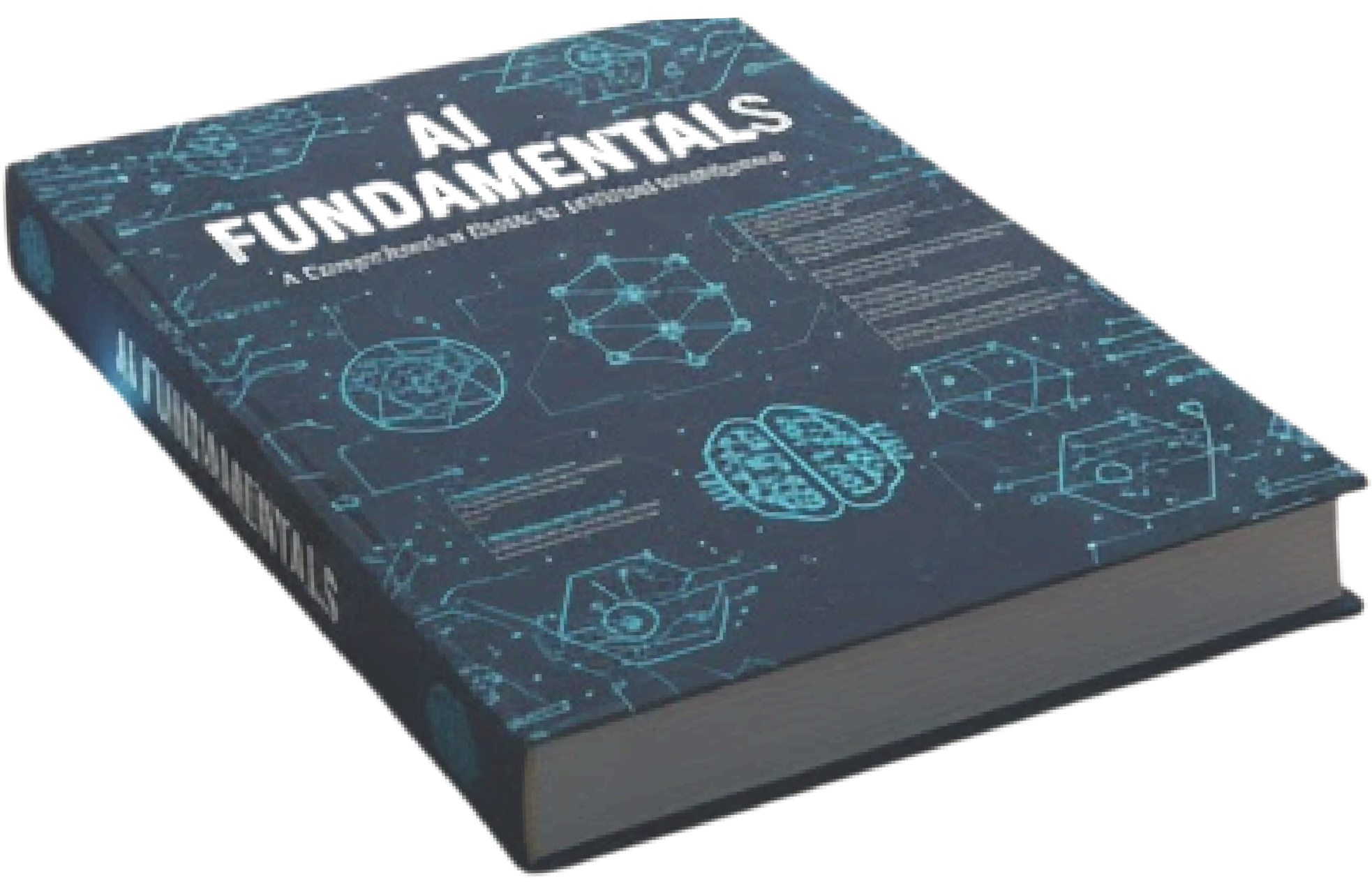
AI FUNDAMENTALS

NỀN TẢNG TRÍ TUỆ NHÂN TẠO

ThS. Bùi Thị Hải

MỤC LỤC:



