sang từng cá nhân, các công ty tiếp nhận gánh nặng đó như một hoạt động kinh doanh sẽ lớn lên. Dù có gấn ngôn ngữ của mạng lưới an sinh, nếu bên trong không có tái phân phối tài chính thật sự, cuối cùng cá nhân vẫn tự trả giá cho mạng lưới an sinh ấy.

Ngay cả khi một chính sách khuyến khích học tập mang mục tiêu đáng giá là xây dựng năng lực cho công dân, vẫn còn một câu hỏi riêng: ví tiền nào sẽ trả chi phí để lấp đầy mục tiêu đó. Những con số cụ thể như doanh số thiết bị học tập hoặc số lượng được đưa vào trường học nằm ngoài các tài liệu nguồn được dùng trong chương này, nên tôi sẽ nói rằng một thị trường đang mở ra, nhưng sẽ không ấn định quy mô của nó. Dấu hiệu cho thấy một thị trường đã mở ra và quy mô của thị trường đó là hai câu hỏi khác nhau. Nhận thấy chính sách tạo ra nhu cầu và tính toán nhu cầu ấy lớn đến đâu đòi hỏi những bằng chứng khác nhau.

Chính sách tiêu dùng còn hướng thẳng hơn về phía thị trường. Các ý kiến triển khai nhằm đẩy nhanh sự phát triển của AI plus consumption đặt các dịch vụ nổi đời sống online và offline, cùng tiêu dùng kết hợp với trí tuệ nhân tạo, vào trung tâm nhiệm vụ. Văn kiện kêu gọi mở rộng các mô hình kinh doanh mới dựa trên sở hữu trí tuệ, kết hợp thương mại, du lịch, văn hóa và thể thao; hỗ trợ thương mại hóa sở hữu trí tuệ trong văn hóa và biểu diễn; cải thiện sự thuận tiện trong thanh toán nước ngoài và thị thực; giảm các bước đặt chỗ tại điểm du lịch; và xây dựng một hệ thống theo dõi dòng tiêu dùng theo thời gian thực bằng dữ liệu lớn. Đằng sau văn kiện này là một thay đổi trong mô hình tăng trưởng. Các nhà phân tích cho rằng khi Trung Quốc rời xa nền kinh tế cũ xoay quanh bất động sản và hạ tầng để chuyển sang tập trung vào new quality productive forces, việc ngăn áp lực giảm phát và bảo vệ nhu cầu trong nước đã trở thành nhiệm vụ của nhà nước.

Các chương trình trợ cấp bổ sung hỗ trợ khi thiết bị cũ được thay bằng thiết bị mới chống đỡ tiêu dùng không chỉ đối với đồ gia dụng và nội thất, mà cả các thiết bị số như smartphone. Xu hướng này đem lại doanh thu cho các nhà sản xuất phần cứng, đồng thời trở thành một kênh đưa các mô hình AI on-device mới nhất vào tay người tiêu dùng. Khi chính phủ dùng tiền thuế để khiến người dân thay thiết bị, các mô hình được cài trong những thiết bị ấy lan vào đời sống hằng ngày của hàng trăm triệu người. Đây là nơi chính sách vĩ mô kích thích tiêu dùng trở thành kênh phân phối cho các mô hình AI.

Một xu hướng liên quan cũng có thể thấy trong sự trỗi dậy của cái gọi là tiêu dùng mới do người trẻ dẫn dắt, kiểu tiêu dùng theo đuổi gu riêng và sự thỏa mãn tâm lý, với các mặt hàng như sản phẩm nhân vật sở hữu trí tuệ và trà sữa ngày càng có chỗ đứng. Các công ty nền tảng như Alibaba, Tencent và ByteDance đưa mô hình của mình vào các điểm tiếp xúc với người tiêu dùng trên thực địa mà chính sách này mở ra. Họ làm vậy bằng cách đặt các mô hình ngôn ngữ lớn của mình vào cửa sổ trò chuyện tìm kiếm và mua sắm, để agent xử lý các dịch vụ gắn với offline như đặt chỗ và giao hàng, rồi theo dõi dữ liệu tiêu dùng kết nối với các mini program bên trong những hệ sinh thái như WeChat. Các biện pháp mở rộng tiêu dùng dịch vụ, do Ministry of Commerce cùng một số bộ khác ban hành, mở sân cho các công ty này bằng cách gọi sự lan rộng của tiêu dùng nổi đời sống online và offline là một nhiệm vụ trung tâm.

Mối quan hệ trong đó chính phủ thiết kế dòng chảy tiêu dùng, còn các công ty đưa mô hình của mình đi trên dòng chảy ấy, giống với cấu trúc hai tầng của chính sách sản xuất. Phương hướng được vạch ra từ bên trên, còn doanh nghiệp thu tiền ở bên dưới. Trong sản xuất, các nhà máy tạo thành tầng dưới; trong giáo dục, đó là trường học và hộ gia đình; trong tiêu dùng, đó là nền tảng và người tiêu dùng. Câu nói về việc xây dựng một hệ thống theo dõi dữ liệu tiêu dùng theo thời gian thực cho thấy rất rõ bản chất của mối quan hệ này. Một chính sách làm tăng tiêu dùng cũng đồng thời đặt nền móng dữ liệu để nhìn vào tiêu dùng, rồi dữ liệu đó lại trở thành nguyên liệu nuôi các mô hình của công ty nền tảng.

Các văn bản về quản lý nhân sự cho thấy xu hướng này cũng đã lan tới lĩnh vực lao động và hành chính. Những ý kiến triển khai về phát triển ứng dụng AI cộng với nguồn nhân lực và an sinh xã hội đặt ra hướng đưa trí tuệ nhân tạo vào giới thiệu việc làm, đào tạo nghề và công tác an sinh xã hội. Khi nhà nước thúc đẩy việc làm và đào tạo lại như một chính sách, nhà nước cần các bộ dữ liệu và mô hình ứng dụng để thực hiện chính sách ấy, và nhu cầu đó lại quay về chính sách giáo dục và chính sách dữ liệu. Ngôn ngữ của đào tạo lại và ngôn ngữ của AI cộng với giáo dục cùng nhìn về một nơi. Khi các quầy hành chính giới thiệu việc làm được thay bằng trí tuệ nhân tạo, hồ sơ của người tìm việc và thông tin tuyển dụng được gom lại với nhau, rồi dữ liệu đó lại trở thành nguyên liệu cho chính sách đào tạo. Sản xuất, giáo dục, tiêu dùng và quản lý nhân sự thuộc các văn bản khác nhau của các bộ khác nhau, nhưng chúng dùng cùng một ngữ pháp.

Nhà nước đặt ra phương hướng và mục tiêu, các đơn vị vận hành hạ tầng cung cấp mô hình và sức mạnh tính toán, còn từng ngành, từng hộ gia đình và từng người lao động trả chi phí trên nền hệ thống đó, đồng thời để lại dữ liệu từ hiện trường. Vòng phản hồi này tạo ra tốc độ lan rộng của trí tuệ nhân tạo ở Trung Quốc, nhưng cũng để lại câu hỏi ai là người gánh chi phí và sức nặng ấy. Văn bản chính sách mở ra thị trường, nhưng trong các thị trường đó, người thật sự mở ví thường là cá nhân và doanh nghiệp. Việc các văn bản nhân sự gắn với an sinh xã hội khiến ta phải suy nghĩ. Khi trí tuệ nhân tạo đi vào bộ máy hành chính quản lý những người có nguy cơ mất việc cao, trí tuệ nhân tạo ấy có thể vừa là công cụ giúp người lao động, vừa là công

cụ phân loại và theo dõi họ.

Phần cuối của chương này thuộc về robot hình người và các mô hình ngành dọc. Để trí tuệ nhân tạo rời khỏi không gian số và tạo ra giá trị trong thế giới vật lý, nó cần hai thứ. Một là một cơ thể có thể tự di chuyển và xử lý đồ vật, tức là robot có trí tuệ hiện thân. Thứ còn lại không phải là tri thức lơ lửng, mà là các mô hình ngành dọc hiểu tường tận công việc của một ngành cụ thể. Trí tuệ hiện thân được nhắc tới trong các báo cáo liên quan đến U.S.-China Economic and Security Review Commission, tức bước chuyển từ large language models sang world models, cho thấy hướng đi của vốn và công nghệ Trung Quốc: từ chatbot nói năng trôi chảy sang các tác nhân hiểu và hành động trong thế giới vật lý. Nhưng phần này phải được đọc sau khi gạt bỏ sự phóng đại. Đằng sau những video trình diễn bóng bẩy, thành tựu của ngành robot vẫn mang những giới hạn rõ ràng.

Việc Trung Quốc lắp đặt nhiều robot công nghiệp hơn hầu hết các quốc gia trên thế giới cho thấy chiều sâu của hạ tầng nước này. Số công ty làm robot hình người cũng tăng nhanh, và một số thống kê cho thấy các công ty Trung Quốc đã vươn lên chiếm một phần trong nền tảng công ty robot hình người toàn cầu, hiện chỉ đứng sau Hoa Kỳ. Tuy vậy, mức lan tỏa thương mại thực tế vẫn được đánh giá là còn ở giai đoạn các đơn hàng ban đầu mới xuất hiện trong những lĩnh vực như logistics, an ninh và giáo dục. Cảnh vài robot nhảy đồng bộ trong một chương trình phát sóng dịp lễ thu hút sự chú ý, nhưng giữa sự kiện đó và việc triển khai thương xuyên trên dây chuyền nhà máy vẫn còn một khoảng cách lớn. Một báo cáo nói rằng một nhà sản xuất xe điện đã đưa robot hình người vào xưởng đúc áp lực, cho chúng thực hiện một quy trình lắp ráp cụ thể trong ba giờ liên tục, và ghi nhận tỷ lệ thành công ở mức hơn 90 phần trăm một chút khi chúng làm việc đồng thời ở cả hai phía của bàn thao tác.

Con số đó vừa là thành tựu, vừa là giới hạn. Tỷ lệ thành công ở mức hơn 90 phần trăm một chút có nghĩa là gần như cứ mười lần thử thì có một lần thất bại, và trong một quy trình sản xuất ô tô đòi hỏi độ nguyên vẹn tuyệt đối, mức đó vẫn còn quá sớm để thay thế hoàn toàn con người. Robot vẫn đang ở giai đoạn thử nghiệm và hỗ trợ. Cùng một robot sẽ hiện ra hoàn toàn khác nhau tùy theo việc con số 90 phần trăm được đọc như ngôn ngữ của thành công hay ngôn ngữ của thất bại. Khác biệt nằm ở đây: video trình diễn chỉ cắt ghép những lần thành công, còn sổ sách của nhà máy ghi lại chi phí của cả một lần thất bại.

Lộ trình của chính Trung Quốc cũng có thể được đọc như một sự thừa nhận điểm này. Lộ trình ấy chia quá trình thành các giai đoạn: một giai đoạn để đột phá các công nghệ cốt lõi và bước vào sản xuất hàng loạt ban đầu, một giai đoạn để mở rộng ứng dụng và tiến tới sản xuất hàng loạt quy mô lớn, và một giai đoạn nhìn xa hơn đến sự lan tỏa và tích hợp với thế giới vật lý. Lộ trình càng đi xa, mục tiêu càng tham vọng, nhưng cũng mang theo nhiều câu hỏi chưa được kiểm chứng hơn. Việc tích hợp trí tuệ nhân tạo tổng quát với thế giới vật lý ở cuối lộ trình không phải là điều có thể xác nhận bằng bằng chứng hiện nay. Nó chỉ nên được đọc như một phương

hướng, một vùng đất chưa được kiểm chứng. Đã có những động thái vốn nhằm gắn kết chuỗi giá trị theo chiều dọc bằng cách mua cổ phần trong các công ty cơ khí chính xác sản xuất linh kiện robot, và cũng có dấu hiệu cho thấy một số công ty đang đặt nền móng bằng cách công bố các bộ dữ liệu huấn luyện mã nguồn mở cho thao tác robot.

Nhưng khoản đầu tư như vậy không có nghĩa là công nghệ đã hoàn tất. Việc mua các công ty linh kiện để gắn phần cứng theo chiều dọc cũng là bằng chứng cho thấy con đường vẫn còn đủ dài đến mức chuỗi giá trị phải được kéo vào bên trong. Việc công bố bộ dữ liệu để cùng xây dựng năng lực nền tảng cũng cho thấy theo chiều ngược lại rằng dữ liệu riêng của từng công ty vẫn chưa đủ. Việc một lộ trình tham vọng và một bản tự đánh giá rằng ngành vẫn ở giai đoạn ban đầu cùng xuất hiện trong một tài liệu cho chúng ta thấy sự cân bằng cần có khi đọc ngành này.

Mô hình ngành dọc cũng gặp những giới hạn rõ ràng trong đời thực. Trí tuệ công nghiệp mà kế hoạch hành động về sản xuất đòi hỏi không phải là kỹ năng hội thoại, mà là khả năng phản ứng theo thời gian thực và độ chính xác. Khi một mô hình dự đoán dòng nhiệt trong lò cao và điều chỉnh quy trình cán được đưa vào nhà máy, điều quan trọng không phải là quy mô tham số, mà là khả năng đưa ra câu trả lời đúng lúc trong các ràng buộc vật lý. Trên sàn sản xuất nơi gang nóng chảy đang đổ ra, một khoảnh khắc chần chừ của trí tuệ nhân tạo có thể dẫn đến tai nạn. Vì vậy, tiêu chuẩn hiệu năng cho các mô hình công nghiệp không được viết bằng độ trôi chảy trong hội thoại, mà bằng độ trễ tính theo mili giây.

Một mô hình lớn hơn không tự động trở thành mô hình tốt hơn. Mô hình tốt là mô hình đáp ứng được các giới hạn thời gian vật lý nghiêm ngặt. Trong thế giới chatbot, người dùng có thể chờ khi câu trả lời đến hơi muộn, nhưng trong nhà máy, độ trễ ấy có thể là ranh giới giữa một lỗi sản phẩm và một tai nạn. Sự khác biệt này cho thấy vì sao trí tuệ nhân tạo cần các tiêu chuẩn riêng khi được gắn vào hệ thống vật lý. Một mô hình đạt điểm cao trong bài kiểm tra chưa chắc vận hành tốt trong nhà máy. Giữa hai việc đó có một ngưỡng cứng gọi là hiệu năng thời gian thực.

Trong môi trường doanh nghiệp, một rào cản nguy hiểm khác đối với việc áp dụng trí tuệ nhân tạo là ảo giác, tức việc bịa ra những sự kiện không tồn tại. Nếu một mô hình xử lý dữ liệu tài chính, nhân sự và logistics tạo ra các con số nghe có vẻ hợp lý nhưng sai, thiệt hại sẽ rơi thẳng vào doanh nghiệp. Khi không có thông tin, mô hình phải trả lời rằng không có thông tin. Việc tự tin đưa ra những con số sai cụ thể trong tình huống ấy là một vấn đề cũng xuất hiện trong các đánh giá hiệu năng thực tế. Vì lý do này, kế hoạch hành động về sản xuất đặt quản trị an toàn lên hàng đầu, kêu gọi thiết lập cơ chế chief data officer chuyên trách dữ liệu, và trên thực tế coi cả quy trình rà soát giấy phép mã nguồn mở, lỗi hỏng bảo mật lẫn retrieval-augmented generation gắn với cơ sở dữ liệu công nghiệp là những yêu cầu cần có.

Tài liệu này viết nhiều lớp biện pháp bảo vệ để trí tuệ nhân tạo có thể được giao phó cho các hệ thống vật lý mà không bị tin tưởng một cách mù quáng. Chính sự tồn tại của các biện pháp bảo vệ này cũng cho thấy văn bản của chính phủ thừa nhận rằng trí tuệ nhân tạo hiện nay chưa đủ đáng tin cậy để tự mình quản lý hiện trường. Chỉ đạo gắn bằng chứng thông qua retrieval-augmented generation và để chief data officer giám sát dữ liệu chẳng khác nào yêu cầu kiểm chứng câu trả lời của trí tuệ nhân tạo bằng cơ sở dữ liệu công nghiệp, thay vì chấp nhận nguyên trạng. Việc các mục tiêu đầy tham vọng và những quy tắc an toàn thận trọng cùng nằm trong một tài liệu cho thấy hai mặt của chính sách này.

Nhìn theo cách này, các văn kiện được bàn trong chương này cùng chỉ về một hướng, dù chúng nói tới những ngành khác nhau. Kế hoạch hành động về sản xuất đặt ra mục tiêu cho bộ dữ liệu và tác nhân ở phía cung. Các văn kiện về giáo dục, tiêu dùng và nhân sự mở cửa thị trường ở phía cầu. Robot hình người và các mô hình chuyên ngành tạo thành cánh cửa cuối cùng để trí tuệ nhân tạo đi xuống thế giới vật chất. Một cơ cấu trong đó nhà nước vạch hướng đi và doanh nghiệp chạy theo hướng đó tạo ra tốc độ. Mô hình giá rẻ, thị trường nội địa rộng và mục tiêu chính xác theo từng ngành làm vòng phản hồi bắt đầu vận hành. Nhưng chính cơ cấu ấy cũng cho thấy gánh nặng rơi vào đâu. Cá nhân gánh chi phí đào tạo lại, doanh nghiệp gánh chi phí tiếp nhận công nghệ, nơi làm việc cùng chia sẻ rủi ro do ảo giác của mô hình, còn robot vẫn đang đứng ở vị trí mà cứ mười lần thử thì có một lần thất bại.

Chúng ta có thể xác nhận những con số mà các văn kiện chính sách đã ghi cho năm 2027 và các năm sau đó. Những con số ấy có đạt được hay không, và nếu đạt được thì ai phải trả giá, vẫn là câu hỏi mà bằng chứng hiện có chưa thể trả lời. Các chính sách này nhắm tới những hệ thống vật lý được đo bằng micromet và mili giây. Việc chúng có thể biến mục tiêu trên giấy thành tỷ lệ thành công trên sàn nhà máy hay không sẽ được trả lời trong vài năm tới bằng số đơn hàng, tỷ lệ lỗi và độ trễ.

Chương 10. Cạnh tranh AI Mỹ-Trung, kiểm soát xuất khẩu và sự đảo chiều trong kiểm soát tiếp cận ở nước ngoài

Tháng 12 năm 2024, một thông báo đăng trên website của Bureau of Industry and Security thuộc U.S. Commerce Department đã làm ngành bán dẫn Trung Quốc khựng lại. Tên văn bản nghe như ngôn ngữ hành chính, nhưng danh sách bên trong thì không. Một trăm bốn mươi công ty Trung Quốc bị đưa cùng lúc vào danh sách các doanh nghiệp bị cấm nhận hàng xuất khẩu bán dẫn tiên tiến, còn 24 loại thiết bị sản xuất tiên tiến và ba công cụ phần mềm cốt lõi bị đặt dưới kiểm soát xuất khẩu. Từ ngày đó, các công ty có tên trong danh sách không còn có thể mua thiết bị, linh kiện hoặc phần mềm thiết kế có nguồn gốc từ Hoa Kỳ qua các kênh chính thức.

Chỉ một thông báo đã buộc hàng trăm công ty phải viết lại kế hoạch mua sắm chỉ sau một đêm. Nhìn lại quá trình danh sách này hình thành, các biện pháp kiểm soát của Hoa Kỳ không xuất hiện một lần với đầy đủ hình dạng. Chúng dày lên từng lớp, từ chip đến thiết bị, rồi đến phần mềm, sau đó là phối hợp với đồng minh. Khi cấu trúc ấy ngày càng dày, phản ứng từ phía bị kiểm soát cũng vượt khỏi việc đổi nguồn mua sắm và chuyển sang xây dựng lại chính ngành công nghiệp. Chương này đi theo mỗi khóa chặt đó.

Báo cáo "Comparison of U.S. and Chinese AI Competitiveness" của Samjong KPMG Economic Research Institute trình bày lịch trình kiểm soát này một cách điềm tĩnh. Nó bắt đầu vào tháng 8 năm 2022, với các hạn chế xuất khẩu sang Trung Quốc đối với GPU H100 tiên tiến của Nvidia. Các công ty Trung Quốc tìm cách đi vòng bằng những mẫu có thông số thấp hơn, và các chip cấp thấp hơn như A800 và H800 đã lắp vào khoảng trống. Tháng 10 năm 2023, Hoa Kỳ chặn luôn lối đi vòng đó. Sau đó là danh sách tháng 12 năm 2024. Cùng với việc này, Hoa Kỳ gây sức ép lên các nhà sản xuất thiết bị như ASML của Hà Lan và Tokyo Electron của Nhật Bản, thu hẹp con đường tiếp cận thiết bị quang khắc tiên tiến của chính Trung Quốc.

Bức tranh hiện ra không phải là quy định xuất khẩu của một quốc gia vận động một mình, mà là quy định của nhiều quốc gia cùng chuyển động. Trọng tâm kiểm soát đi ngược lên thượng nguồn, từ chip thành phẩm đến thiết bị làm ra những con chip đó, rồi từ thiết bị đến phần mềm dùng để thiết kế thiết bị ấy. Điều đó đáng được chú ý. Các biện pháp kiểm soát không dừng ở việc chặn bán hàng. Chúng chạm vào khả năng làm ra sản phẩm ngay từ đầu. Khi cách vượt qua ngưỡng quản lý bằng chip thông số thấp hơn bị chặn, lựa chọn còn lại không còn là đi vòng qua ngưỡng ấy nữa. Đó là dựng một cánh cửa mới ở nơi khác.

Một số công ty tìm cách lách qua tấm lưới này bằng cách bẻ cong quy định, nhưng con đường đó cũng phải trả giá đắt. Theo "Hana China Weekly" của Hana Securities, vào tháng 3 năm 2026, công ty thiết bị bán dẫn Applied Materials của Mỹ đã đồng ý nộp khoảng 252 triệu USD tiền

phạt vì cáo buộc xuất khẩu trái phép thiết bị sang SMIC của Trung Quốc. Các điều tra viên phát hiện công ty đã gửi thiết bị sản xuất tại Mỹ tới công ty con ở Hàn Quốc để lắp ráp, rồi chuyển tiếp sang SMIC tại Trung Quốc qua một tuyến đường phức tạp. Cuối cùng việc này bị phát hiện. Khoản phạt này cho thấy ngay cả những tuyến đường trong đó thiết bị đi qua một nước rồi được lắp ráp ở một nước khác cũng đã trở thành mục tiêu giám sát của Mỹ. Tấm lưới quản lý giờ đây đã được đan chặt xuyên biên giới.

Chính phủ Mỹ còn đi thêm một bước và đưa vào diện xem xét cả kế hoạch giới hạn số chip Nvidia H200 mà mỗi công ty Trung Quốc được mua. Thay vì chặn hẳn việc mua, kế hoạch này đặt trần cho lượng chip có thể mua. Giữa cấm đoán và cho phép, một tay nắm mới mang tên kiểm soát khối lượng đã xuất hiện. Một khi tầng trung gian này nằm giữa lệnh cấm hoàn toàn và thương mại tự do, bên kiểm soát có thể dùng đầu ngón tay điều chỉnh cả mức tồn kho của phía bên kia. Khi các biện pháp kiểm soát trở nên chính xác hơn, việc các công ty Trung Quốc có thể mua gì từ nước ngoài, và với số lượng bao nhiêu, ngày càng trở thành một biến số do quyết định của Washington đặt ra.

Đến thời điểm này, chiến lược mua chip từ nước ngoài trên thực tế đã trở thành một con đường bị chặn. Vì vậy phản ứng của Trung Quốc vượt ra ngoài việc thay đổi cách mua sắm và chuyển sang một câu hỏi khác: cần sản xuất gì trong nước? Tín hiệu xuất hiện ở quy mô vốn nhà nước. Tháng 5 năm 2024, chính phủ Trung Quốc khởi động giai đoạn ba của National Integrated Circuit Industry Investment Fund, trị giá 344 tỷ nhân dân tệ. Dòng tiền này chảy vào sản xuất của Trung Quốc trong các lĩnh vực năng lực foundry, thiết bị bán dẫn và vật liệu tiên tiến. Nếu Big Fund và dự án Eastern Data, Western Computing đã bàn ở chương trước tạo thành bản thiết kế phần cứng, thì quỹ giai đoạn ba có thể được hiểu là nơi áp lực bên ngoài từ các lệnh trừng phạt đẩy nhanh bản thiết kế đó.

Khi chính phủ định hướng và vốn nhà nước chống lưng, các công ty bắt đầu thử nghiệm thiết bị Trung Quốc và vật liệu Trung Quốc trên nền tảng đó. Quyết định tự sản xuất trong nước những thứ trước đây mua từ nước ngoài không phải là phán đoán của một công ty riêng lẻ. Đó là kết quả của chính sách quốc gia và hoạt động mua sắm của doanh nghiệp cùng vận hành. Ở đây có thể thấy khoảng cách giữa cơ quan nhà nước và doanh nghiệp thu hẹp lại như thế nào. Khi một quỹ nhà nước rót tiền vào foundry và vật liệu, dòng tiền ấy đi xuống việc mở rộng các dây chuyền sản xuất cụ thể của doanh nghiệp và những nỗ lực chuyển các quy trình cụ thể sang sản xuất tại Trung Quốc. Trong lúc các biện pháp kiểm soát đóng cánh cửa từ bên ngoài, vốn nhà nước gieo hạt cho một chuỗi cung ứng thay thế ở bên trong.

Trong dòng chảy đó, Huawei nổi lên như phương án thay thế trong nước để lấp khoảng trống mà Nvidia để lại. Huawei phát hành dòng bán dẫn AI riêng của mình, Ascend 910, và chuyển các chip vốn được sản xuất bởi TSMC của Đài Loan sang xưởng đúc SMIC của Trung Quốc. Ở phía phần mềm, Huawei cũng chuẩn bị kiến trúc tính toán và framework riêng để không bị ràng

buộc vào CUDA, môi trường phát triển của Nvidia. Chỉ đổi con chip sang chip Trung Quốc là chưa đủ. Chẳng nào các nhà phát triển vẫn quen với hệ sinh thái phần mềm của Nvidia, người ta sẽ không chuyển sang dù phần cứng đã thay đổi.

Từ bỏ công cụ quen thuộc và chuyển sang công cụ mới cần thời gian và chi phí học lại. Nếu các chi phí đó cao, chip Trung Quốc có thể chất đống trong kho, còn việc phát triển thực tế vẫn diễn ra trong hệ sinh thái của người khác. Vì vậy, Huawei chọn xây dựng cả con chip lẫn môi trường phần mềm chạy trên nó thành một hệ thống Trung Quốc thống nhất. Tính toán ở đây là các nhà phát triển chỉ chuyển sang khi độc lập phần cứng và độc lập phần mềm được buộc chặt với nhau.

Dù vậy, các con số ở đây cần được xử lý thận trọng. Đã có nhiều tuyên bố về hiệu năng suy luận và thể hệ tiến trình của chip Ascend, nhưng các tài liệu hiện có chưa đủ để kết luận chắc chắn rằng những chip này ngang với các chip cao cấp của Nvidia về hiệu năng. Điều có thể xác nhận là chip Trung Quốc đã bắt đầu được dùng trong các dịch vụ thực tế bên trong Trung Quốc, và việc tiếp nhận này tăng tốc trong điều kiện bị trừng phạt. Một câu so sánh hiệu năng đặt cạnh nhau và một câu nói rằng chip đã bắt đầu được dùng trong nước có trọng lượng khác nhau.

Câu thứ nhất chứa cái bẫy của benchmark và tiêu chuẩn đo lường, trong khi câu thứ hai để lại dấu vết có thể kiểm chứng qua mức độ tiếp nhận và doanh số. Văn bản này chỉ chấp nhận câu thứ hai như một sự thật. Những tuyên bố về hiệu năng ngang bằng có thể là tín hiệu được định hình bởi lợi ích của bên công bố, nên an toàn hơn là chưa xem chúng như điều đã ngã ngũ cho đến khi xuất hiện tài liệu được kiểm chứng trong cùng điều kiện.

Bài phát biểu chính của Huawei tại một sự kiện học thuật về bán dẫn tổ chức ở Thượng Hải năm 2026 cho thấy vì sao sự thận trọng này lại quan trọng. Sau khi không còn có thể đưa vào các thiết bị quang khắc EUV dùng để sản xuất chip tiên tiến, Huawei đã rời khỏi con đường truyền thống là làm cho transistor nhỏ hơn về mặt vật lý. Thay vào đó, họ đưa ra một lời giải mới: xếp chip lên cao theo nhiều lớp bằng thiết bị cũ hơn để tăng mật độ. Công ty cũng trình bày lộ trình đạt mật độ có thể so sánh với một số thế hệ nhất định trong vài năm tới, ngay cả khi không có EUV. Đây gần với một tuyên bố về mục tiêu tương lai hơn là một kết quả đã được chứng minh. Điều này đáng chú ý vì các lệnh trừng phạt đã đẩy Huawei vào một lối đi vòng mà nếu không có sức ép ấy, có lẽ họ sẽ không chọn. Nhưng sẽ an toàn hơn nếu không đọc các con số trong lộ trình như hiệu năng đã hoàn tất.

Lộ trình của một doanh nghiệp cũng là tín hiệu gửi tới thị trường, cơ quan quản lý và chính phủ của chính doanh nghiệp đó. Vì vậy, không ai có thể bảo đảm rằng năm mục tiêu và các chỉ tiêu mục tiêu sẽ thành hiện thực đúng như đã nêu. Cách đúng để xử lý tài liệu này là đọc nó như một cảnh cho thấy các biện pháp kiểm soát đã bẻ hướng công nghệ ra sao, chứ không phải như bằng chứng rằng kết quả đã nằm chắc trong tay. Giữa việc đặt mục tiêu và đạt được mục tiêu là

một khoảng thời gian thực hiện vẫn chưa được lấp đầy.

Giới hạn phần cứng cũng làm thay đổi chiến lược phần mềm. Con đường mà các tập đoàn công nghệ lớn của Mỹ đi theo nghiêng về việc đổ thêm nhiều năng lực tính toán vào các mô hình lớn hơn. Một báo cáo của U.S.-China Economic and Security Review Commission cho thấy sự tương phản này qua quy mô chi tiêu vốn. Trong khi các công ty lớn của Mỹ chi ít nhất \$350 tỷ vào năm 2025 và hơn \$400 tỷ vào năm 2026 cho trung tâm dữ liệu và tính toán, khoản đầu tư liên quan của các công ty điện toán đám mây Trung Quốc vẫn dưới \$40 tỷ. Khoảng cách gần gấp mười lần. Nếu cuộc đua được quyết định bằng lượng tính toán thuần túy, Trung Quốc sẽ gặp một đối thủ rất khó. Nếu không thể đổ thêm tiền, họ phải đổi cách cạnh tranh, và các công ty Trung Quốc đã chọn con đường đó. Thay vì chiến đấu theo cùng luật chơi trong một trận đấu mà vốn xuất phát chênh nhau mười lần, họ chuyển cuộc đấu sang một sân có luật khác.

Báo cáo giải thích sự dịch chuyển này qua hai vòng lặp. Các biện pháp kiểm soát của Mỹ nhằm vào quyền tiếp cận các chip tiên tiến cần để huấn luyện mô hình, tức vòng lặp tính toán. Trung Quốc bắt lợi trong vòng lặp này. Các công ty Trung Quốc vì thế chuyển trọng tâm sang các mô hình mở không đòi hỏi lượng tính toán khổng lồ, và sang vòng lặp vật lý, nơi nhà máy và công trường liên tục tạo ra dữ liệu. Họ đưa các mô hình nhỏ hơn, được điều chỉnh cho các điểm sản xuất, vào các ngành công nghiệp thật, rồi thu thập dữ liệu phát sinh ở đó và sửa lại mô hình. Con đường này làm giảm sự lệ thuộc vào chip tiên tiến và đi vòng qua đúng điểm mà các biện pháp kiểm soát nhắm tới. Nền tảng sản xuất khổng lồ của Trung Quốc trở thành một vũ khí ở đây.

Nhiều nhà máy hơn nghĩa là nhiều dữ liệu hơn, và nhiều dữ liệu hơn nghĩa là có nhiều mô hình hơn để sửa đổi dựa trên dữ liệu đó. Trong khi các biện pháp kiểm soát của Hoa Kỳ đang đóng cánh cửa mang tên năng lực tính toán, Trung Quốc bước qua cánh cửa ở ngay bên cạnh, mang tên dữ liệu. Việc một con đường vẫn tương đối rộng vì thiết kế của các biện pháp kiểm soát chỉ nhằm vào một lối đi cũng cho thấy giới hạn của kiểm soát xuất khẩu như một công cụ. Sự lan rộng của sản xuất AI+ và các tác nhân công nghiệp đã bàn ở chương trước có thể được đọc cùng với điều này, như tuyến đường thực tế đưa dữ liệu vào vòng lặp vật lý.

Cái tên đưa chiến lược hiệu quả này ra thế giới là DeepSeek. Mô hình của DeepSeek, được công bố vào đầu năm 2025, chọn thay đổi cấu trúc thay vì huy động sức tính toán áp đảo, và được đánh giá là đã tiến gần tới hiệu năng tiên tiến với chi phí chỉ bằng một phần nhỏ so với các mô hình hàng đầu của Hoa Kỳ. Cú sốc mà DeepSeek tạo ra không nằm nhiều ở thứ hạng hiệu năng, mà ở khả năng nó cho thấy: ngay cả với ít tài nguyên hơn, một công ty vẫn có thể đứng gần nhóm dẫn đầu. Các nhà quan sát cũng ghi nhận cách DeepSeek len vào những khoảng trống mà lệnh trừng phạt để lại, dùng các chip Nvidia cũ hơn mà họ đã có sẵn để huấn luyện, trong khi chạy suy luận cho dịch vụ trên chip do Trung Quốc sản xuất. Cách phân chia huấn luyện và suy luận cho các loại chip khác nhau cho thấy họ đã đọc rất kỹ những kẽ hở còn lại sau

các lệnh trừng phạt.

Đằng sau đó là một phán đoán: huấn luyện cần chip tiên tiến hơn, nhưng lượng chip đã tích trữ có thể giúp giữ vững tuyến này, còn suy luận ở giai đoạn dịch vụ có thể được xử lý bằng chip Trung Quốc. Tuy vậy, các nguồn khác nhau không thống nhất về tuyên bố rằng mô hình đã đạt điểm ở mức huy chương vàng trong một cuộc thi cụ thể, cũng như về ngày công bố chính xác, nên văn bản này không xem những chi tiết đó là sự thật đã được chốt. Thứ hạng benchmark và kết quả thi có thể thay đổi theo phương pháp đo và thời điểm phát hành. Cách chính xác hơn là chỉ chấp nhận hướng đi rộng hơn, rằng Trung Quốc đã tiến gần nhóm dẫn đầu, còn thứ hạng hoặc ngày tháng cụ thể thì để riêng ra.