LỜI GIỚI THIỆU

Tôi viết những dòng này không phải để kể về một "làn sóng" công nghệ, mà để gọi tên một thay đổi đang diễn ra rất khẽ - nhưng ảnh hưởng ngày càng sâu - trong đời sống của mỗi chúng ta. AI xuất hiện trong những điều tưởng như vụn vặt: một gợi ý tìm kiếm đúng lúc, một câu chữ được hoàn thiện khi ta còn gõ dở, một bản tin hiện lên vừa khớp mối quan tâm, một cuộc gọi bị chặn vì nghi ngờ rác. Ban đầu, nó chỉ giống như một tiện ích. Nhưng càng quan sát, tôi càng thấy AI không chỉ giúp ta nhanh hơn; nó có thể tác động đến cách ta nhìn thế giới và cách thế giới đáp lại ta - ta đọc gì, tin gì, chọn gì, học gì, và đôi khi cả cơ hội ta nhận được hay đánh mất. Vì vậy, câu hỏi quan trọng không còn là "AI có thông minh đến đâu", mà là "mình hiểu nó đến mức nào để vẫn giữ được quyền chủ động". Khi công nghệ đi nhanh, sự chủ động ấy không phải đặc quyền của người làm kỹ thuật; đó là một năng lực mới mà ai cũng cần để không vô tình để những quyết định được tạo ra thay mình.

Với tôi, AI chỉ thật sự có ý nghĩa khi nó làm cho con người tốt hơn trong công việc và tử tế hơn trong lựa chọn: giúp việc khó trở nên nhẹ nhàng, biến quy trình rối rắm thành minh bạch, và biến dữ liệu rời rạc thành căn cứ cho những quyết định công bằng, kịp thời. Trên hành trình cùng VN168 đưa AI đến gần hơn với cơ quan, doanh nghiệp và cộng đồng, tôi gặp rất nhiều người chung một tâm thế: tò mò nhưng dè dặt, muốn học nhưng ngại bắt đầu, muốn dùng nhưng sợ "mình không đủ nền". Và chính điều đó trở thành lý do của cuốn sách này - một điểm khởi đầu gần gũi, mạch lạc và thực tế, để bạn bước vào thế giới AI mà không bị ngợp, hiểu bản chất thay vì chỉ học mẹo, và quan trọng nhất: thấy rõ rằng AI là công cụ, còn bạn mới là người cầm lái.

Và để bắt đầu, chúng ta sẽ đi từ những viên gạch nền tảng nhất trong phần "NỀN TẢNG TRÍ TUỆ NHÂN TẠO": AI là gì và không phải là gì, vì sao nó "học" được, dữ liệu và thuật toán đóng vai trò ra sao, mô hình hoạt động như thế nào và vì sao AI có thể trả lời rất xuất sắc nhưng cũng có thể sai rất thuyết phục. Khi hiểu những nguyên lý cốt lõi ấy, bạn sẽ không còn nhìn AI như một khối mù mờ khó nắm bắt, mà như một hệ thống có logic - có giới hạn - và có cách để ta sử dụng một cách an toàn, hiệu quả, và chủ động.

Tôi viết cuốn sách này vì tin rằng muốn đi xa với AI, ta phải bắt đầu bằng nền tảng. Nền tảng không phải để thuộc lòng, mà để hiểu bản chất: AI đang giải quyết loại vấn đề gì, mạnh ở đâu, yếu ở đâu, vì sao có lúc nó "thông minh" đến đáng kinh ngạc nhưng cũng có lúc sai rất thuyết phục; dữ liệu, mô hình, tối ưu, và ngữ cảnh đã tạo nên "trí tuệ máy" như thế nào; và quan trọng nhất, con người cần đặt những câu hỏi nào để làm chủ công nghệ thay vì bị công nghệ dẫn dắt. Tôi luôn quan niệm: nếu con người cần "vaccine" để bảo vệ sức khỏe, thì khoa học công nghệ và chuyển đổi số chính là "vaccine" để phát triển bền vững. Nhưng "vaccine" chỉ phát huy tác dụng khi chúng ta hiểu, tin, và biết cách sử dụng đúng. AI cũng vậy.

Trong những trang tiếp theo, bạn sẽ bắt gặp các gợi ý liên quan đến lập trình và triển khai, bởi AI không thể tách rời thực hành. Tuy nhiên, tôi chủ ý không sa đà vào việc viết một chương trình "hoàn chỉnh", cũng không biến cuốn sách thành cuộc tranh luận dài về ngôn ngữ lập trình hay thư viện công cụ. Điều tôi muốn giữ lại cho bạn là phần cốt lõi hơn: "xương sống" của tư duy AI - hiểu bài toán trước khi chọn thuật toán, hiểu dữ liệu trước khi huấn luyện mô hình, hiểu vì sao tối ưu là nhịp tim của học máy, hiểu cách đánh giá để không bị ảo tưởng bởi con số, và hiểu những giới hạn cần được tôn trọng để AI không trở thành một chiếc hộp đen đầy tự tin nhưng thiếu đáng tin.

Tôi mong AI Fundamentals không chỉ giúp bạn "biết thêm một lĩnh vực", mà trao cho bạn một năng lực mới: năng lực đọc, hiểu và đối thoại với AI bằng tư duy rõ ràng; năng lực chọn đúng việc để dùng AI và biết khi nào không nên dùng; năng lực xây dựng niềm tin dựa trên hiểu biết, thay vì dựa trên cảm tính, hào nhoáng hay sự phóng đại.

Nếu phải gói gọn lý do tôi chắp bút trong một câu, thì đó là: tôi muốn góp một viên gạch nền để AI trở thành "tiện ích đời thường", nhưng được sử dụng bằng hiểu biết và trách nhiệm. Khi nền tảng đủ vững, chúng ta không chỉ "dùng AI", mà còn có thể kiến tạo bằng AI - cho doanh nghiệp, cho chính quyền, cho cộng đồng, và cho tương lai số của Việt Nam.

Chúc bạn một hành trình học tập hứng khởi, bền bỉ và nhiều khám phá cùng AI Fundamentals - Nền tảng Trí tuệ Nhân tạo.

ThS. Bùi Thị Hải
Chủ tịch HĐQT - Công ty Cổ phần Công nghệ VN168

1. TRÍ TUỆ LÀ GÌ?

Dưới góc nhìn thực dụng, "trí tuệ" thường được hiểu là năng lực đạt mục tiêu trong một thế giới phức tạp: biết quan sát, suy luận, học hỏi, dự đoán, rồi chọn hành động phù hợp khi thông tin không đầy đủ và nguồn lực có hạn (thời gian, năng lượng, dữ liệu, công cụ). Định nghĩa này nghe có vẻ rộng, nhưng chính sự "rộng" ấy lại phản ánh đúng bản chất: trí tuệ không phải là một nút bấm duy nhất, mà là một bó năng lực đan xen.

Trí tuệ không chỉ là "biết điều"

Một ngộ nhận phổ biến là đồng nhất trí tuệ với lượng kiến thức. Nhưng kiến thức giống như "kho đồ", còn trí tuệ giống như "cách dùng đồ": biết cái gì quan trọng, dùng lúc nào, bỏ qua lúc nào, và xoay sở ra sao khi không có đủ thứ mình muốn. Một người có thể thuộc rất nhiều công thức nhưng lúng túng trước một tình huống đời thực; ngược lại, có người ít chữ nhưng vẫn đưa ra quyết định sắc bén nhờ hiểu bối cảnh, hiểu người, hiểu rủi ro.

Vì vậy, nếu chỉ dùng những bài test thiên về ngôn ngữ, logic hình thức hay trí nhớ để kết luận "ai thông minh hơn", ta dễ rơi vào cái bẫy: đo cái dễ đo, rồi nhầm nó là cái quan trọng nhất.

Thước đo nào cũng có "điểm mù"

Các thước đi trí tuệ (từ bài kiểm tra năng lực đến benchmark AI) thường phải chọn một "đại diện" nào đó: tốc độ giải toán, độ chính xác trả lời, khả năng viết mạch lạc, năng lực lập kế hoạch...Nhưng mỗi đại diện đều có điểm mù:

- Hiệu suất trong phòng thí nghiệm không đảm bảo năng lực ngoài đời
- Một hệ thống có thể giỏi "làm bài", nhưng kém ở nhận biết giới hạn, tự kiểm tra sai, hoặc hành xử an toàn khi gặp tình huống lạ.
- Khi ta tối ưu mạnh một thước đo, hệ thống có xu hướng học cách thắng bài kiểm tra hơn là học bản chất vấn đề. Nói đơn giản: cứ có "đề thi" là sẽ có "luyện tủ".

Điều này đặc biệt quan trọng với AI: một mô hình có thể đạt điểm rất cao trên một bộ đánh giá cụ thể, nhưng lại thất bại ở những biến

thể nhỏ hoặc những tình huống "người dùng thật" hay gặp.

Cái bẫy lấy con người làm chuẩn

Lấy con người làm chuẩn có hai dạng "lệch" thường gặp.

Thứ nhất, lệch vi hình thức. Ta dễ đánh giá cao những gì giống con người (nói năng trôi chảy, hài hước, có vẻ "hiểu ý"), và đánh giá thấp những dạng thông minh không mang dáng dấp người. Nhưng tự nhiên đã cho thấy trí tuệ có nhiều hình: khả năng định hướng của chim di cư, chiến lược săn mồi của bạch tuộc, hay cách một đàn kiến tối ưu đường đi - đều là "thông minh" theo định nghĩa giải quyết vấn đề, dù không cần ngôn ngữ như ta.