Các lệnh trừng phạt cũng bắt đầu tạo ra nhu cầu trong nước. Để phòng trước sự bất định từ các hạn chế của Hoa Kỳ, ByteDance đã dùng một phần lớn chi tiêu vốn năm 2025 để mua trước chip Trung Quốc. Phép tính rất rõ: nếu không chắc sau này công ty có thể mua được hay không, thì nên tích trữ khi vẫn còn có thể. MetaX, một công ty GPU của Trung Quốc, đã nâng mục tiêu doanh thu quý I năm 2026 so với năm trước khi nhu cầu đối với GPU Trung Quốc cho trung tâm dữ liệu và huấn luyện tăng lên. Khi con đường mua từ nước ngoài bị thu hẹp, nhu cầu nổi lên trong nước, và nhu cầu đó chảy ngược thành doanh thu cho các công ty linh kiện Trung Quốc. Các biện pháp kiểm soát của Hoa Kỳ rốt cuộc đã góp phần nuôi lớn một hệ sinh thái thay thế bên trong Trung Quốc, cho thấy kiểm soát xuất khẩu không phải lúc nào cũng chỉ tạo ra kết quả mà nó nhắm tới.

Khi nguồn mua sắm từ bên ngoài bị cắt đứt, nguồn mua sắm từ bên trong sẽ lớn lên. Một chuỗi cung ứng Trung Quốc được xây dựng theo cách này sẽ không còn rung lắc nhiều như trước khi lệnh trừng phạt tiếp theo xuất hiện. Các biện pháp kiểm soát nhằm ghìm Trung Quốc lại, nhưng điều xuất hiện lại là bảng doanh thu dày hơn cho các công ty linh kiện Trung Quốc. Một biện pháp cắt đứt hoạt động mua sắm của một quốc gia đã tạo ra một thị trường được bảo đảm cho các nhà cung cấp bên trong chính quốc gia đó. Kiểm soát và tự lực bắt đầu đẩy nhau đi lên.

Nếu đó là câu chuyện của phía bị kiểm soát, thì sự đảo chiều trong nhan đề chương này xuất hiện ngay sau đó. Khi các mô hình giá rẻ, hiệu năng cao được xây dựng dưới lệnh trừng phạt bắt đầu được ưa chuộng ở nước ngoài, chính Trung Quốc lại bắt đầu bàn xem có nên thu hẹp cánh cửa để các mô hình ấy rời khỏi đất nước hay không. Theo một bài của Times of India dẫn lại thông tin từ Reuters, vào giữa năm 2026, Bộ Thương mại Trung Quốc đã triệu tập lãnh đạo của các công ty lớn như Alibaba, ByteDance và Zhipu để thảo luận về giới hạn đối với quyền truy cập từ nước ngoài vào các mô hình AI tiên tiến của Trung Quốc. Tin này trước hết nên được hiểu là một cuộc thảo luận đang được xem xét, chưa phải một quy định đã hoàn tất.

Nguồn ở đây là một bản tin truyền thông, và điều bản tin mô tả không phải là một biện pháp đã có hiệu lực, mà là dấu hiệu cho thấy các bộ ngành và công ty liên quan đã ngồi lại để bàn về

những phương án có thể có. Văn bản này xem điểm đó như một tín hiệu về hướng đi, chứ không phải một chính sách đã được quyết định. Việc một cuộc thảo luận như vậy xuất hiện, việc Trung Quốc, nước từng nói về sự cởi mở thông qua mã nguồn mở, có thể tự mình thu hẹp cánh cửa ấy, chính là sự đảo chiều cốt lõi mà chương này bàn tới. Nếu cởi mở từng là một chiến lược, thì rút lại chiến lược ấy cũng có thể là một chiến lược.

Đang sau cuộc thảo luận là những lợi ích mà Trung Quốc đã giành được nhờ sự cởi mở. Một báo cáo của U.S.-China Economic and Security Review Commission nói rằng Trung Quốc đã nhắm tới Hugging Face, kho mô hình công khai lớn nhất thế giới, trong nỗ lực vượt qua các hạn chế điện toán của Hoa Kỳ. Báo cáo cho biết đến nửa cuối năm 2025, lượt tải xuống các mô hình Trung Quốc đã vượt các mô hình Hoa Kỳ, và dòng Qwen của Alibaba đã hình thành một hệ sinh thái lớn với hơn 100.000 mô hình phái sinh. Những con số đó cần được kiểm tra lại, nhưng hướng đi thì rõ. Thông qua chiến lược phát hành công khai, Trung Quốc đã kéo các nhà phát triển trên khắp thế giới về phía các mô hình của mình, và các mô hình Trung Quốc đã bén rễ tại các địa điểm phát triển ở nhiều quốc gia.

Các mô hình được phát hành miễn phí lan rộng qua biên giới, và khi ngày càng nhiều mô hình phái sinh được xây dựng trên nền tảng đó, ảnh hưởng của nhà phát triển mô hình gốc cũng tăng theo. Khi các nhà phát triển xây dựng dịch vụ riêng của họ trên một mô hình nền tảng, những dịch vụ ấy thừa hưởng phương pháp và thói quen của mô hình. Theo nghĩa này, tính mở không phải là một mất mát, mà là một chiến lược để giành thị phần. Trong một lĩnh vực mà Trung Quốc không thể dẫn đầu bằng năng lực tính toán, việc phát hành công khai đã giúp nước này có vị trí rộng hơn trong hệ sinh thái nhà phát triển toàn cầu.

Vấn đề là tính mở này chỉ vận hành theo một chiều. Báo cáo chỉ ra sự bất cân xứng được cài sẵn trong cách quản trị của Trung Quốc. Các mô hình Trung Quốc lớn lên bằng cách hấp thụ dữ liệu mở từ khắp thế giới và phản hồi từ các cộng đồng toàn cầu, nhưng luật an ninh dữ liệu của Trung Quốc lại chặn rất chặt việc dữ liệu rời khỏi lãnh thổ, khiến các nhà phát triển bên ngoài không có con đường tương đương để tiếp cận dữ liệu bên trong Trung Quốc theo cùng cách đó. Cánh cửa để tiếp nhận thì rộng, còn cánh cửa để đưa ra thì hẹp. Dữ liệu của thế giới chảy vào, nhưng dữ liệu của Trung Quốc không chảy ra. Đó là một lối đi chỉ mở ở một phía. Sự bất cân xứng này không xuất hiện tình cờ. Nó bắt nguồn từ một hệ thống quy tắc được thiết kế quanh an ninh dữ liệu.

Các chính sách về bộ dữ liệu của National Data Administration và những quy tắc an ninh dữ liệu đã bàn ở các chương trước hoạt động ở đây như một cơ chế chỉ mở một cánh cửa ra thế giới bên ngoài. Cuộc thảo luận mới nhất của Ministry of Commerce có thể được hiểu là đang đẩy sự bất cân xứng này thêm một bước. Sau khi giành được trí tuệ tiên tiến và thị phần toàn cầu thông qua trọng số mở, Trung Quốc đang chuyển sang ngăn năng lực cốt lõi ấy chảy ngược về các quốc gia cạnh tranh. Cách phán đoán này khớp từ cả hai đầu: tăng trưởng nhờ tính mở,

rời thu hẹp tính mở ấy để giữ thành quả ở bên trong.

Logic nâng đỡ sự đổi hướng này có thể được tìm thấy trong cách nhìn riêng của Trung Quốc về an ninh. Các nghiên cứu học thuật về triết lý quản lý AI của Trung Quốc giải thích rằng Trung Quốc không xem artificial intelligence chỉ là một công cụ giúp kinh tế, mà là một công nghệ có thể làm rung chuyển chính trị và toàn xã hội. Từ góc nhìn này, thuật toán và mô hình AI là sản phẩm thương mại, nhưng chúng cũng là đối tượng mà nhà nước phải quản lý. Bài nghiên cứu truy nguyên gốc rễ của cách nghĩ này tới một khái niệm an ninh rộng, lấy an ninh của nhân dân làm cơ sở và an ninh chính trị làm nền tảng.

Trong cách nghĩ xem kinh tế và công nghệ nằm trong khuôn khổ an ninh này, câu hỏi có nên mở mô hình ra thế giới bên ngoài không rời khỏi phạm vi thương mại và trở thành một vấn đề an ninh. Nhìn theo cách đó, việc phát hành vô điều kiện một mô hình gần với trình độ cao nhất thế giới chẳng khác gì trao đi một tài sản chiến lược.

Cuộc thảo luận về việc xem hành vi đánh cắp sở hữu trí tuệ là vấn đề an ninh quốc gia cũng có thể được hiểu trong cùng khuôn khổ ấy. Điểm này cũng cần thận trọng. Một số tài liệu so sánh AI với các chương trình phát triển vũ khí quốc gia trước đây, nhưng sự so sánh đó là cách diễn giải của nhà phân tích, không phải ngôn ngữ chính sách đã được xác lập. Văn bản này không biến sự so sánh đó thành một khẳng định chắc chắn. Sự tồn tại của một khuôn khổ an ninh để xử lý công nghệ mang một trọng lượng riêng; lối nói ví cách nghĩ ấy với một chương trình vũ khí cụ thể lại mang một trọng lượng khác.

Các tính toán kinh tế cũng đang vận hành cùng lúc. Theo các tường thuật, những cuộc thảo luận của Ministry of Commerce không chỉ bao gồm việc thu hẹp quyền tiếp cận mô hình từ nước ngoài, mà còn cả việc hạn chế đầu tư nước ngoài vào các startup AI trong China. Điều này dường như nhằm ngăn vốn bên ngoài giành ảnh hưởng thông qua cổ phần trong các công ty AI đầy triển vọng của China. Bên trong China, các công ty lớn đang rót những khoản tiền lớn vào hạ tầng cloud và mô hình. Nếu các mô hình tiên tiến mà họ xây dựng bị rò rỉ ra nước ngoài với giá thấp, việc thu hồi khoản đầu tư đó sẽ khó hơn.

Cuộc thảo luận tìm cách thu hẹp cả cánh cửa dẫn tới mô hình như một sản phẩm hoàn chỉnh lẫn cánh cửa dẫn tới công ty đã tạo ra mô hình như một tài sản. Có thể đọc một phần cuộc thảo luận này như một phán đoán rằng China trước hết nên tăng quy mô nhờ mở cửa, rồi sau đó thu hẹp cánh cửa để giữ lợi nhuận ở trong nước. Khi logic an ninh và logic kinh tế cùng chỉ về một hướng, kết luận rằng cánh cửa nên được thu hẹp sẽ dễ trở nên cứng rắn hơn. Khi lập luận bảo vệ tài sản quốc gia trùng khít với phép tính bảo vệ lợi nhuận doanh nghiệp, chính sách thường trượt về một hướng.

Một số người đọc sự đảo chiều này qua lăng kính những bước đi qua lại với Hoa Kỳ. Cùng báo cáo đó nói rằng Hoa Kỳ cũng đã áp dụng các biện pháp để ngăn công dân nước ngoài tiếp cận các mô hình tiên tiến của mình, và đặt cuộc thảo luận của Trung Quốc trong khung phản ứng

lại. Vì đây là một tập hợp tình huống được báo chí ghi nhận, cách đọc chính xác hơn không phải là coi nó như một quan hệ nhân quả đã được xác lập, mà là một mô thức trong đó mỗi bên phản chiếu biện pháp của bên kia. Khi một bên đóng cửa, bên kia cũng đóng cửa, và lần đóng cửa ấy lại kéo theo một lần đóng cửa khác từ phía đối diện. Hệ sinh thái mô hình từng mở đang bị chia cắt theo biên giới quốc gia. Qua nhiều nguồn khác nhau, có thể thấy quản trị AI đang tách thành các phe đặt đối mới lên trước, đặt quản lý lên trước, hoặc đặt chủ quyền số do nhà nước dẫn dắt lên trước.

Điều vẫn chưa rõ là sự chia tách này sẽ đặt ra những điều kiện nào cho Thế giới thứ ba, nghĩa là ai sẽ được tiếp cận AI và với cái giá nào. Bằng chứng theo từng quốc gia vẫn còn mỏng, nên văn bản này để ngỏ câu hỏi đó thay vì xem nó như một dự báo đã chắc chắn. Khi ranh giới giữa mở và đóng đang được vẽ lại theo biên giới quốc gia, những tài liệu hiện có vẫn chưa trả lời được các nước nằm ngoài những đường ranh ấy sẽ phải đứng trước cánh cửa nào. Với bằng chứng hiện có trong tay, chưa thể nói bao nhiêu quốc gia sẽ được cho phép tiếp cận với điều kiện họ chấp nhận trung tâm dữ liệu và hạ tầng theo kiểu Trung Quốc, hoặc những điều kiện đó sẽ đi kèm mức giá nào.

Các công ty Hàn Quốc đứng ở đâu giữa lúc bàn cờ lớn hơn đang rung chuyển như vậy? Một báo cáo của Samjong KPMG cho rằng cuộc đối đầu Hoa Kỳ-Trung Quốc, nghịch lý thay, đang mở ra thị trường cho ngành bán dẫn và AI của Hàn Quốc. Khi hai nước chạy đua xây dựng trung tâm dữ liệu và cụm tính toán riêng, nền tảng nhu cầu AI toàn cầu đang mở rộng, và điều đó tạo ra khoảng trống cho các công ty Hàn Quốc có năng lực tiên tiến về bộ nhớ và sản xuất bán dẫn. Báo cáo cũng nhận định một hướng đi có thể nằm ở các giải pháp hướng tới doanh nghiệp, gắn AI với những ngành mà Hàn Quốc vốn đã mạnh như sản xuất, y tế và tài chính, thay vì trực diện đối đầu với Big Tech của Hoa Kỳ và Trung Quốc trong một cuộc chiến về quy mô.

Nhận định ở đây là, thay vì tham gia cuộc đua đổ lượng vốn khổng lồ vào vị trí của một mô hình lớn đa dụng, Hàn Quốc có con đường thực tế hơn khi điều chỉnh mô hình theo dữ liệu và quy trình của những ngành mà nước này đã hiểu rõ. Vị trí nằm giữa các cường quốc, tương đối gần với trung lập, cho Hàn Quốc khoảng trống để đóng vai trò điều phối trong các cuộc thảo luận về tiêu chuẩn quốc tế và quản trị. Khi hai khối thúc ép các quy tắc riêng của mình, một nơi không hoàn toàn thuộc về bên nào có thể tự nó tạo ra một tiếng nói.

Ở phía bên kia của cơ hội là rủi ro mang tính cấu trúc. Nếu hai nước coi AI là tài sản quốc gia và nghiêng về chủ nghĩa bảo hộ, các biện pháp kiểm soát xuất khẩu, giới hạn đầu tư và hạn chế quyền tiếp cận dữ liệu sẽ thu hẹp không gian hợp tác của các công ty Hàn Quốc vốn phụ thuộc vào chuỗi cung ứng toàn cầu. Báo cáo thường niên của Tencent nêu rằng căng thẳng địa chính trị có thể dẫn tới kiểm soát xuất khẩu và giới hạn đầu tư trong một số lĩnh vực, cùng với các hạn chế đối với hoạt động tại địa phương. Đoạn này cho thấy rủi ro tương tự cũng áp dụng trực tiếp với các công ty đa quốc gia của Hàn Quốc. Trong thời đại mà ngay cả một tập đoàn lớn của

Trung Quốc cũng ghi căng thẳng địa chính trị là rủi ro kinh doanh trong báo cáo của mình, không có lý do gì để nghĩ rằng căng thẳng ấy sẽ bỏ qua các công ty Hàn Quốc đang làm việc xuyên biên giới.

Còn một mối lo khác: nếu Hàn Quốc, vốn đã đi sau về quy mô vốn và hạ tầng, không xây được đủ mức độc lập trong hệ sinh thái của mình, nước này có thể bị kéo vào hai khối công nghệ khổng lồ. Trước các công ty Mỹ và Trung Quốc có khả năng hút nhân tài bằng mức lương cao hơn và điều kiện làm việc tự do hơn, tình trạng các nhà nghiên cứu và nhà phát triển Hàn Quốc rời đi vẫn là một nỗi lo sâu sắc. Khi con người rời đi, sẽ còn ít bàn tay hơn để xây dựng thể hệ công nghệ tiếp theo. Khi số bàn tay ấy giảm xuống, nền tảng cho sự độc lập cũng mỏng đi. Tài liệu này cũng chỉ ra các điểm yếu bên trong: hỗ trợ đào tạo AI và hệ thống quản lý trong các tổ chức vẫn ở mức thấp, và khoảng cách giới trong lực lượng nhân tài AI còn rộng.

Khi đánh giá vị thế của Hàn Quốc, cần nhìn kỹ khoảng cách giữa các chỉ báo sáng sủa và thực địa. Theo bản tóm tắt tiếng Hàn của Artificial Intelligence Index Report, Hàn Quốc đứng gần nhóm dẫn đầu thế giới với 14.31 bằng sáng chế AI trên mỗi 100,000 dân, và tỷ lệ sử dụng AI tạo sinh tăng từ 25.9% trong nửa đầu năm 2025 lên 30.7% trong nửa cuối năm. Số luật liên quan đến AI được ban hành cũng đạt 17, đưa Hàn Quốc vào nhóm cao trong nhiều quốc gia.

Nếu xét theo mật độ tri thức và thể chế, bức tranh trông có vẻ tiên tiến. Ba chỉ báo đều đi lên: Hàn Quốc nộp nhiều bằng sáng chế, người dân tiếp nhận AI tạo sinh nhanh, và Quốc hội bận rộn ban hành các luật liên quan. Nhưng khi hạ tầm mắt xuống các nhà máy và hiện trường công nghiệp, cảnh tượng thay đổi. Nơi mà các chỉ báo chỉ tới và nơi các nhà máy đang đứng quay về hai hướng khác nhau.

Một khảo sát tháng 11 năm 2025 của Korea Chamber of Commerce and Industry, được Korea Institute for Advancement of Technology giới thiệu, cho thấy trong số 504 doanh nghiệp sản xuất Hàn Quốc tham gia, 82.3% chưa thể đưa công nghệ AI vào nơi làm việc. Lý do là tiền và con người. Trong toàn bộ số người trả lời, 73.6% nói chi phí đầu tư AI là gánh nặng, còn 80.7% không chỉ không có chuyên gia AI mà cũng không có kế hoạch tuyển dụng hoặc đào tạo nội bộ. Trên hết, kỳ vọng của họ thấp.

Chỉ 39.4% doanh nghiệp kỳ vọng sẽ có kết quả từ chuyển đổi AI, trong khi 60.6% cho rằng hiệu quả sau khi áp dụng sẽ bị giới hạn. Đây là bức tường ba lớp: họ không thể áp dụng vì chi phí nặng, họ không thể áp dụng vì thiếu người, và họ không áp dụng vì cho rằng ngay cả khi làm thì hiệu quả cũng không lớn. Việc hai con số, vị trí dẫn đầu thế giới về bằng sáng chế và tỷ lệ 82.3% chưa áp dụng tại nơi làm việc, cùng tồn tại trong một quốc gia cho thấy thung lũng mà Hàn Quốc phải vượt qua rộng đến mức nào.

Nếu phần trước là cuộc vật lộn giữa chính phủ Trung Quốc và các doanh nghiệp Trung Quốc trước các lệnh trừng phạt, câu chuyện của Hàn Quốc đặt ra một bài toán khác: làm thế nào để thu hẹp khoảng cách giữa các chỉ số và sàn nhà máy từ bên ngoài cuộc cạnh tranh đó. Ở đây

xuất hiện một đối chiếu rất rõ. Trung Quốc dùng dữ liệu từ các cơ sở sản xuất như mỗi nổi vật chất giúp các mô hình của mình lớn lên, trong khi Hàn Quốc, dù có nền tảng sản xuất tiên tiến, vẫn có hơn tám mươi phần trăm nhà máy đứng ngoài AI. Ngay cả giữa các cường quốc sản xuất, dữ liệu sẽ trở thành tài sản hay tiếp tục bị khóa lại tùy thuộc vào việc ai xây được cây cầu giữa sàn nhà máy và mô hình trước.

Chính sách đang dịch chuyển theo hướng lấp thung lũng này. Năm 2025, Ministry of Trade, Industry and Energy mở rộng ý tưởng trước đó về sản xuất tự chủ thành nhà máy AI và khởi động một quan hệ hợp tác giữa công nghiệp, học giới và nghiên cứu. Ministry of SMEs and Startups công bố kế hoạch cung cấp 12,000 nhà máy thông minh lấy AI làm trung tâm vào năm 2030 và nâng tỷ lệ áp dụng AI trong các nhà sản xuất nhỏ và vừa lên 10%. Hướng đi đã rõ, nhưng vì các khảo sát cho thấy chi phí và nhân lực là những bức tường ngay tại hiện trường, kế hoạch này chỉ chuyển thành kết quả nếu các doanh nghiệp nhỏ và vừa có được điều kiện để gánh chi phí áp dụng thật sự, và nếu những ví dụ tích lũy đủ để họ tận mắt thấy hiệu quả sau khi áp dụng.

Đặt ra các con số mục tiêu và giảm gánh nặng cho xưởng sản xuất là hai việc khác nhau. Để mục tiêu 12.000 nhà máy thông minh thật sự kéo giảm tỷ lệ chưa áp dụng AI ở mức 82,3%, cần có những cơ chế đứng giữa hai điểm đó, giúp doanh nghiệp vượt qua hai bức tường là chi phí và con người. Việc bản thân mục tiêu áp dụng 10% vẫn được đặt ở mức thấp cũng có thể đọc theo chiều ngược lại: chính phủ thừa nhận rằng bức tường ngoài hiện trường cao đến mức ấy.

Khi đọc chương này cùng với các chương trước, một đường nét chung hiện ra. Các biện pháp kiểm soát xuất khẩu của Hoa Kỳ tìm cách chặn hoạt động mua sắm của Trung Quốc, nhưng Trung Quốc đã đổi sân chơi bằng cách chuyển sang sản xuất bằng năng lực trong nước và nâng hiệu suất, thay vì phụ thuộc vào mua sắm. Khi những mô hình được xây dựng theo cách đó lan ra khắp thế giới, Trung Quốc lại quay về với các cuộc thảo luận về việc tự thu hẹp cánh cửa ấy. Trong nhịp qua lại này, nơi kiểm soát gọi thêm kiểm soát, còn mở cửa lại kéo đóng cửa quay về, Hàn Quốc nắm những lá bài mạnh về bằng sáng chế và chất bán dẫn, nhưng vẫn đứng ở một nơi mà 82,3% xưởng sản xuất chưa đưa AI vào. Khi cánh cửa của hai cường quốc lần lượt khép lại, câu hỏi Hàn Quốc sẽ từ bỏ điều gì và bảo vệ điều gì trước từng cánh cửa vẫn chưa có lời đáp.

Bên cạnh câu hỏi đó còn có một câu hỏi lớn hơn: chu kỳ này sẽ dừng ở đâu, khi năng lực tự chủ do lệnh trừng phạt tạo ra trở thành sức mạnh để chịu đựng các lệnh trừng phạt tiếp theo, còn thị phần do mở cửa tạo ra lại trở thành lý do để đóng cửa? Việc cuộc thảo luận về đóng cánh cửa đối với các mô hình nhân danh tài sản quốc gia sẽ quay lại dưới dạng quy định thật sự, hay chỉ dừng ở một tín hiệu trong giai đoạn rà soát, là điều không ai có thể khẳng định chắc chắn cho đến khi văn kiện chính thức tiếp theo xuất hiện.

Chương 11. Những giới hạn đã được xác nhận, những rủi ro chưa được kiểm chứng, và Trung Quốc sẽ đưa trí tuệ quân sự cùng AGI đi xa đến đâu

Ở cuối cùng của tài liệu mục lục, có một mục không xuất hiện trong bất kỳ chương nào khác. Tiêu đề của mục đó là "Những ranh giới không được vượt qua khi viết". Sau đó là năm dòng ngắn. Không viết vai trò điều phối của He Lifeng như một sự thật đã được xác nhận. Không viết rằng chip Ascend và chip do Mỹ sản xuất có cùng hiệu năng. Không khẳng định, chỉ dựa trên siêu dữ liệu video mà không có bản chép lời, rằng đến năm 2035 Quân Giải phóng Nhân dân sẽ đạt được trí tuệ hóa, vũ khí robot và vị thế dẫn đầu về AI. Không coi những mục có cách ghi gốc bằng tiếng Trung được suy đoán như thể chúng đã chắc chắn. Không đặt các phát biểu từ nguồn YouTube trong dấu ngoặc kép.

Đó là những ranh giới mà người viết cuốn sách tự vạch ra cho mình. Sách thường không cho độc giả thấy những ranh giới như vậy. Chúng giữ chúng lặng lẽ ở phía sau và chỉ đưa những câu văn đã được gọt giũa ra phía trước. Chương này đưa các ranh giới ấy ra trước mắt. Mục đích là giải thích vì sao phần chính của cuốn sách không dựng các chương sự kiện riêng cho hai câu hỏi mà người ta thường đặt ra về trí tuệ nhân tạo Trung Quốc: AI làm chiến tranh thay đổi ra sao, và liệu một trí tuệ vượt con người có đang đến hay không. Trong các chương trước, chúng ta đã đi theo văn kiện của chính phủ và công bố của doanh nghiệp, rồi thấy khoảng cách giữa các cơ quan nhà nước và những công ty dẫn đầu tạo ra tốc độ như thế nào.

Chúng ta đã lần theo một cấu trúc trong đó Quốc vụ viện và Ủy ban Phát triển và Cải cách Quốc gia đặt ra phương hướng, Cục Quản lý Không gian mạng Trung Quốc thẩm định mô hình, nội dung và thủ tục đăng ký, Bộ Công nghiệp và Công nghệ Thông tin cùng Cục Quản lý Tiêu chuẩn hóa Trung Quốc đặt ra các tiêu chuẩn an toàn, còn Cục Dữ liệu Quốc gia chuyển các tập dữ liệu sang ngôn ngữ của chính sách công nghiệp. Chương này bàn về những gì có thể và không thể xác nhận phía sau tốc độ do cấu trúc ấy tạo ra, và tài liệu chưa được xác nhận nên được xử lý ra sao.

Cần nói trước một điểm. Chương này có tính chất khác với các chương trước. Nếu các chương trước tích lũy những sự thật đã được xác nhận trong văn kiện chính phủ và công bố của doanh nghiệp, thì chương này bắt đầu ở rìa của khối sự thật ấy. Rìa là nơi tài liệu trở nên mỏng hơn, các tuyên bố chạy trước sự thật, còn kỳ vọng và tuyên truyền thay thế cho văn kiện. Trong các câu chuyện về trí tuệ nhân tạo, những chủ đề tạo ra nhiều hào hứng nhất lại thường tụ lại ở rìa này. Chiến tranh tương lai và trí tuệ vượt con người là hai trong số đó.

Một cuốn sách chỉ có hai cách xử lý ranh giới này. Nó có thể thổi phồng tư liệu mỏng thành một chương trông có vẻ hợp lý, hoặc có thể thừa nhận tư liệu mỏng đúng là mỏng và chỉ ghi lại những gì đã được xác nhận. Phần chính văn đã chọn cách thứ hai. Tôi sẽ giải thích lựa chọn đó

qua ba điểm: vì sao các tuyên bố chưa được xác nhận không được biến thành các chương sự kiện, vì sao những dấu bóng đã được xác nhận lại được ghi nhận như phản chứng, và sự kiện với tuyên bố được tách ra như thế nào trong đồng tư liệu.

Cụm từ quân sự thông minh hóa thường xuất hiện trên tiêu đề như thể quá trình ấy đã hoàn tất. Bức tranh được vẽ ra là trí tuệ nhân tạo chọn mục tiêu, drone tự đưa ra phán đoán, và robot di chuyển theo đội hình. Muốn biến bức tranh đó thành một chương sự kiện, chúng cứ phải nâng đỡ được nó. Nhưng trong các tư liệu dùng cho phần chính văn, phần lớn những câu có thể nâng đỡ bức tranh ấy lại dựa trên siêu dữ liệu video không có bản chép lời. Y011 được ghi là một video nói về năng lực trí tuệ nhân tạo của Trung Quốc từ góc nhìn an ninh quốc gia Hoa Kỳ, nhưng tính đến tháng 7 năm 2026 vẫn chưa thu được bản chép lời, và bản thân tệp chỉ là một ghi chú siêu dữ liệu gồm tiêu đề, kênh và mô tả công khai.

Nếu một văn bản như thế rút ra một số hiệu đơn vị, năm triển khai, hoặc loại vũ khí cụ thể rồi đưa vào phần chính văn như sự kiện đã chắc chắn, nó trao uy quyền của chữ in cho một câu không có chứng cứ. Cuốn sách này sẽ không làm vậy. Cùng một logic ấy giải thích vì sao mục lục vạch ranh giới quanh quá trình thông minh hóa của People's Liberation Army vào năm 2035, vũ khí robot, và vai trò lãnh đạo về AI. Ba chủ đề này nằm dưới một quy tắc: không đưa ra khẳng định chắc chắn dựa trên siêu dữ liệu video nếu không có bản chép lời.

Những gì có thể xác nhận thì hẹp và phải được xử lý thận trọng. Sự kiện nền tảng rằng Hoa Kỳ và Trung Quốc đều đang làm việc với trí tuệ nhân tạo trong quốc phòng xuất hiện trong nhiều nguồn. Các báo cáo so sánh năng lực cạnh tranh của Hoa Kỳ và Trung Quốc trình bày bức tranh lớn hơn: cả hai nước đều coi AI là một trục cạnh tranh quốc gia. Nhưng chính những báo cáo ấy không nói rằng AI quân sự đã thay thế hoàn toàn sức mạnh chiến đấu. Vận hành thử nghiệm và công việc hạ tầng cơ bản khác với hoàn tất trên chiến trường. Hoạt động quân sự đòi hỏi độ tin cậy vật lý với mức dung sai lỗi rất thấp, và vài màn trình diễn có kiểm soát chưa đủ để cho thấy độ tin cậy ấy đã được bảo đảm.

Vì vậy, chủ đề này vẫn là một giới hạn đã được xác nhận. AI đang đi vào quốc phòng, nhưng các tư liệu hiện có chưa cho phép đưa ra phán đoán chắc chắn về mức độ trưởng thành của nó hoặc phạm vi triển khai thực tế. Rủi ro chưa được xác nhận bắt đầu từ đây. Nói rằng một điều chưa được xác nhận không có nghĩa là không có nguy hiểm. Việc thiếu kiểm chứng, trong khi các quốc gia vẫn tiếp tục tiến nhanh, tự nó đã là một rủi ro. Phần chính không phóng đại cũng không xóa bỏ rủi ro đó. Nó ghi nhận rủi ro đúng như bản chất của nó: chưa được xác nhận.

Ở đây cần thêm một sự phân biệt nữa. Từ chối biến AI quân sự thành một chương gồm những sự kiện đã ổn định không có nghĩa là phủ im lặng lên chủ đề này. Điều đó có nghĩa ngược lại. Nhiệm vụ của chương này là tách những gì vẫn còn ở giai đoạn trình diễn khỏi những gì đã hoàn tất ngoài thực địa, và tách các năm mục tiêu trên lộ trình khỏi những triển khai đã thật sự diễn ra. Nếu năm mục tiêu của một lộ trình được viết như thể đó là ngày hoàn tất, độc

giả sẽ đọc một tương lai chưa tới như thể nó đã hiện diện. Một câu đặt ra mục tiêu và một câu báo cáo thành tựu dùng những thì khác nhau.

Câu thứ nhất diễn đạt một ý định. Câu thứ hai báo cáo một việc đã làm xong. Khi hai câu đứng cạnh nhau với cùng cỡ chữ, độc giả dễ nhầm ý định thành báo cáo. Phần chính chọn cách ngăn sự lẫn lộn đó về thì. Phương hướng vẫn là phương hướng, mục tiêu vẫn là mục tiêu, và những triển khai đã được xác nhận vẫn là những triển khai đã được xác nhận. Khi các vị trí đó được giữ tách biệt, bức tranh gây ấn tượng mạnh trở nên mỏng hơn, nhưng những câu còn lại chỉ mang đúng sức nặng mà nguồn tư liệu có thể chịu được.

Các tuyên bố về trí tuệ vượt quá con người, nói cách khác là trí tuệ nhân tạo tổng quát, cũng được xử lý theo cách tương tự. Lý do đầu tiên khiến chủ đề này khó trở thành một chương gồm những sự kiện đã ổn định là ý nghĩa của chính thuật ngữ đó khác nhau giữa các quốc gia. Các thảo luận ở U.S. thường hình dung một hệ thống vượt con người trong mọi lĩnh vực nhận thức, đặt trọng tâm vào cuộc đo sức trí tuệ giữa con người và máy móc. Ngôn ngữ chính sách của Trung Quốc không đi theo nguyên xi câu chuyện đó. Nghiên cứu về quản trị AI của Trung Quốc chỉ ra rằng các nhà hoạch định chính sách Trung Quốc không chỉ xem AI là một công cụ phát triển.

Họ nhìn trí tuệ nhân tạo như một công nghệ có thể làm rung chuyển an ninh trong chính trị, kinh tế và xã hội nếu bị quản lý kém. Bên dưới cách nhìn đó là quan niệm an ninh quốc gia tổng thể, lấy an ninh của nhân dân làm đường hướng chỉ đạo và đặt an ninh chính trị làm nền tảng. Trong khung này, thiên lệch thuật toán hoặc sự lệch mục tiêu không chỉ được xếp vào lỗi kỹ thuật, mà còn là rủi ro mang tính hệ thống.

Khác biệt này không phải là chơi chữ. Nếu câu chuyện của Mỹ hướng tới cuộc cạnh tranh về tổng lượng năng lực nhận thức, ngôn ngữ chính sách của Trung Quốc lại nghiêng về mục tiêu phát triển và việc điều chỉnh mô hình cho phù hợp với các tình huống sử dụng trong từng ngành. Hai bên nói về cùng một trí tuệ nhân tạo, nhưng một bên nói về các hệ thống vượt con người, còn bên kia nói về những công cụ nâng năng suất sản xuất. Khi hai ngôn ngữ này chồng lên nhau và lưu hành như thể chúng là một thuật ngữ, người đọc rất dễ nhầm các mục tiêu khác nhau thành cùng một thành tựu.