Thứ hai: lệch vì giá trị. Con người không chỉ thông minh; con người còn có đạo đức, cảm xúc, ý nghĩa sống, mục tiêu dài hạn gắn với cộng đồng, Khi ra gộp tất cả thành một chữ "trí tuệ", ta vô tình biến trí tuệ thành một khái niệm vừa mô tả năng lực, vừa phán xét giá trị. Trong khi đó, một hệ thống có thể rất giỏi tối ưu mục tiêu nào đó nhưng lại hoàn toàn "mù" trước các giá trị mà con người coi trọng. Vì vậy, "thông minh hơn" không tự động đồng nghĩa "tốt hơn".

Trí tuệ hẹp, trí tuệ rộng, và câu hỏi về "tổng quát"

Một cách nhìn hữu ích là phân biệt:

- Trí tuệ hẹp: cực giỏi trong một phạm vi (chơi cờ, nhận dạng ảnh, tối ưu logistics).
- Trí tuệ rộng: biết chuyển kỹ năng sang tình huống mới, học nhanh khi đổi luật chơi, tự sửa sai, và phối hợp nhiều năng lực cùng lúc.

Nếu coi trí tuệ là "khả năng thích nghi", thì điều làm ta quan tâm nhất không phải AI có làm tốt một bài cụ thể hay không, mà là: khi bối cảnh đổi khác, AI còn giữ được năng lực không? Đây cũng là lý do những câu hỏi về tổng quát hóa, độ tin cậy, và kahr năng tự giám sát trở thành trung tâm trong thảo luận về AI tiên tiến.

Từ những điểm trên, chương sách thường muốn người đọc mang theo một kết luận thực tế: Trí tuệ là đa chiều, thước đo luôn có giới hạn, và "giống con người" không phải chuẩn duy nhất. Khi đã chấp nhận điều đó, ta sẽ dễ hiểu vì sao AI có thể phát triển theo nhiều hướng khác nhau.

2. AI CỔ ĐIỂN

Trong giai đoạn đầu của ngành Trí tuệ nhân tạo, nhiều nhà nghiên cứu tin rằng nếu ta muốn tạo ra "trí tuệ". ta nên bắt đầu từ thứ con người dùng để suy nghĩ: khái niệm, ngôn ngữ và lập luận. Từ niềm tin đó hình thành một truyền thống lớn của AI, thường được gọi là AI cổ điển hay AI ký hiệu. Ý tưởng cốt lõi rất trực quan: thay vì để máy "tự học" như cách trẻ em lớn lên, ta sẽ mô tả tri thức thành các ký hiệu rõ ràng và các luật suy luận, rồi để máy vận hành trên những ký hiệu ấy như một người giải bài toán logic.

Ở đây, "ký hiệu" có thể là một mệnh đề như "Trời mưa", "Bệnh nhân sốt", "Xe ở giao lộ", hoặc một cấu trúc biểu diễn quan hệ như "A là cha của B", "X là nguyên nhân của Y". Còn "luật" là những câu kiểu"nếu...thì...", cho phép hệ thống đi từ dữ kiện đến kết luận. Khi tri thức được viết thành quy tắc, ta có thể sửa luật nếu thấy sai, và ta có thể giải thích lại đường đi của lập luận cho người khác. Chính sự rõ ràng này khiến AI cổ điển từng được kỳ vọng như con đường thẳng nhất đê tiên tới trí tuệ.

Một động cơ quan trọng khác của AI cổ điển là ước mơ "chuẩn hóa thế giới" thành những bài toán có thể giải bằng suy luận. Nếu ta mô hình hóa được tình huống thành các trạng thái, rồi mô tả được hành động như những phép biến đổi hợp lệ, thì "ra quyết định" sẽ trở thành bài toán đi từ trạng thái ban đầu tới trạng thái mục tiêu. Từ đó, trung tâm kỹ thuật của AI cổ diển thường xoay quanh tìm kiếm trong không gian trạng thái: hệ thống thử các khả năng, loại các lựa chọn vô ích, và tiến dần tới lời giải.

Nghe thì đơn giản, nhưng ngay lập tức ta gặp một vấn đề kinh điển: số khả năng có thể bùng nổ nhanh đến mức "thứ hết" là bất khả thi. Chỉ cần thêm vài lực chọn ở mỗi bước, tổng số đường đi đã tăng theo cấp số nhân. Vì vậy, AI cổ điển sớm phát triển một lớp "mẹo" mang tính chiến lược, gọi là heuristic: các tiêu chí định hướng giúp hệ thống ưu tiên những nhánh hứa hẹn, giống như khi con người giải mê cung ta không đi bừa mà nhìn dấu hiệu, đoán hướng, và chọn đường "có vẻ đúng". Heuristic không đảm bảo luôn đúng tuyệt đối, nhưng nó

làm điều quan trọng hơn: giúp bài toán trở nên giải được trong thời gian hữu hạn.