Đó là lý do cuốn sách này không mô tả trí tuệ nhân tạo tổng quát như một tương lai đã được định sẵn. Khi một thuật ngữ có nghĩa chưa ổn định được đặt vào vị trí lẽ ra dành cho một sự kiện đã được xác nhận, nghĩa của câu sẽ thay đổi tùy theo quốc gia của người dùng thuật ngữ đó. Nguyên tắc nêu trong tài liệu từ khóa cốt lõi, tức tránh những cách diễn đạt có nghĩa thay đổi theo quốc gia của người nói, cũng áp dụng ở đây. Trí tuệ nhân tạo tổng quát là một ví dụ tiêu biểu của vấn đề ấy.

Trong số các công trình được nghiên cứu đó trích dẫn có một nhan đề so sánh trí tuệ nhân tạo tổng quát với chương trình 'hai bom, một vệ tinh' trước đây, tức việc phát triển bom nguyên

tử và vệ tinh nhân tạo. Điều này không cho thấy mức độ trưởng thành của công nghệ, mà cho thấy bối cảnh trong đó công nghệ được tiếp nhận. Trong tranh luận chính sách của Trung Quốc, trí tuệ nhân tạo tổng quát ít tồn tại như một đối tượng đã được kiểm chứng bằng điểm chuẩn, mà giống một phần của ngôn ngữ ý thức hệ về sự tồn vong quốc gia và quyền bá chủ hơn. Nếu cuốn sách này mô tả mức độ trưởng thành của trí tuệ nhân tạo tổng quát như một sự thật, nó sẽ rơi vào sai lầm là trao tính chính danh học thuật cho một mục tiêu tương lai chưa được kiểm chứng.

Điều đó sẽ khiến người đọc nhầm những kỳ vọng được tạo ra từ phần chồng lấn giữa câu chuyện mà doanh nghiệp dùng để thu hút đầu tư và câu chuyện mà nhà nước huy động nhân danh an ninh thành những sự thật đã được tài liệu xác nhận. Ở chương trước, chúng ta đã thấy mô hình quản trị kiểu Trung Quốc đặt phát triển và an ninh cạnh nhau. Khi thuật ngữ trí tuệ nhân tạo tổng quát di chuyển trong các văn kiện Trung Quốc giữa một công cụ phát triển và một biểu tượng an ninh, nó mang cùng lúc hai khuôn mặt. Nếu ép hai khuôn mặt ấy thành một sự thật duy nhất, bức tranh sẽ mờ đi.

Vì lý do đó, quân sự hóa thông minh và trí tuệ nhân tạo tổng quát không được dành các mục sự kiện độc lập trong cuốn sách này. Thay vào đó, chúng được đặt vào hai ô: giới hạn đã xác nhận và rủi ro chưa xác nhận. Giới hạn đã xác nhận là những sự kiện hẹp mà tài liệu hiện có có thể nâng đỡ. Các văn kiện xác nhận phương hướng và tham vọng của cả hai công nghệ, nhưng không xác nhận rằng chúng đã hoàn tất trên chiến trường hay thay thế trí tuệ con người. Rủi ro chưa xác nhận là mối nguy sinh ra chính từ việc khoảng trống này vẫn còn trống.

Khi có tài liệu chính thức mới, phần này có thể được viết lại bất cứ lúc nào. Hiện giờ, nó vẫn để mở. Để một khoảng trống mở không giống với việc bịa ra điều gì đó, và văn bản này cố gắng tôn trọng sự khác biệt ấy. Một đoạn văn tự nghĩ ra thứ cần lấp vào khoảng trống có thể trông vững chắc trong lúc đọc, nhưng sẽ sụp đổ nguyên khối khi tài liệu mới xuất hiện. Một đoạn văn để khoảng trống mở có thể tạo cảm giác còn thiếu trong lúc đọc, nhưng khi tài liệu mới đến, các sự kiện có thể được đặt ngay vào đó.

Khi gạt những tuyên bố lớn sang một bên, một cái bóng đã được xác nhận đang hình thành bên dưới chúng trở nên nhìn thấy rõ. Cái bóng đầu tiên là việc làm. Sự bất an mà các nhà hoạch định chính sách Trung Quốc cảm thấy trước trí tuệ nhân tạo và việc làm xuất hiện rõ trong ghi chép của các nhà quan sát. Sự bất an này có gốc rễ lịch sử. Nó kéo dài trở lại ký ức về làn sóng thất nghiệp lớn do cải cách doanh nghiệp nhà nước năm 1999 gây ra. Khi đó, chính phủ nhanh chóng mở rộng giáo dục đại học để giữ thanh niên ở lại trong trường và ngăn người thất nghiệp tràn vào thị trường lao động cùng một lúc.

Ký ức ấy vẫn phủ bóng lên ngôn ngữ chính sách. Truyền thông nhà nước tiếp tục lặp lại rằng trí tuệ nhân tạo sẽ không lấy đi sinh kế của người dân, và Kế hoạch 5 năm lần thứ 15 trình bày học tập suốt đời như một mạng lưới an toàn cho người lao động. Nhưng mạng lưới an toàn này

không bao gồm việc phân phối lại nguồn lực tài khóa. Thay vì mở rộng trợ cấp thất nghiệp, nó gắn với việc bảo người dân tự học kỹ năng mới và tự tìm cách sống sót hơn.

Cần đọc điều này cùng với ngôn ngữ phát triển đã thấy ở chương trước thì bức tranh mới đầy đủ. Chính phủ đang đẩy trí tuệ nhân tạo xuống nền kinh tế thực, mở ra các địa điểm trong sản xuất, giáo dục và tiêu dùng. Cũng chính phủ ấy đang cân nhắc cú sốc thất nghiệp mà tốc độ này có thể tạo ra, nhưng thay vì hấp thụ cú sốc đó bằng tài chính công, họ lại chuyển nó sang việc học tập của từng cá nhân. Bàn tay thúc đẩy phát triển và bàn tay chuyển cú sốc sang cá nhân thuộc về cùng một cơ thể.

Sự bất an này cũng đi vào ngôn ngữ quản lý. Artificial Intelligence Safety and Governance Framework 2.0, được công bố vào tháng 9 năm 2025, đã nói thẳng rằng trong khi phần lợi ích thuộc về vốn, công nghệ và dữ liệu đang tăng lên, giá trị của lao động với tư cách một yếu tố sản xuất đang giảm xuống, và nhu cầu lao động truyền thống đang giảm mạnh. Vì vậy, một tiêu chuẩn quốc gia đã được tạo ra, yêu cầu các công ty tiến hành đánh giá tác động việc làm khi áp dụng trí tuệ nhân tạo tạo sinh.

Tiêu chuẩn này vượt ra ngoài một khuyến nghị và đặt thêm gánh nặng mới lên các công ty. Nó đi theo hướng khiến các công ty đang hưởng lợi từ trí tuệ nhân tạo khó sa thải người lao động ngay lập tức hơn, đồng thời làm chi phí quản lý tăng lên tương ứng. Lập trường hai mặt của chính phủ, vừa tăng tốc phát triển vừa đo cú sốc đối với xã hội, xuất hiện ở đây. Một văn bản mang tên đánh giá tác động việc làm cho thấy các tiêu chuẩn an toàn của Standardization Administration of China đã thấy ở chương trước là ngôn ngữ của chính sách công nghiệp, đồng thời cũng là ngôn ngữ của chính sách lao động. Tiêu chuẩn không chỉ đo mức độ an toàn của công nghệ. Chúng còn cố đo vị trí của những con người mà công nghệ có thể đẩy sang một bên.

Cái bóng thứ hai là thứ lao động đang chống đỡ sự phát triển này từ bên dưới. Để trí tuệ nhân tạo giải quyết vấn đề và nhận diện ảnh, phải có người dạy cho máy, từng bước một, cách tri giác của con người vận hành ra sao. Công việc ấy được gọi là gán nhãn dữ liệu. Nó không phải lao động chân tay, cũng không phải lao động cảm xúc. Nó là lao động nhận thức. Một nghiên cứu học thuật đã đi vào bên trong một công ty dữ liệu trí tuệ nhân tạo để xem lao động gán nhãn dữ liệu bị kiểm soát như thế nào. Quy mô của công ty và vị trí của những người gán nhãn lệch nhau rất xa. Nhóm gán nhãn được khảo sát trong nghiên cứu chỉ gồm một số ít lao động thuê ngoài. Thực tập sinh nhận 100 đến 200 nhân dân tệ mỗi ngày, nhân viên thuê ngoài nhận khoảng 4.000 đến 5.000 nhân dân tệ mỗi tháng, và họ bị loại khỏi các đợt đánh giá hiệu suất. Ở phía trên họ, các lập trình viên thuật toán và nhân viên kiểm thử tạo thành những tầng riêng, rồi vạch ra ranh giới phân công lao động.

Công việc của họ là làm cho máy đọc được sách bài tập giáo dục. Họ vẽ khung quanh hình ảnh của các biểu thức số học, gán nhãn 'bốn phép toán số học', rồi gán nhãn 'giải phương trình'

cho những biểu thức khác. Ban quản lý hạ thấp công việc ấy, nói rằng ngay cả học sinh tiểu học cũng làm được. Nhưng biến năng lực nhận biết tự nhiên của con người thành dữ liệu chuẩn hóa để máy tính đọc được là lao động nhận thức mệt mỏi. Nghiên cứu này nhìn nhận ban quản lý như đang cố ép cả cách nghĩ của người gán nhãn phải khớp với máy.

Nếu những người khác nhau nhìn cùng một hình ảnh theo những cách khác nhau, thuật toán sẽ hỏng. Vì vậy công ty dùng các hướng dẫn gán nhãn chi tiết để đưa phán đoán của con người về cùng một chuẩn cố định. Những chỉ dẫn như vẽ khung theo hình dạng quy định quanh từng dòng chữ, và bảo đảm một khung không nằm hoàn toàn bên trong một khung khác, đã huấn luyện lại mắt và tay của con người để đi theo quy tắc của chương trình.

Khi có lỗi xảy ra trong lúc làm việc, lỗi ấy được nêu công khai trong nhóm chat để người lao động tự sửa. Trong các cuộc phỏng vấn tuyển dụng, công ty nhấn mạnh kỹ năng giao tiếp nhằm lọc bỏ những người hay do dự. Đến giai đoạn nâng tốc độ, công ty còn kìm nén cả những nghi ngờ mà người gán nhãn có về các quy tắc. Nghiên cứu ghi lại một cảnh: một người gán nhãn cảm thấy các hướng dẫn tự mâu thuẫn với nhau, rồi tự nhủ đừng nghĩ quá sâu và vẫn làm theo quy tắc.

Sự kiểm soát này không phải lúc nào cũng diễn ra trơn tru. Khi trưởng nhóm đặt chỉ tiêu hằng ngày theo người làm nhanh nhất và nói rằng sẽ kiểm tra nhật ký công việc, các nhân viên gán nhãn đã tụ tập trong giờ ăn trưa để bàn cách phản ứng. Họ thống nhất rằng mỗi người chỉ hoàn thành đến một số lượng nhiệm vụ nhất định rồi dừng lại, không làm vượt quá mức đó, lấy lý do hình ảnh phức tạp để làm chậm tiến độ. Đó là một hành động phản kháng lặng lẽ của những người bị đối xử như linh kiện máy móc, một nỗ lực giành lại quyền tự chủ. Cảm giác xa lạ với công việc cũng lộ ra trong lời nói. Các nhân viên gán nhãn nói rằng họ chẳng khác gì công nhân trên dây chuyền lắp ráp hay trong một nhà máy điện tử lớn. Một học sinh trường nghề được phân công làm gán nhãn dữ liệu cho xe tự lái nói mắt mình đau và rất mệt, nếu việc đó không phải là yêu cầu để tốt nghiệp thì em sẽ không làm.

Hai tháng sau khi đợt điền dã của nghiên cứu kết thúc, toàn bộ nhân viên gán nhãn trong nhóm đó đã nghỉ việc, và công ty nhanh chóng lấp chỗ trống bằng những người mới. Câu chuyện ở đây là phía sau hiệu suất rực rỡ của các mô hình, vẫn có một cấu trúc tiếp tục vận hành, trong đó lao động ở tầng dưới được thay thế với chi phí rẻ khi công ty lớn lên.

Nếu đặt cảnh này cạnh các văn kiện ở chương trước, khoảng cách sẽ hiện rõ. Cục Dữ liệu Quốc gia nói về dữ liệu bằng ngôn ngữ của chính sách công nghiệp: bộ dữ liệu chất lượng cao và cộng hưởng tham số. Ở tuyến đầu của bộ dữ liệu ấy là đôi tay của những người lao động kiếm 100 yuan mỗi ngày bằng việc vẽ các khung hộp. Trong văn kiện chính sách, bộ dữ liệu trông như những tài nguyên được sắp xếp sạch sẽ. Nhưng ở nơi các tài nguyên đó được tạo ra, con người phải uốn cách nghĩ của mình theo quy tắc, chạy theo chỉ tiêu, và cuối cùng cả một nhóm rời đi chỉ sau hai tháng.

Cũng như phần chính đã lần theo khoảng cách giữa các cơ quan nhà nước và doanh nghiệp, phần này phải lần theo khoảng cách giữa văn kiện chính sách và lao động tại hiện trường. Nếu khoảng cách đó biến mất, dữ liệu sẽ được đọc như thể nó tự xuất hiện. Nếu khoảng cách đó vẫn được nhìn thấy, chúng ta có thể thấy lao động của ai đã trao giá trị cho dữ liệu. Bên cạnh bất kỳ câu nào nói về tốc độ của trí tuệ nhân tạo Trung Quốc, lập luận phản biện này phải ghi lại cái giá mà lực lượng lao động nâng đỡ tốc độ ấy đã trả.

Cái bóng thứ ba là chi phí của quản lý và kiểm soát. Kiểu quản trị Trung Quốc, vốn đặt phát triển và an ninh đi cùng nhau, tạo ra các ngưỡng cửa cho doanh nghiệp. Các dự thảo quy định về trí tuệ nhân tạo sinh đặt lên những doanh nghiệp cung cấp dịch vụ nhiều nghĩa vụ, trong đó có đăng ký và quản lý nội dung. Những nghĩa vụ này là công cụ bảo đảm an toàn, nhưng chúng cũng là chi phí mà doanh nghiệp phải gánh. Một công ty càng phải đi qua nhiều thủ tục trước khi đưa một mô hình ra phục vụ công chúng, ngưỡng cửa ấy càng trở nên nặng nề, trước hết là với các công ty nhỏ hơn. Chương này nhắc lại rằng tốc độ phát triển đã thấy ở chương trước được tạo ra trên chính ngưỡng cửa đó.

Từ góc nhìn phản biện, cần đọc lại hai mặt trong vai trò của Cyberspace Administration of China khi thẩm tra mô hình và nội dung: một rào chắn vì an toàn và một chi phí đối với doanh nghiệp. Mỗi thủ tục mà dự thảo quy định yêu cầu chỉ là một dòng văn bản pháp lý, nhưng với những người phải tuân thủ, nó trở thành công việc thực tế cần nhân sự, thời gian và thẩm định. Nếu các chương trước đọc những quy định này từ góc độ cân bằng giữa phát triển và an ninh, chương này đọc lại cùng các quy định ấy từ góc độ gánh nặng doanh nghiệp. Tùy theo góc đọc, cùng một văn bản có thể trở thành bức tường bảo vệ hoặc một ngưỡng cửa.

Sự kiểm soát ấy còn có một mặt khác, vượt ra ngoài lao động và thị trường. Một báo cáo của một tổ chức nhân quyền quốc tế về quyền lực đàn áp của trí tuệ nhân tạo bàn đến việc các quyền tự do dân sự bị thu hẹp ra sao khi AI được dùng làm công cụ giám sát và kiểm duyệt. Khi ngôn ngữ phát triển chảy trơn tru qua các văn kiện chính sách công nghiệp, báo cáo này ghi lại từ phía bên kia cách cùng một công nghệ vận hành tại những nơi giám sát và kiểm duyệt. Lý do phần chính đặt báo cáo này bên cạnh câu chuyện về tốc độ của trí tuệ nhân tạo Trung Quốc như một lập luận phản biện là vì nếu mặt trái của tốc độ bị xóa đi, bức tranh chỉ còn một nửa.

Chỉ khi ngôn ngữ phát triển và hồ sơ về đàn áp được đặt cạnh nhau, độc giả mới có thể tự cân nhắc vấn đề. Đây là nơi phần này nối với cặp song hành phát triển và an ninh ở chương trước. Chương này gom vào một chỗ hai hướng mà kiểu quản trị nói đồng thời về phát triển và an ninh hiện ra: với doanh nghiệp, nó là chi phí tuân thủ; với công dân, nó là sự thu hẹp tự do. Cùng một công nghệ, cùng các cơ quan nhà nước, cùng các quy định ấy được đọc như công cụ phát triển trong ngôn ngữ công nghiệp, và như công cụ kiểm soát trong ngôn ngữ nhân quyền. Cuốn sách phải mang cả hai cách đọc ấy nếu không muốn chỉ kể một nửa câu chuyện.

Sau khi ghi lại những vùng mờ đã được kiểm chứng này, câu hỏi còn lại thuộc về phương pháp. Khi các sự kiện đã được xác nhận và những tuyên bố chưa được kiểm chứng nằm chung trong cùng một tập tài liệu, một cuốn sách nên tách hai loại ấy như thế nào? Cách làm ở đây là đặt mỗi sự kiện vào đúng vị trí của nó theo trọng lượng của nguồn. Khung lập luận dựa trên văn bản chính thức và công bố của doanh nghiệp. Quy định quản lý của chính phủ, hồ sơ chứng khoán và số liệu thống kê từ các tổ chức nghiên cứu có công khai phương pháp được xếp ở tuyến đầu của căn cứ sự kiện. Các cơ quan truyền thông lớn và báo cáo phân tích cung cấp bối cảnh, nhưng tự thân chúng không trở thành những khẳng định đã chốt trong các chương sự kiện. Blog cá nhân, bình luận không ký tên và bài đăng chưa được kiểm chứng bị loại khỏi nền tảng dùng để đưa ra tuyên bố sự kiện.

Một con số chính xác lấy từ một bài đăng ẩn danh về một chỉ số mà công ty chưa từng công bố có thể trông có vẻ đúng, nhưng nó không có căn cứ, nên bị bỏ ra ngoài. Quy tắc viết yêu cầu xây dựng chứng cứ theo thứ tự văn bản chính phủ, công bố của doanh nghiệp, báo cáo nghiên cứu và báo cáo chính sách, vì thứ tự ấy giữ được hệ thống trọng lượng của nguồn. Một con số được ghi tới chữ số thập phân tạo cho độc giả cảm giác tin cậy. Khi cảm giác tin cậy ấy không có nền tảng, độ chính xác trở thành lớp bao bì che giấu điều sai.

Quy tắc xử lý tư liệu video trong văn bản này rất chặt. Các tệp YouTube được liệt kê trong phần tư liệu là ghi chú siêu dữ liệu, không phải bản chép lời. Tính đến tháng 7 năm 2026, chúng chỉ chứa tên kênh, tiêu đề và phần mô tả công khai, không có bản chép lời đã được bảo đảm. Các video về tổng mức đầu tư toàn quốc cho hạ tầng trí tuệ nhân tạo, các video về những vòng gọi vốn nối tiếp của các công ty trí tuệ nhân tạo Trung Quốc, và các video đánh giá năng lực từ góc nhìn an ninh quốc gia của Hoa Kỳ đều vẫn ở tình trạng này. Vì vậy, văn bản này không đưa các con số từ những video ấy vào hồ sơ chứng cứ, cũng không trích dẫn phát biểu từ chúng trong dấu ngoặc kép. Các tệp này chỉ còn là manh mối để khảo sát, cho thấy những chủ đề nào đang được bàn tới và ở đâu. Khi cần, chúng được đánh dấu là đối tượng phải kiểm tra thêm, nhưng chúng không tạo ra các khẳng định đã chốt.

Ngay cả khi một người trong video khẳng định ưu thế của một công nghệ nào đó hoặc tác động của một chính sách, khẳng định ấy vẫn nằm trong nhóm tuyên bố chưa được kiểm chứng. Chính các nguyên tắc do bộ tư liệu nêu ra cũng đi theo hướng ấy. Sự kiện về chính sách, doanh nghiệp và công nghệ lấy văn bản chính thức và tài liệu đăng ký làm chứng cứ ưu tiên, trong khi tư liệu video bị hạ xuống thành tài liệu hỗ trợ, chỉ cho thấy sự tồn tại, tiêu đề và phạm vi chủ đề của chúng. Nếu cần một trích dẫn chính xác từ video, nguồn gốc phải được kiểm tra lại, và trước khi việc kiểm tra ấy hoàn tất, phát biểu đó không được chuyển sang như một sự kiện. Con số càng lớn, chẳng hạn tổng mức đầu tư toàn quốc, quy tắc này càng phải được áp dụng nghiêm. Những con số lớn rất dễ được trích dẫn, và một khi đã được trích dẫn, chúng đi vào các bài viết khác rồi đông cứng lại như thể là sự thật.

Kỷ luật tương tự cũng được áp dụng khi xử lý các báo cáo kỹ thuật từ doanh nghiệp và viện nghiên cứu. Một tuyên bố trong phần mở đầu của báo cáo chỉ được xem là sự thật khi nó khớp với các kết quả ở phần sau. Các nhà phát triển thường trộn những mục tiêu đầy tham vọng vào báo cáo của họ, chẳng hạn như một trí tuệ tương lai sẽ giải quyết các vấn đề của nhân loại. Những mục tiêu như vậy có thể được nêu như động cơ, nhưng ranh giới phải được vạch rõ: chúng chưa được đạt tới. Công việc này cũng bao gồm việc nêu ra các giả định và giới hạn ẩn sau những kết quả mạnh. Phần nói về giới hạn, nơi người viết hỏi liệu một kết quả có còn đúng trong các điều kiện khác hay không, không được để trống. Chính kỷ luật này là lý do phần bàn về benchmark trong chương trước được viết một cách thận trọng.

Văn bản này không viết rằng một con chip do Trung Quốc sản xuất có hiệu năng ngang với một con chip do U.S. sản xuất, và cũng không tuyên bố rằng một mô hình nào đó đã vượt qua mô hình hàng đầu của U.S., bởi ngay cả các thước đo cho diffusion cũng chưa được chuẩn hóa. Việc cố định thứ hạng trong những điều kiện như vậy gần với việc đưa ra một tuyên bố hơn là nêu một sự thật. Mục lục đặt riêng một đường cấm đối với các tuyên bố về ngang bằng hiệu năng giữa chip Ascend và chip do U.S. sản xuất cũng vì lý do đó. Điểm benchmark thay đổi khi điều kiện thay đổi. Nếu bộ bài toán dùng để đo khác nhau, hoặc nếu thang đo dùng để đếm khác nhau, cùng một mô hình có thể nhận một thứ hạng khác. Khi những con số chưa ổn định như vậy bị biến thành thứ hạng cố định, một phép đo nhất thời sẽ đông cứng thành một sự thật lâu dài.

Xử lý xung đột giữa các nguồn cũng là một phần của phương pháp này. Các nguồn có thẩm quyền thường đưa ra những con số khác nhau cho cùng một vấn đề. Khi điều đó xảy ra, văn bản không âm thầm chọn một con số rồi in ra. Thay vào đó, nó nói rõ các nguồn thống nhất, mâu thuẫn, hay ngay từ đầu chỉ có một nguồn. Khi chúng mâu thuẫn, văn bản đưa ra một khoảng bao trùm giá trị nhỏ nhất và lớn nhất, kèm theo phương pháp dùng để tạo ra khoảng đó. Nó cũng để lại một dấu hiệu cho thấy giá trị ấy thuộc phạm vi ước tính, không phải sự thật đã chốt: (estimate). Nguyên tắc tương tự được áp dụng khi ngày tháng trong dòng thời gian khác nhau giữa các nguồn. Văn bản không xóa một ngày, mà ghi cả hai ngày cạnh nhau.

Khi các nguồn không thống nhất về thời điểm một đạo luật liên quan đến dữ liệu có hiệu lực hoặc thời điểm một văn kiện quy hoạch được công bố, văn bản ghi cả hai ngày thay vì chọn một ngày làm kết luận cuối cùng, để độc giả tự thấy sự không nhất quán. Từ cách ghi song song đó, độc giả có thể biết điều gì vẫn chưa được xác lập. Một văn bản âm thầm chỉ chọn một con số có thể trông trơn tru, nhưng độc giả không có cách nào nhìn thấy lựa chọn bị giấu sau sự trơn tru ấy. Một văn bản giữ lại cả hai con số có thể trông gồ ghề, nhưng nó trao cho độc giả chất liệu cần thiết để tự phán đoán.

Cùng một phương pháp ấy chi phối cách xử lý các thuật ngữ gốc tiếng Trung. Nhiều tên công ty, thuật ngữ chính sách và tên cơ quan nhà nước vẫn chưa có dạng tiếng Trung gốc được xác

nhận. Nếu văn bản in những mục ấy như thể dạng gốc của chúng đã được chốt, một dạng chưa kiểm chứng sẽ trở thành sự thật nhờ sức nặng của chữ in. Vì vậy, văn bản chỉ ghi tiếng Trung gốc đối với những tên đã được xác nhận, và không dùng các dạng tạm suy đoán như thuật ngữ đã ổn định. Quy tắc ấy cũng áp dụng với chức danh của một cá nhân. Việc không coi các vai trò điều phối dựa trên tin đồn là sự thật đã xác nhận phản ánh sự thận trọng cần có khi xử lý tên tổ chức và vị trí của con người.

Các tổ chức chỉ được mô tả đến phạm vi vai trò đã được xác nhận trong tài liệu, còn con người chỉ được mô tả đến phạm vi chức danh đã được xác nhận. Ngay cả khi nhiều nơi nói rằng một người nào đó điều phối công việc trí tuệ nhân tạo trong một ủy ban nào đó, văn bản vẫn không gắn vị trí ấy vào tên người đó chừng nào nhận định ấy còn ở mức tin đồn. Một khi tên và chức danh đã được đặt cạnh nhau trên trang in, cặp ghép ấy sẽ đi vào các bài viết khác, bất kể nó từng được xác nhận hay chưa.

Đến đây, lý do phương pháp này thuộc về chương này trở nên rõ ràng. Việc không biến trí tuệ hóa quân sự và trí tuệ nhân tạo tổng quát thành các chương mang tính sự kiện, việc đặt việc làm, lao động gắn nhãn dữ liệu và chi phí tuân thủ quy định cạnh nhau như bằng chứng đối trọng, và việc để siêu dữ liệu video ra ngoài phần trích dẫn không phải là những quyết định rời rạc. Chúng xuất phát từ một kỷ luật: chỉ viết điều đã được xác nhận trong đúng phạm vi nó được xác nhận, nói rõ rằng điều chưa được xác nhận vẫn là chưa được xác nhận, và không ép khoảng trống giữa hai bên phải khép lại. Điều này cũng giải thích vì sao văn bản đã theo dõi khoảng cách giữa các cơ quan nhà nước và các doanh nghiệp dẫn đầu ngay từ đầu. Khoảng cách ấy tạo ra tốc độ, nhưng nó cũng đánh dấu ranh giới giữa điều tài liệu xác nhận và điều vẫn là kỳ vọng hoặc tuyên truyền.

Định hướng do Quốc vụ viện và Ủy ban Cải cách và Phát triển Quốc gia đặt ra, quy trình đăng ký của Cục Quản lý Không gian mạng Trung Quốc, các tiêu chuẩn của Bộ Công nghiệp và Công nghệ Thông tin cùng Cục Tiêu chuẩn hóa Trung Quốc, và các bộ dữ liệu của Cục Dữ liệu Quốc gia đều nằm ở những vị trí mà tài liệu có thể xác nhận. Trên nền tảng đó, tốc độ vận hành của các công ty có thể được kiểm chứng, nhưng tuyên bố rằng tốc độ này trực tiếp dẫn đến hoàn tất về quân sự hoặc đến trí tuệ vượt con người vẫn nằm ngoài tài liệu. Việc kẻ một đường ranh giữa sự kiện nằm trong tài liệu và kỳ vọng nằm ngoài tài liệu là vai trò mà phần chính của chương này đã đảm nhận.

Hãy trở lại với năm dòng được viết ở cuối cùng của tài liệu mục lục. Những dòng cảm đoán ấy có thể trông giống các quy tắc làm cuốn sách hẹp lại, nhưng thực ra chúng là điều kiện khiến cuốn sách đáng tin. Vì có những câu chúng ta đã chọn không viết, những câu chúng ta đã chọn viết mới có trọng lượng. Các chương trước đi theo ngôn ngữ của phát triển, còn chương này ghi lại mặt sau của ngôn ngữ ấy như bằng chứng ngược. Vì vậy, chúng tạo thành mặt trước và mặt sau của cùng một cuốn sách. Chỉ đọc một mặt, ta chỉ thấy tốc độ. Chỉ đọc mặt kia, ta chỉ thấy

bóng tối.

Chỉ khi đặt hai mặt cạnh nhau, nhiệt độ thật mới hiện ra ở nơi các cơ quan nhà nước, doanh nghiệp và dữ liệu đan vào nhau. Câu hỏi còn lại được để mở cho người đọc. Đến ngày những tài liệu chính thức mới xuất hiện và rủi ro chưa được xác nhận hôm nay trở thành sự thật đã được kiểm chứng, chúng ta sẽ viết lại câu ấy dựa trên căn cứ nào? Và ngay cả khi đó, như bây giờ, liệu chúng ta có thể đưa ranh giới giữa điều đã được xác nhận và điều chưa được xác nhận ra phía trước, rồi cho người đọc thấy hay không?

Chương 12. Những lựa chọn của Hàn Quốc: Vẽ lại khoảng cách giữa chính phủ, doanh nghiệp và dữ liệu

Hãy hình dung ba tài liệu được xếp chồng trên bàn. Ở dưới cùng là một báo cáo của Samjong KPMG so sánh năng lực cạnh tranh về AI giữa United States và China. Lật trang bìa, một bức tranh quen thuộc hiện ra. Một bên là đoạn viết rằng chỉ riêng trong quý I năm 2025, các startup AI của U.S. đã thu hút lượng lớn vốn đầu tư mạo hiểm tư nhân. Bên cạnh đó là đoạn viết rằng China đang bù vào sự co lại của đầu tư tư nhân bằng vốn nhà nước và hỗ trợ quá trình chuyển sang sản xuất bán dẫn Chinese-made.

Phía trên tài liệu đó là một văn bản giải thích chiến lược AI+ của chính phủ China, và trên cùng là một báo cáo tóm tắt các xu hướng chính sách AI gần đây của China. Khi một nhà hoạch định chính sách Korean đặt ba trang này cạnh nhau, câu hỏi đầu tiên không nên là có sao chép quốc gia đó hay không. Câu hỏi phải là các cơ quan chính phủ, doanh nghiệp và hạ tầng dữ liệu của chúng ta hiện đang cách nhau bao xa, và chúng ta nên thiết lập lại khoảng cách ấy như thế nào. Chương này giữ chặt một câu hỏi đó.

Mở các văn kiện chính sách của China ra, điều đầu tiên hiện lên khá rõ. Họ không nói về AI chỉ như một chatbot hay một công cụ tìm kiếm. Văn kiện chiến lược AI+ dùng thứ ngôn ngữ đưa AI xuống tận sàn nhà máy và nền móng của nền kinh tế thực. Báo cáo tóm tắt các xu hướng chính sách gần đây cũng cho thấy cách chính phủ biến các vấn đề công nghiệp thành ngôn ngữ chính sách, theo một đường nối từ 2017 Next Generation AI Development Plan đến chiến lược AI+, rồi đến nền kinh tế thông minh của 15th Five-Year Plan. Chính phủ đặt hướng đi, hạ tầng dữ liệu do nhà nước xây dựng cung cấp khung vật lý, và các doanh nghiệp vận hành nền kinh tế thực trên nền đó.

Việc ba yếu tố này không di chuyển tách rời, mà tiếp tục như một dòng chảy thống nhất, nằm bên dưới tốc độ của China. Korea nên nhìn không phải vẻ hào nhoáng của từng công nghệ riêng lẻ, mà là cách China xâu chính sách, hạ tầng và doanh nghiệp thành một tuyến thực thi liền mạch. Thay vì đưa ra những lời nói bay bổng về artificial general intelligence, China biến các vấn đề tại hiện trường công nghiệp thành câu chữ trong văn kiện chính sách. Năng lực dịch chuyển đó vẫn còn mỏng trong các văn kiện chính sách của Korean.

Chúng ta cần nhìn kỹ hơn một chút xem việc chuyển dịch này trông như thế nào. Văn kiện chiến lược AI+ chứa một kế hoạch gắn AI với nhiều ngành. Nó gắn AI với sản xuất, giáo dục, tiêu dùng và quản trị hành chính. Mỗi lần như vậy, chính phủ viết lại các vấn đề của ngành đó bằng chính ngôn ngữ của ngành đó. Văn kiện chính sách ghi rõ quy trình nào cần được nâng lên mức nào, những bộ dữ liệu nào cần được thu thập và thu thập ra sao. Như văn kiện về xu hướng chính sách gần đây chỉ ra, kiểu kế hoạch này không đến từ đầu ngón tay của một bộ duy nhất.

Nó đi xuống như một cơ chế phối hợp nhiều tầng, chạy từ Quốc vụ viện đến chính quyền địa phương.

Đó là một cấu trúc trong đó nơi đặt ra phương hướng và nơi thực thi đã được nối với nhau ngay trong văn kiện. Điều Hàn Quốc có thể học không phải là quy mô ngân sách, mà là độ chính xác của việc chuyển dịch này. Năng lực biến ngôn ngữ của hiện trường thành câu chữ của chính sách có thể được học riêng, không phụ thuộc vào quy mô nguồn lực tài chính.

Nếu mở cấu trúc phối hợp này thêm một chút, ta có thể thấy một nơi đặt ra phương hướng và một nơi khác đặt ra ngưỡng kiểm soát. Quốc vụ viện và Ủy ban Phát triển và Cải cách Quốc gia đứng ở vị trí xác lập phương hướng lớn, trong khi Cục Quản lý Không gian mạng Trung Quốc quản lý những mô hình nào có thể được phát hành dưới dạng dịch vụ công thông qua thủ tục đăng ký và thẩm định. Bộ Công nghiệp và Công nghệ Thông tin cùng Cục Quản lý Tiêu chuẩn hóa Trung Quốc phụ trách các quy tắc an toàn và tiêu chuẩn công nghiệp. Cục Dữ liệu Quốc gia, cơ quan xử lý dữ liệu, đã bước vào dòng chảy này và biến dữ liệu rải rác thành ngôn ngữ của chính sách công nghiệp.