Từ nền tảng đó, AI cổ điển nở rộ thành nhiều nhánh ứng dụng. Một nhánh nổi bật là lập kế hoạch: hệ thống không chỉ tìm một bước đi kế tiếp, mà xây cả chuỗi hành động là lập suy luận logic và chứng minh định lý, nơi máy được dùng để suy diễn từ tiên đề tới kết quả. Và một nhánh gây tiếng vang lớn trong công nghiệp là hệ chuyên gia: thay vì mô phỏng trí tuệ chung, người ta gom tri thức của chuyên gia trong một lĩnh vực hẹp (y khoa, chuẩn đoán thiết bị, cấu hình hệ thống...) thành khi luật. Khi nhận dữ liệu đầu vào, hệ thống áp luật để đưa ra kết luận hoặc gợi ý, đôi khi còn kèm theo phần giải thích "vì sao".

Điểm mạnh của thế hệ AI này nằm ở sự "có cấu trúc". Khi bài toán đủ sạch, đủ rõ, và miền tri thức tương đối ổn định, AI ký hiệu hoạt động rất đẹp. Nó giống như một chiếc máy suy luận đặt trong căn phòng gọn gàng: mọi thứ có nhãn, có vị trí, có quy tắc, và chỉ cần làm đúng quy trình là đi tới kết quả. Hơn nữa, vì lời giải có được tạo ra từ các luật và bước suy luận, người ta có thể kiểm tra, tranh luận, và thậm chí kiểm toán hệ thống theo cách gần với khoa học và kỹ thuật truyền thống. Với nhiều môi trường đòi hỏi trách nhiệm giải trình cao, ưu điểm "giải thích được" này không hề nhỏ.

Tuy nhiên, đời sống thật hiếm khi là một căn phòng gọn gàng. Nó là một chợ đông, nơi mọi thứ mơ hồ, nhiều ngoại lệ, nhiều tín hiệu nhiễu, và thông tin thường thiếu hụt. Cái khó của AI cổ điển nằm ở chỗ: để hệ thống suy luận, ta phải cung cấp cho nó một mô hình thế giới đủ đầy. Nhưng thế giới lại quá rộng và quá "lỏng" để viết thành luật một cách trọn vẹn. Chỉ riêng tri thức thường thức cũng đã là một đại dượng: cái cốc rơi thì vỡ, nước đổ thì ướt, người đang nói có thể đùa hoặc nói thật, một câu "được" có thể là đồng ý hoặc miễn cưỡng...Những điều con người xử lý tự nhiên lại rất khóc biến thành quy tắc rõ ràng mà không tạo ra vô số ngoại lệ.

Từ đây xuất hiện một nút thắt nổi tiếng: "nút thắt thu thập tri thức". Nếu muốn hệ thống thông minh hơn, ta phải viết thêm luật; viết

thêm luật thì phải gặp chuyên gia, phải chuẩn hóa ngôn ngữ, phải xử lý tình huống biên, và phải đảm bảo luật mới không mâu thuẫn với luật cũ. Càng nhiều luật, hệ thống càng nặng nề, vừa khó mở rộng vừa khó bảo trì. Đáng ngại hơn, trong những hệ lớn, một thay đổi nhỏ có thể tạo hiệu ứng dây chuyền khiến kết quả lệch ở những nơi không ai ngờ tới. Hệ thống trở nên "mong manh": nó chạy tốt trong điều kiện giống lúc thiết kế, nhưng chỉ cần gặp một tình huống hơi khác là lúng túng, hoặc suy luận theo cách rất kỳ quặc.

Ngoài ra còn một vấn đề sâu hơn: AI cổ điển thường giả định rằng thế giới có thể được mô tả bằng những thực thể rạch ròi và những quan hệ rõ ràng, trong khi thực tế đầy những vùng xám. Con người không chỉ suy luận bằng luật, mà còn dựa vào trực giác, kinh nghiệm, cảm giác về ngữ cảnh, và một loại "hiểu ngầm" khó gọi tên. Khi ta nói "trời lạnh", ta không đưa ra con số chính xác nhưng người nghe vẫn hiểu nên mặc thêm áo. Khi ta nói "để lát", ta không hẹn một mốc thời gian, nhưng cả hai bên vẫn phối hợp được. Những thứ như vậy khó "đóng gói" vào các luật cứng mà không làm mất đi tinh tế của đời sống.