Ở đây có một điểm cần thận trọng. Không nên coi những câu chuyện về một cá nhân cụ thể cầm đầu chỉ huy cho sự phối hợp này là sự thật đã được xác nhận. Các văn kiện mà đoạn viết này dựa vào cho thấy cách bố trí các cơ quan và thủ tục, chứ không phải những câu chuyện anh hùng về cá nhân. Nơi đặt ra phương hướng, nơi đặt ra ngưỡng kiểm soát và nơi tập hợp dữ liệu được liên kết với nhau trong các văn kiện. Chính sự sắp xếp đó là khung vận hành của quản trị kiểu Trung Quốc.

Bài học thứ hai đáng rút ra là cách nhìn rằng hạ tầng cũng là vấn đề của không gian và quyền lực. Ví dụ tiêu biểu được nêu trong báo cáo của Samjong KPMG là dự án Eastern Data, Western Computing. Khi dữ liệu đổ ra từ miền đông tập trung công nghiệp, công việc tính toán được chuyển về miền tây, nơi đất đai và điện năng dồi dào hơn, để xử lý. Đó là một cuộc phân bổ lại hạ tầng ở quy mô quốc gia. Câu hỏi đặt trung tâm dữ liệu ở đâu không chỉ được để mặc cho từng doanh nghiệp chọn địa điểm. Trung Quốc vẽ bản đồ này cùng với lưới điện và đất đai của cả lục địa. Hàn Quốc có lãnh thổ hẹp và đang gặp nút thắt trong việc bảo đảm nguồn điện.

Đã có tranh luận về việc đưa trung tâm dữ liệu về các địa phương, nhưng một thiết kế cấp quốc gia đặt nguồn điện, mạng thông tin liên lạc và nhu cầu công nghiệp lên cùng một bản đồ vẫn còn rời rạc. Hàn Quốc không thể sao chép nguyên xi bức tranh của Trung Quốc, nhưng có thể tiếp nhận cách nhìn hạ tầng như sự kết hợp giữa không gian, điện năng và nhu cầu. Các báo cáo sắp xếp dòng chảy của công nghiệp Hàn Quốc cũng chỉ ra rằng cách nhìn tính toán và điện năng cùng nhau có sức nặng lớn trong cuộc cạnh tranh năm 2026.

Điểm thứ ba là con đường của các mô hình mở và hiệu quả chi phí. Báo cáo của U.S.-China Economic and Security Review Commission tóm tắt lựa chọn của Trung Quốc dưới các biện pháp kiểm soát xuất khẩu chip tiên tiến của Mỹ như một lựa chọn đi theo con đường khác, thay

vì mở rộng bằng chi tiêu vốn không giới hạn. Báo cáo đánh giá rằng DeepSeek đã điều chỉnh kiến trúc của mình và tạo ra hiệu năng có thể dùng được với chi phí thấp hơn nhiều so với các mô hình hàng đầu của Mỹ. Cũng báo cáo đó nói rằng Qwen của Alibaba đã mở rộng hệ sinh thái mô hình mở, với nhiều mô hình phái sinh tỏa ra trên các kho lưu trữ công khai.

Ở đây báo cáo đưa ra một khung nhìn. Báo cáo xem chiến lược mở của Trung Quốc là một cách tăng quyền kiểm soát công nghiệp. Khi các mô hình được mở, nhà phát triển trên khắp thế giới xây dựng các mô hình phái sinh trên nền tảng đó, và dòng chảy này quay trở lại dưới dạng phản hồi, làm mở rộng các tiêu chuẩn và công cụ của các công ty Trung Quốc.

Đối với các công ty Hàn Quốc không có nguồn vốn khổng lồ, việc nhảy thẳng vào cuộc đua quy mô của các mô hình đóng khổng lồ có xác suất thành công rất thấp. Con đường thực tế hơn là xây dựng trên nền các mô hình mở mạnh, rồi cạnh tranh bằng các giải pháp doanh nghiệp nhẹ hơn, được điều chỉnh cho những lĩnh vực mà Hàn Quốc có thế mạnh, như sản xuất và y học. Điều Trung Quốc cho thấy không phải là tài làm ra mọi thứ với giá rẻ, mà là thái độ thiết kế đường vòng khi bị ràng buộc. Thái độ ấy không dừng lại ở câu chuyện của một công ty, DeepSeek. Nó nổi với chất liệu của cả một ngành, nơi các công ty đi sau rất nhanh như Moonshot, MiniMax và Zhipu nhanh chóng bắt kịp bằng cách đi theo dòng chảy của các mô hình mở.

Nhưng tốc độ kiểu Trung Quốc cũng mang theo những điều Hàn Quốc nên gạt bỏ. Điều đầu tiên là một dạng quản trị nói về phát triển và an ninh cùng lúc, trong khi cán cân nghiêng nặng về phía an ninh. Các nghiên cứu về quản trị AI ghi nhận rằng các nhà hoạch định chính sách Trung Quốc nhìn AI như một công nghệ có khả năng lật đổ, có thể làm rung chuyển hệ thống. Theo cách nhìn đó, các dịch vụ AI tạo sinh phải đăng ký trước khi phát hành ra công chúng, còn cả nội dung lẫn mô hình đều trở thành đối tượng quản lý. Đó là một cấu trúc nói về phát triển, nhưng đặt an toàn chính trị lên trước. Nghiên cứu mô tả giới lãnh đạo Trung Quốc có một cách nhìn toàn diện, chia an ninh thành nhiều lớp, trong đó an toàn chính trị được đặt đầu tiên.

Vì vậy, dù mô hình của một công ty có năng lực đến đâu, nó vẫn bị xếp vào nhóm rủi ro nếu vượt qua ranh giới kiểm soát chính trị. Khung này luôn kéo theo phê phán rằng về thực chất, nó là kiểm duyệt và đàn áp. Điều Hàn Quốc nên học không phải là kiểm soát nhân danh an ninh, mà là sự thẳng thắn khi nói về phát triển và rủi ro trong cùng một văn bản. Cách vượt qua ranh giới đó rồi đóng lại sự cởi mở và quyền biểu đạt không phải là con đường dành cho Hàn Quốc, một nước thuộc phe tự do.

Cần nói thêm một điểm nữa. Cơ chế quản trị của Trung Quốc vận hành bằng cách đặt các ngưỡng đối với doanh nghiệp. Những mô hình được phát hành dưới dạng dịch vụ phải trải qua đăng ký, còn nội dung thì chịu sự rà soát. Các ngưỡng này vừa là tay nắm để chính phủ điều khiển ngành, vừa là chi phí mà doanh nghiệp phải gánh. Nếu trong những văn bản nói đồng thời về phát triển và an ninh, chúng ta không nhìn xem ai chịu chi phí tuân thủ, và mức kiểm

soát ấy thu hẹp biểu đạt và thông tin đến đâu, chúng ta sẽ chỉ thấy các con số tăng trưởng mà bỏ lỡ gánh nặng nằm bên dưới chúng.

Khi Hàn Quốc thiết kế quy định, điều then chốt là giữ thái độ viết về phát triển và rủi ro cùng lúc, đồng thời không để cán cân nghiêng về kiểm soát. Nhà hoạch định chính sách phải tính đến cả hai mặt của đồng xu và thẩm định: chúng có thể là thiết bị an toàn, nhưng cũng có thể trở thành rào cản gia nhập thị trường.

Điều thứ hai cần loại bỏ là thói quen chuyển chi phí xã hội sang cho doanh nghiệp và cá nhân. Nỗi lo AI sẽ làm xáo trộn việc làm cũng rất mạnh ở Trung Quốc. Các chương trước đã xem như bằng chứng phản chứng khuynh hướng nhà nước đáp lại nỗi lo ấy không phải bằng tái phân phối và lưới an sinh, mà bằng sự tự sinh tồn của cá nhân, trong khi đáp lại doanh nghiệp bằng cách đặt lên họ nghĩa vụ bảo vệ việc làm. Lao động tăng đáng kể trong việc làm sạch dữ liệu, sự kiểm soát công việc gián tiếp, và chi phí tuân thủ quy định mà doanh nghiệp gánh chịu đều bị che khuất sau những kết quả lấp lánh.

Khó có thể xem những bóng tối này là các con số đã được xác lập, và nhiều điều vẫn chưa được xác nhận. Điều rõ ràng là Hàn Quốc không được lặp lại thói quen chạy theo tốc độ trong khi đẩy các chi phí ấy sang người lao động và các doanh nghiệp nhỏ và vừa. Một con đường bền hơn là nhà nước cùng chia sẻ gánh nặng thông qua các quỹ chuyển đổi và kinh phí đào tạo lại. Nếu các con số tăng trưởng và đời sống của những người chống đỡ các con số ấy không được tính trong cùng một tài liệu, các chỉ báo và nơi làm việc sẽ dần tách xa nhau.

Điều thứ ba cần loại bỏ là sự thổi phồng công nghệ che giấu các giới hạn vật lý. Các tuyên bố về hiệu năng của Trung Quốc xoay quanh tự chủ bán dẫn phải được đọc cẩn trọng. Cuốn sách này không nói rằng phần cứng được thiết kế để lách kiểm soát xuất khẩu có hiệu năng ngang với các chip hàng đầu của Mỹ. Những con số đặt mật độ lên trước nhưng che đi chi phí nhiệt và điện năng tiêu thụ rất dễ biến thành màn trình diễn chính trị. Chúng ta cũng nên tránh các khẳng định chắc chắn rằng các mô hình Trung Quốc đã gần bắt kịp hoặc vượt qua các mô hình Mỹ trong bảng xếp hạng benchmark.

Lý do là các chỉ số dùng để đo mức độ lan tỏa vẫn chưa ổn định thành một chuẩn chung. Sức mạnh của Hàn Quốc trong ngành bán dẫn đến từ tỷ lệ thành phẩm và kỹ thuật quy trình minh bạch. Cách bảo vệ thế mạnh của Hàn Quốc là đặt hiệu suất có thể kiểm chứng và thực tế tài chính lên trước, thay vì bị cuốn vào cuộc cạnh tranh phóng đại quanh các chỉ số. Vì vậy, các báo cáo ngành của Hàn Quốc vẫn đọc năm 2026 của Trung Quốc như một giai đoạn công nghệ, đồng thời nhấn mạnh rằng phải nhìn vào giới hạn của benchmark và bức tranh tài chính thực tế.

Bây giờ hãy vượt qua cách cân nhắc từng điểm một, điều gì cần học và điều gì cần bỏ, để chuyển sang một cách đọc AI Trung Quốc rõ hơn. Nếu chỉ nhìn AI Trung Quốc qua bảng xếp hạng doanh nghiệp, chúng ta sẽ bỏ lỡ một nửa bức tranh. Chỉ đếm chi phí của DeepSeek, số

lượng mô hình phái sinh từ Qwen, hay điểm benchmark của một công ty không cho thấy hệ thống này vận hành ra sao. Bức tranh chỉ hiện ra khi ba trục được đặt chồng lên nhau: cơ quan nhà nước, doanh nghiệp, và hạ tầng dữ liệu. Nếu AI Mỹ lớn lên trong tay khu vực tư nhân ở Silicon Valley và vốn đầu tư mạo hiểm, thì AI Trung Quốc chuyển động như một cơ chế lớn, nơi ba trục này ăn khớp với nhau.

Trục thứ nhất, các cơ quan nhà nước, làm nhiều hơn việc ban hành quy định. Nó đứng ở vị trí chỉ huy, gom dữ liệu và vốn đang phân tán về một mối. Trung Quốc đã lập National Data Administration như một cơ quan chuyên trách về dữ liệu. Theo văn kiện chiến lược AI+, vào tháng 8 năm 2025, National Data Administration đã triệu tập một cuộc họp tại Guiyang với các bộ ngành trung ương, gồm Ministry of Industry and Information Technology, một số chính quyền địa phương, và các công ty liên quan đến dữ liệu, để bàn về việc phát triển Digital China. Cuộc họp tập trung vào cách kết nối các mạng hành chính phân tán với dữ liệu doanh nghiệp.

Tại cuộc họp này, việc xây dựng một mạng tài nguyên tính toán nối các vùng phía đông và phía tây, được gọi là compute power network, hay suanliwang, đã được đưa vào chương trình nghị sự cho những năm tới. Điều đó có nghĩa là chính phủ đứng ở vị trí viết bản nhạc. Như các tài liệu xu hướng chính sách gần đây giải thích, chính sách AI của Trung Quốc không đi xuống dưới dạng quy định từ một bộ duy nhất, mà là sự phối hợp nhiều tầng chạy từ State Council đến chính quyền địa phương. Ở cuối sự phối hợp đó, các cơ quan như National Data Administration biến dữ liệu thành ngôn ngữ của chính sách.

Mệnh lệnh này hình thành qua quy hoạch từ trên xuống. Theo một báo cáo của Samjong KPMG, một số bộ ngành, trong đó có National Data Administration, đã cùng ban hành các kế hoạch hành động nhiều năm về dữ liệu và đưa vào đó các mục tiêu định lượng nhằm đưa ngành dữ liệu lên một quy mô đã định trong một thời hạn rõ ràng. Báo cáo cũng ghi nhận rằng một quỹ đầu tư công nghiệp mạch tích hợp quy mô lớn đã được khởi động lại để dùng vốn nhà nước hỗ trợ Trung Quốc chuyển sang sản xuất chất bán dẫn tại Trung Quốc. Trên bàn cờ này, các mô hình của những công ty tư nhân như Alibaba và Tencent được tiếp cận, dưới sự phê chuẩn của chính phủ, dữ liệu công nghiệp của các doanh nghiệp nhà nước.

Đó là cách một vòng phản hồi dữ liệu mà các công ty phương Tây khó có được được tạo ra. Đây là cấu trúc trong đó hai vòng mà U.S.-China Economic and Security Review Commission mô tả, vòng mã nguồn mở nơi các mô hình công khai lưu chuyển và vòng phản hồi nơi dữ liệu quay về từ các hiện trường công nghiệp, chồng lên nhau trong tay chính phủ. Sự mở ở đây không phải là mở vì bản thân sự mở. Đó là sự mở quay trở lại dưới dạng kiểm soát công nghiệp.

Trục thứ hai, hạ tầng dữ liệu, đem lại cho kế hoạch này một thân thể vật chất. East Data, West Computing cho thấy nơi quy hoạch của chính phủ và hạ tầng ăn khớp với nhau. Đó là hạ tầng quốc gia chuyển nhu cầu tính toán từ miền đông nhiều dữ liệu sang miền tây, nơi điện năng và đất đai dồi dào hơn. Nhưng giữa kế hoạch và thực tế có một khoảng cách đau đớn. Báo cáo của

U.S.-China Economic and Security Review Commission chỉ ra rằng tỷ lệ vận hành thực tế tại các trung tâm dữ liệu xây ở miền tây vẫn ở mức thấp. Các công ty ở miền đông vẫn chưa vượt qua được bức tường độ trễ truyền thông mà họ phải vượt qua để suy luận thời gian thực.

Câu trả lời nảy sinh từ vấn đề này là mạng lưới sức mạnh tính toán đã bàn ở phần trước. Đó là nỗ lực giải quyết nút thắt hạ tầng một lần nữa thông qua mối liên kết giữa chính phủ và doanh nghiệp. Chỉ khi nhìn cả tầm vóc lớn của kế hoạch lẫn tỷ lệ vận hành thấp tại hiện trường, chúng ta mới đọc được cùng lúc sức mạnh và điểm yếu trong hạ tầng của Trung Quốc. Điều này cũng phản chiếu trực tiếp sang Hàn Quốc. Vẽ hạ tầng trên quy mô lớn và làm cho hạ tầng ấy thật sự vận hành là hai bài toán khác nhau.

Trục thứ ba là doanh nghiệp. Họ cắm vào hạ tầng này và làm dịch chuyển sản xuất vật chất. Bức tranh sẽ bị méo nếu chỉ nhìn các công ty AI và công ty sản xuất của Trung Quốc như những doanh nghiệp phần mềm bán điểm truy cập đám mây. Họ kết nối trực tiếp với hạ tầng dữ liệu do chính phủ đặt nền và vận hành nhà máy. Thuật ngữ chính sách model-data resonance, nghĩa là mô hình lớn và dữ liệu kết hợp để tạo ra cộng hưởng công nghiệp, chính là tên gọi được đặt cho mối nối này. Ở đây, một cấu trúc hình thành, trong đó thiết bị và các nút dữ liệu tại hiện trường công nghiệp đi qua các trung tâm doanh nghiệp rồi nối vào mạng hạ tầng dữ liệu quốc gia.

Doanh nghiệp không đổi mới một mình. Họ kết hợp với hạ tầng và tự làm lại mình để đáp ứng các mục tiêu do chính phủ đặt ra: bộ dữ liệu công nghiệp và agent công nghiệp. Đây cũng là lý do các báo cáo ngành của Hàn Quốc đến năm 2026 vẫn đọc Trung Quốc như một sân khấu công nghệ. Cạnh tranh không chỉ được định hình bởi năng lực của từng công ty, mà còn bởi cách công ty ấy được gắn với hạ tầng quốc gia.

Khi ba trục này được đặt chồng lên nhau, ta thấy rõ sức mạnh và nguy hiểm của hệ thống AI Trung Quốc đến từ cùng một nơi. Khi National Data Administration tập hợp các bộ và viết bản nhạc, Eastern Data, Western Computing cùng mạng sức mạnh tính toán trở thành những nhạc cụ ở quy mô lục địa, tạo ra điện và năng lực tính toán, trong khi Alibaba và các doanh nghiệp sản xuất thuộc sở hữu nhà nước làm chuyển động máy móc trên sàn nhà máy và trình diễn tác phẩm. Cây đuă chỉ huy của bản giao hưởng này được nắm bởi một bản song tấu giữa phát triển và an ninh. Cấu trúc tạo ra sức mạnh và cấu trúc mời gọi kiểm soát không tách rời nhau. Đây là điều không được bỏ sót khi nhìn vào AI Trung Quốc. Nếu tách nó ra, ta sẽ đánh mất nó; chỉ khi nhìn chung lại, nó mới hiện hình. Chỉ đếm tham số của một công ty, chỉ đọc văn bản quản lý của chính phủ, hoặc chỉ đo quy mô trung tâm dữ liệu đều không nắm được toàn bộ. Chỉ khi đọc ba loại tài liệu cùng nhau, đường nét của bộ máy này mới lộ ra.

Vậy điều gì hiện ra khi góc nhìn này quay trở lại Hàn Quốc? Hàn Quốc là một quốc gia có mật độ đổi mới cao. Nước này giữ vị trí hàng đầu thế giới về bằng sáng chế liên quan đến AI và bộ nhớ bán dẫn, và tỷ lệ người dùng AI tạo sinh đã tăng nhanh. Nhưng có một nghịch lý: mật độ ấy

chưa đi tới sàn nhà máy. Các khảo sát về doanh nghiệp sản xuất Hàn Quốc cho thấy nhiều công ty vẫn chưa đưa AI vào, với lý do chi phí và thiếu nhân lực. Ngay cả các công ty lớn cũng thường trì hoãn việc áp dụng vì gánh nặng chi phí và nghi ngờ về hiệu quả.

Một mặt độ bằng sáng chế hàng đầu thế giới và một tỷ lệ áp dụng thấp trong thực địa đang cùng diễn ra trong cùng một quốc gia. Ở nơi Trung Quốc đã xâm chính phủ, hạ tầng và doanh nghiệp thành một tuyến rồi đưa tuyến đó xuống nhà máy, Hàn Quốc lại có một cấu trúc trong đó ba trục này đứng khá xa nhau. Hàn Quốc đi trước về tài sản tri thức như bằng sáng chế và bài báo khoa học, nhưng cây cầu đưa những tài sản ấy vào năng suất của nền kinh tế thực vẫn chưa được xây xong.

Thu hẹp khoảng cách này là trọng tâm trong lựa chọn của Hàn Quốc. Có thể đặt hướng đi thành ba đường. Một là xây dựng năng lực của các cơ quan nhà nước trong việc chuyển mục tiêu sang ngôn ngữ của công nghiệp. Điều này có nghĩa là năng lực viết bằng ngôn ngữ của hiện trường rằng quy trình nào trong ngành nào phải được nâng lên mức nào, để văn bản quản lý và chương trình hỗ trợ không vận hành tách rời nhau. Sức mạnh thể hiện trong các văn kiện chính sách gần đây của Trung Quốc không nằm ở ngân sách khổng lồ, mà nằm ở độ chính xác khi các vấn đề tại nơi sản xuất được chuyển thành ngôn ngữ chính sách. Năng lực chuyển ngữ này có thể học được, tách khỏi quy mô ngân sách. Việc khác là về hạ tầng dữ liệu ở quy mô lớn như một tổ hợp của không gian, điện năng và nhu cầu.

Địa điểm đặt trung tâm dữ liệu, lưới điện và nhu cầu công nghiệp cần được chồng lên cùng một bản đồ, đồng thời bài học về tỷ lệ vận hành thấp của Trung Quốc phải được đánh dấu từ trước. Các trung tâm dữ liệu miền tây Trung Quốc đã cho thấy rằng làm cho cơ sở vật chất thật sự vận hành còn khó hơn xây chúng thật lớn. Việc còn lại là hạ ngưỡng tiếp cận dữ liệu công và dữ liệu công nghiệp để doanh nghiệp có thể kết nối các mô hình mở với dữ liệu công nghiệp. Ở giai đoạn này, điều then chốt là giữ cân bằng giữa mở dữ liệu và bảo vệ thông tin cá nhân, an ninh, mà không để cán cân ấy nghiêng sang kiểu kiểm soát của Trung Quốc.

Cần thêm một lời thận trọng ở đây. Những tuyên bố về trí tuệ hóa quân sự của Trung Quốc hoặc vị thế dẫn đầu trong trí tuệ nhân tạo tổng quát không phải là những phạm vi mà văn bản này xem như sự thật đã được xác quyết. Những gì tài liệu hiện có xác nhận chỉ dừng ở các giới hạn đã được kiểm chứng và các rủi ro chưa được xác nhận; ngoài phạm vi đó, văn bản không đưa ra khẳng định chắc chắn. Tốt hơn là không viết, khi chưa có chứng cứ được kiểm chứng, những câu nói rằng một mốc quân sự nào đó sẽ đạt được vào một năm nào đó, rằng một hệ thống vũ khí nào đó sẽ hoàn tất, hoặc rằng một chủ thể nào đó sẽ giữ vị trí dẫn đầu trong trí tuệ nhân tạo tổng quát. Chính sách của Hàn Quốc cần đề phòng việc bị những hình dung phóng đại như vậy về tương lai kéo vào chỗ lãng phí ngân sách và sự chú ý của công chúng.

Có những diễn đàn đặt câu hỏi Trung Quốc nên được nhìn nhận đến mức nào như một mối đe dọa và đến mức nào như một cơ hội, trong đó có các hội nghị học thuật đánh giá lại giá trị

chiến lược của Trung Quốc. Tuy vậy, trong cuốn sách này, những cuộc thảo luận ấy vẫn là những vấn đề cần được kiểm chứng thêm. Đặt tài liệu và công bố đã được xác minh ở phía trước, còn các tuyên bố chưa được xác minh ở phía sau, tự thân nó đã là một thái độ mà Korea nên học. Kỷ luật phân biệt lập luận nào trở thành sự thật và lập luận nào vẫn còn là câu hỏi bỏ ngỏ chính là thứ duy trì niềm tin vào chính sách.

Hãy mở rộng kỷ luật đó thêm một bước. Các văn kiện chính sách của Trung Quốc và các công bố của doanh nghiệp phản chiếu lẫn nhau. Khi một văn kiện của chính phủ đặt ra các mục tiêu như bộ dữ liệu công nghiệp và tác nhân công nghiệp, các công bố của doanh nghiệp và báo cáo ngành cho thấy dấu vết của việc những mục tiêu ấy được chuyển thành hoạt động kinh doanh thật. Các phân tích bên ngoài, như báo cáo của Samjong KPMG và U.S.-China Economic and Security Review Commission, cho thấy chi phí và giới hạn mà văn kiện chính phủ không nhắc đến. Tỷ lệ vận hành thấp, các chỉ số chưa trở thành chuẩn mực, và cơ chế quản trị nghiêng về kiểm soát là những ví dụ.

Đây là lý do các nhà hoạch định chính sách của Korea phải đọc ba loại tài liệu này cùng nhau. Nếu chỉ đọc văn kiện chính phủ, bức tranh trông rất lớn. Nếu chỉ đọc các phân tích bên ngoài, rủi ro trông bị phóng đại. Chỉ khi hai lớp tài liệu được đọc chồng lên nhau, quy mô thật của tốc độ Trung Quốc mới hiện ra. Cách đọc nhiều lớp này là một cách để đánh giá Trung Quốc, và cũng là một cách để Korea soi lại chính sách của chính mình.

Cách đọc nhiều lớp ấy cũng phải được áp dụng cho bằng chứng trái chiều nếu muốn đánh giá công bằng. Chi phí thấp của DeepSeek và hệ sinh thái phát sinh rộng của Qwen được ghi nhận trong các báo cáo như những sự thật rõ ràng. Nhưng lời phê phán rằng những thành quả ấy được xây trên kiểm duyệt và đàn áp, gánh nặng tuân thủ của doanh nghiệp, và sự kiểm soát đối với lao động tầng đáy cũng có trọng lượng tương tự. Nếu các câu viết về thành tựu và các câu viết về bóng tối không nằm cạnh nhau trên cùng một trang, tốc độ Trung Quốc chỉ còn phần sáng, còn phần tối thì biến mất. Vì vậy, khi Korea chọn điều cần học, Korea phải viết bằng cùng một bàn tay không chỉ danh sách những gì sẽ tiếp nhận, mà cả danh sách những gì sẽ từ chối. Korea nên lấy tốc độ thực thi vươn tới nền kinh tế thật, đồng thời lọc bỏ cách đối xử với con người như vật tiêu hao và các cơ chế kiểm soát đóng kín biểu đạt.

Khi chương này khép lại, nó để lại cho độc giả vài câu hỏi còn mở hơn là một câu trả lời cố định. Câu hỏi đầu tiên liên quan đến vốn và hạ tầng. Chưa ai biết điều gì sẽ chạm giới hạn trước: cuộc chiến tiêu hao theo kiểu vốn của Hoa Kỳ, hay cuộc huy động hạ tầng nhà nước theo kiểu Trung Quốc. Khoản chi khổng lồ của Big Tech Hoa Kỳ có thể mệt mỏi trước, hoặc hạ tầng của Trung Quốc, bị đè nặng bởi nút thắt tỷ lệ vận hành thấp, có thể rung lắc trước. Giữa hai mô hình đó, Hàn Quốc, nơi không có vốn khổng lồ cũng không có lãnh thổ khổng lồ, sẽ tạo ra chỗ đứng riêng bằng loại vốn nào và loại hạ tầng nào? Đây là câu hỏi đầu tiên.

Câu hỏi thứ hai liên quan đến thế giới vật lý. Vẫn chưa được xác nhận khi nào AI ngày nay có thể xử lý ổn định các điều kiện không đều đặn của sàn nhà máy. Hàn Quốc có một trong những mật độ bằng sáng chế cao nhất thế giới, vậy vì sao nước này vẫn không thể tự mình vượt qua rào cản áp dụng ngay trên các sàn nhà máy của mình? Nếu sức mạnh về bằng sáng chế và bộ nhớ không thể chuyển thành tỷ lệ áp dụng tại hiện trường, mật độ ấy chỉ còn là một con số trên chỉ báo. Cách Trung Quốc xâu chính phủ, hạ tầng và doanh nghiệp lại với nhau rồi đưa xuống hiện trường đặt ra cho Hàn Quốc một câu hỏi như tấm gương, bởi ba trục ấy vẫn còn tách rời ở Hàn Quốc.

Câu hỏi thứ ba liên quan đến con người. Phía sau những kết quả chói sáng là lao động ở tầng nền để làm sạch dữ liệu và nỗi sợ mất việc. Nếu chính phủ, doanh nghiệp và xã hội không muốn đẩy chi phí này sang người lao động và các doanh nghiệp nhỏ và vừa, mỗi bên sẽ gánh phần nào, và gánh bằng cách nào? Cách Trung Quốc chuyển chi phí này sang cá nhân đang tự xoay xở và sang doanh nghiệp như một gánh nặng quản lý không phải là con đường Hàn Quốc nên lặp lại. Quản trị của Hàn Quốc nên tách khỏi Trung Quốc ở đúng điểm nó đặt cả các con số tăng trưởng và cuộc sống của những người chống đỡ các con số ấy vào cùng một tài liệu.

Câu hỏi thứ tư liên quan đến quan hệ với thế giới bên ngoài. Khi các mô hình tiên tiến được đối xử như vũ khí và quyền tiếp cận đang bị khóa lại, Hàn Quốc sẽ đứng ở đâu giữa các khối dữ liệu, và đứng bằng cách nào? Nếu Hàn Quốc không thể ngồi tại chiếc bàn nơi các quy tắc được viết ra, nước này sẽ vẫn ở vị trí đi theo những đường kẻ do người khác vạch. Nếu Hàn Quốc đã đọc được cấu trúc trong đó chiến lược mở của Trung Quốc quay trở lại thành ưu thế công nghiệp, thì Hàn Quốc cũng phải xây dựng cách riêng để tính toán sự mở cửa và lợi ích cùng nhau.

Những câu hỏi này không phải là lời tu từ. Chúng là những hóa đơn đã đến hạn phải trả. Đo xem các cơ quan nhà nước, doanh nghiệp và hạ tầng dữ liệu đang đứng cách nhau bao xa, rồi đặt lại khoảng cách ấy, không phải là việc của riêng một bộ hay một công ty. Nơi đầu tiên phải tìm câu trả lời là có được tầm nhìn để thấy ba trục ấy cùng một lúc. Các văn kiện của Trung Quốc đã vẽ tầm nhìn đó thành một bức tranh, trong khi các văn kiện của Korea vẫn còn rải rác trên ba trang. Chương này bắt đầu bằng ba văn kiện xếp chồng trên một chiếc bàn, nên câu hỏi cuối cùng cũng xuất phát từ đó. Liệu chúng ta có thể đặt bản đồ các văn kiện chính phủ, công bố của doanh nghiệp và hạ tầng dữ liệu của mình lên cùng một trang, rồi đọc chúng như một tổng thể không? Vẫn chưa rõ bàn tay nào, trên chiếc bàn nào, trong bộ nào, sẽ bắt đầu cách đọc xếp lớp ấy.

Để đặt kỷ luật cho cách đọc xếp lớp, trước hết chúng ta phải giữ vững chuẩn mực để quyết định câu nào nên được xem là sự thật, câu nào nên để lại như một câu hỏi mở. Quy mô vốn đầu tư mạo hiểm được nêu trong báo cáo của Samjong KPMG và việc khởi động lại quỹ đầu tư ngành mạch tích hợp đều đứng trên nền tảng có công bố và văn kiện chống đỡ. Ngược lại,

những nhận xét về tỷ lệ vận hành thấp của các trung tâm dữ liệu miền tây, hoặc việc các chỉ số khuếch tán chưa thể ổn định thành chuẩn mực, lại trở thành thước đo để kiểm tra lại các con số hiệu quả. Nếu các nhà hoạch định chính sách của Korea tự rèn cho mình cách đọc hai loại câu này cạnh nhau, họ có thể cùng một tay tách ra cả vẻ lớn lao lẫn điểm yếu trong tuyên bố của bất kỳ quốc gia nào. Việc rèn luyện này không dừng ở chuyện đánh giá Trung Quốc. Nó quay trở lại nhiệm vụ kiểm tra các kế hoạch do chính các bộ của chúng ta ban hành.

Trở lại điểm đó, câu hỏi đầu tiên cần đặt ra là liệu các văn kiện chính sách của chúng ta có đang dịch các vấn đề tại hiện trường công nghiệp sang ngôn ngữ của chính những ngành ấy hay không. Các văn kiện về xu hướng chính sách gần đây của Trung Quốc cho thấy một hình thức phối hợp theo tầng, chạy từ State Council xuống đến chính quyền địa phương, và ở cuối chuỗi phối hợp đó, các cơ quan xử lý dữ liệu biến những nguồn lực rời rạc thành ngôn ngữ chính sách. Những nơi đặt phương hướng, những nơi đặt ngưỡng, và những nơi thu thập dữ liệu được nối với nhau ngay trong văn kiện. Cách sắp xếp ấy là nền móng của tốc độ.

Nếu các văn kiện của Korea không nói rõ, bằng ngôn ngữ của hiện trường, quy trình nào phải được nâng lên mức nào, và nếu các chương trình hỗ trợ cùng quy định pháp lý đi theo hai đường riêng, ba trục ấy sẽ vẫn ở xa nhau. Độ chính xác của việc dịch vấn đề có thể học được, tách khỏi quy mô nguồn vốn. Việc học ấy có diễn ra hay không sẽ quyết định liệu Korea có thể xây cây cầu đưa mật độ bằng sáng chế và bài báo khoa học thành tỷ lệ áp dụng trên sàn nhà máy hay không.

Cách đọc theo nhiều lớp ấy cũng phải được áp dụng cân xứng với danh sách những gì cần bỏ lại phía sau. Nếu chi phí thấp của DeepSeek và hệ sinh thái phái sinh rộng của Qwen là sản phẩm của một thái độ biết thiết kế đường vòng trong điều kiện bị ràng buộc, thì việc những thành tựu ấy xuất hiện ở ngưỡng đăng ký và thẩm định, dưới các cơ chế kiểm soát nội dung là đối tượng quản lý, và trên một quy mô đặt an toàn chính trị lên trước, cũng phải được đặt cùng trọng lượng. Korea nên giữ sự trung thực khi ghi lại tiến bộ và rủi ro trong cùng một văn bản, đồng thời cũng phải rèn được bàn tay biết lọc ra những khoảnh khắc cán cân nghiêng về kiểm soát. Bài học thật sự là đặt cả hai chuyển động ấy trong một phán đoán duy nhất. Nếu câu viết về thành tựu và câu viết về cái bóng của nó không đứng cạnh nhau trên cùng một trang, thì chỉ còn lại danh sách những điều cần học, còn danh sách những điều phải từ chối sẽ biến mất.

Support This AI Library Book

If this book was useful, you can support future translations, public editions, and open AI knowledge projects through PayPal.



PayPal

[**https://paypal.me/kimkjj**](https://paypal.me/kimkjj)

Scan the QR code or open the link above.