Vì thế, AI cổ điển có thể được xem như một giai đoạn mà con người cố gắng "viết thế giới thành quy tắc" để máy có thể suy nghĩ. Nó đã tạo ra nhiều thành tựu quan trọng và đặt nền móng cho các khái niệm cốt lõi như biểu diễn tri thức, suy luận, lập kế hoạch và tìm kiếm. Nhưng chính khi đụng vào sự bừa bộn của thực tại, hướng tiếp cận này bộc lộ giới hạn: tri thức quá rộng để viết hết, ngoại lệ quá nhiều để liệt kê đủ, và bối cảnh quá tinh vi để đóng khung bằng vài dòng "nếu… thì…". Những giới hạn ấy không làm AI cổ điển "thất bại" theo nghĩa vô giá trị; chúng chỉ cho thấy một điều: để bước ra đời thực, AI cần thêm một nguồn sức mạnh khác ngoài luật lệ - khả năng học từ dữ liệu và tự hình thành biểu diễn, tức con đường sẽ dẫn ta sang các phương pháp hiện đại hơn ở những phần tiếp theo.

3. TRIẾT HỌC AI

Ở mốc này, câu chuyện về AI bắt đầu chuyển từ "làm được" sang "là gì". Một hệ thống có thể trả lời trôi chảy, lập kế hoạch, thậm chí tranh luận như người - nhưng liệu nó thật sự hiểu điều mình nói, hay chỉ đang tạo ra ảo giác hiểu biết bằng cách ghép nối những mẫu ngôn ngữ đúng? Đường ranh giữa "mô phỏng" và "hiểu" vì thế trở thành một trong những chủ đề triết học trung tâm của AI.

Có một lập trường thực dụng cho rằng nếu hành vi đã đủ thuyết phục và năng lực đã đủ mạnh, thì trong đời sống thực ta khó phân biệt được "hiểu" với "bắt chước". Người dùng quan tâm kết quả: AI có giúp chẩn đoán tốt hơn không, có viết mã hiệu quả không, có hỗ trợ học tập không. Khi mọi dấu hiệu bên ngoài đều cho thấy một hệ thống "biết điều nó đang làm", câu hỏi về "thật sự hiểu" dường như trở nên xa xỉ, hoặc ít nhất không còn là câu hỏi cấp bách.

Nhưng lập trường đối diện phản biện rằng thao tác đúng không đồng nghĩa với hiểu. Một hệ thống có thể xử lý ký hiệu, tuân theo quy tắc, tạo ra câu trả lời hợp lý mà vẫn hoàn toàn không có "ý nghĩa" ở bên trong. Nó giống như một quy trình dịch thuật hoàn hảo về hình thức: đầu vào là câu hỏi, đầu ra là câu trả lời; nhưng không có trải nghiệm chủ quan, không có "tôi đang nói về điều này", không có điểm tựa nội tại để gọi là hiểu. Từ góc nhìn này, AI có thể rất giỏi trong việc tạo ra biểu hiện của hiểu biết mà không cần bất kỳ thứ gì giống với hiểu biết theo nghĩa con người.

Từ tranh luận đó, khái niệm ý thức bước vào như một câu hỏi vừa hấp dẫn vừa khó nắm. Ý thức không chỉ là xử lý thông tin; nó thường được gắn với trải nghiệm chủ quan - cảm giác "có một ai đó" đang ở đây. AI có thể mô phỏng ngôn ngữ của cảm xúc ("tôi buồn", "tôi lo"), có thể phản hồi như một người đang đồng cảm, nhưng điều đó không tự động nói lên rằng nó thực sự có trải nghiệm. Nếu ta gọi những mô phỏng ấy bằng cùng một từ như cảm xúc con người, ta dễ nhầm lẫn giữa biểu hiện bề mặt và bản chất bên trong - một nhầm lẫn không chỉ mang tính học thuật, mà có thể kéo theo hệ quả xã hội.

Bởi khi AI đi vào đời sống, con người có xu hướng nhân hóa. Một

hệ thống nói năng lịch thiệp dễ được coi là đáng tin; một hệ thống phản hồi chắc chắn dễ được coi là có thẩm quyền; một hệ thống biết an ủi dễ được coi là "hiểu mình". Từ đó xuất hiện nguy cơ: người dùng gán cho AI vị thế giống chuyên gia, tin vào kết luận của nó như "chân lý", hoặc đặt niềm tin cảm xúc vào một thực thể vốn không có nghĩa vụ đạo đức như con người. Sự nhầm lẫn này còn có thể bị khuếch đại bởi thiết kế sản phẩm: giọng điệu tự tin, giao diện thân thiện, và khả năng đối thoại liên tục khiến AI giống một chủ thể hơn là một công cụ.

Vì vậy, triết học AI trong cuốn sách này không nhằm kéo người đọc vào sự mơ hồ, mà để cung cấp một bộ phân biệt rõ ràng khi bước ra thực tiễn. Một lớp là năng lực: hệ thống làm được gì, tốt đến đâu, có giới hạn nào. Lớp thứ hai là ý nghĩa: những lời nó nói có "hiểu" theo nghĩa nào, hay chỉ là mô hình dự đoán tạo ra câu chữ hợp lý. Và lớp thứ ba là trách nhiệm: khi hệ thống gây hại, ai chịu trách nhiệm - người dùng, tổ chức triển khai, nhà phát triển, hay nhà quản lý?

Càng tiến gần ứng dụng thực tế, ba lớp này càng phải tách bạch. Năng lực mạnh không tự động kéo theo hiểu biết; hiểu biết (nếu có) không tự động kéo theo đạo đức; và đạo đức không tự động xuất hiện chỉ vì hệ thống "nói như người". Nếu không phân biệt, ta dễ rơi vào hai cực đoan nguy hiểm: hoặc thần thánh hoá AI như một thực thể có quyền uy, hoặc xem nhẹ rủi ro vì nghĩ nó chỉ là phần mềm vô hại. Triết học, ở đây, đóng vai trò như một hàng rào nhận thức: giúp ta sử dụng AI đúng vị trí - khai thác sức mạnh của nó, nhưng không trao nhầm niềm tin, và không để trách nhiệm bị hòa tan vào "máy nói vậy".

4. AI HIỆN ĐẠI

Nếu AI cổ điển cố gắng "viết thế giới thành quy tắc", thì AI hiện đại chọn cách để thế giới tự "in dấu" lên mô hình thông qua dữ liệu. Cách nghĩ này dựa trên một quan sát đơn giản: phần lớn tri thức hữu ích của con người không nằm trong những câu định nghĩa gọn gàng, mà nằm trong vô số ví dụ - nhìn nhiều lần thì biết đó là mèo, nghe nhiều lần thì phân biệt được giọng nói, sống đủ lâu thì cảm nhận được ngữ cảnh trong một lời hứa hay một câu đùa. Mạng nơ-ron và học sâu trở thành trung tâm vì chúng có khả năng học ra các mẫu tinh vi từ trải nghiệm, thay vì đòi ta phải mô tả mọi thứ bằng lời. Về mặt cơ chế, mô hình được đặt vào một bài toán dự đoán: nó thử đoán câu tiếp theo, nhãn của bức ảnh, hướng di chuyển kế tiếp, hoặc hành động tối ưu trong một tình huống; rồi dựa vào sai số giữa dự đoán và thực tế mà điều chỉnh hàng triệu đến hàng tỷ tham số. Quá trình này lặp đi lặp lại như cách mài một lưỡi dao: mỗi lần mài rất nhỏ, nhưng đủ lâu thì sắc. Đáng chú ý là nhiều năng lực "trông như hiểu" lại xuất hiện từ những mục tiêu học tưởng như khiêm tôn, đặc biệt khi mô hình được huấn luyện trên dữ liệu lớn và đa dạng: nó học được không chỉ một kỹ năng đơn lẻ, mà một kho khuôn mẫu để suy luận gần đúng, liên tưởng và "điền vào chỗ trống" theo cách hữu dụng. Từ đây, ý tưởng về các mô hình nền tảng (foundation models) trở nên tự nhiên: thay vì xây một hệ thống cho một bài toán, người ta huấn luyện một mô hình tổng quát trên lượng dữ liệu khổng lồ, rồi tinh chỉnh để nó phục vụ các mục tiêu cụ thể - viết, dịch, tóm tắt, lập trình, hỏi đáp, hỗ trợ quyết định. Song song với học từ dữ liệu, AI hiện đại còn mở rộng theo hướng "tác nhân" (agents): không chỉ dự đoán, mà biết hành động để đạt mục tiêu. Khi có cơ chế thưởng–phạt, mô hình có thể học chiến lược thông qua thử - sai; khi được cung cấp công cụ, nó có thể tìm thông tin, lập kế hoạch, gọi API, phối hợp nhiều bước để hoàn thành một nhiệm vụ dài hơi. Từ góc nhìn "từ dưới lên", trí tuệ vì thế dần được hiểu như một hiện tượng nổi lên: không nhất thiết phải được cài sẵn thành các quy tắc, mà có thể xuất hiện khi hệ thống đủ lớn, đủ dữ liệu, đủ vòng lặp tối ưu hóa - giống như một khu phố đông đúc dần

hình thành nếp sống, dù không ai viết ra toàn bộ "luật" của nó ngay từ đầu.

Nhưng cũng chính vì sức mạnh ấy đến từ thống kê và tối ưu hóa, AI hiện đại mang một kiểu rủi ro rất khác với AI cổ điển: nó có thể đúng "phần lớn thời gian" mà vẫn sai ở những điểm chí mạng. Mô hình thường tỏa sáng trong vùng dữ liệu quen thuộc - những gì nó đã thấy, hoặc rất giống cái đã thấy - nhưng khi bối cảnh đổi khác, nó có thể trượt khỏi đường ray mà không tự nhận ra: một thay đổi nhỏ trong cách diễn đạt, một môi trường lạ, một nhóm người dùng khác, hoặc một đối thủ cố tình "gài" đầu vào là đủ để làm lộ ra các điểm yếu. Vấn đề vì thế không chỉ là mô hình có mạnh hay không, mà là nó có đáng tin hay không: nó có tổng quát hóa tốt không, có ổn định trước nhiễu không, có bị lệch vì dữ liệu thiên lệch không, có biết từ chối khi không chắc không, có thể giải thích hoặc ít nhất là truy vết lý do ở mức vận hành không. Và quan trọng hơn, AI hiện đại không tồn tại như một khối độc lập; nó là một hệ sinh thái gồm dữ liệu, cách gán nhãn hoặc tự giám sát, mục tiêu tối ưu hóa, quy trình huấn luyện, cách đánh giá, và cách triển khai ngoài đời. Chỉ cần một mắt xích sai - dữ liệu bị "bẩn", thước đo đánh giá không phản ánh thực tế, mục tiêu huấn luyện khuyến khích trả lời trơn tru hơn là trả lời đúng, hoặc triển khai thiếu giám sát - thì một mô hình trông rất tốt trong phòng thí nghiệm có thể trở thành nguồn lỗi ở quy mô xã hội. Khi AI được đưa vào sản phẩm, các vòng phản hồi còn phức tạp hơn: người dùng thay đổi hành vi theo hệ thống, hệ thống lại học từ dữ liệu mới, và thế giới thật bị "định hình" một phần bởi chính các dự đoán của AI. Vì vậy, câu chuyện của AI hiện đại trong một cuốn sách không thể chỉ là câu chuyện về thuật toán; nó phải là câu chuyện về việc xây dựng năng lực đi cùng kỷ luật an toàn: kiểm thử trong điều kiện gần thực tế, theodõi sau triển khai, thiết kế cơ chế giảm thiểu sai hỏng, và hiểu rằng "học được" không đồng nghĩa với "đáng giao quyền" - đặc biệt trong những lĩnh vực mà một sai lệch nhỏ có thể lan thành hậu quả lớn.

5. ROBOT

Robot đưa trí tuệ nhân tạo bước ra khỏi "thế giới của văn bản và dữ liệu" để đối mặt với một thực tại cứng rắn hơn nhiều: trọng lực, ma sát, vật cản, sai số đo lường, pin cạn, bánh xe trượt, cánh tay kẹt, và những tình huống không bao giờ lặp lại y hệt. Trong môi trường số, một mô hình có thể sai mà không gây hậu quả tức thì; còn trong môi trường vật lý, sai một chút có thể thành va chạm, rơi vỡ, chấn thương. Vì thế, trí tuệ của robot không còn được hiểu chủ yếu như "trả lời đúng", mà là "hành động đúng" dưới ràng buộc thời gian và rủi ro: biết dừng đúng lúc, biết né an toàn, biết chọn phương án đủ tốt thay vì tối ưu tuyệt đối, biết tự điều chỉnh khi cảm biến nhiễu hoặc khi thế giới lệch khỏi dự đoán. Cũng từ đây, người ta thấy rõ một khác biệt quan trọng: robot không chỉ cần nhận thức (perception) để hiểu mình đang ở đâu và đang thấy gì, mà còn cần điều khiển (control) để biến hiểu biết đó thành chuyển động ổn định, và cần lập kế hoạch (planning) để nối nhiều hành động nhỏ thành một mục tiêu dài hơi. Ba lớp này phải phối hợp nhịp nhàng; chỉ một lớp "giỏi" là chưa đủ, vì một robot nhìn rất tốt nhưng điều khiển kém vẫn có thể gây nguy hiểm, và một robot điều khiển mượt nhưng hiểu sai bối cảnh vẫn có thể đi lạc hoặc làm sai việc.

Phần robot cũng làm nổi bật giá trị của những kiến trúc dựa trên hành vi và phản xạ - một kiểu thông minh thực dụng, ưu tiên an toàn và tốc độ hơn là suy luận dài dòng. Thay vì chờ một bộ não trung tâm tính toán mọi thứ rồi mới hành động, robot thường được thiết kế như một tập các hành vi chồng lớp: lớp thấp phản ứng nhanh với nguy hiểm gần (tránh vật cản, giữ thăng bằng), lớp giữa lo các mục tiêu ngắn (đi đến điểm A, cầm nắm vật B), lớp cao xử lý kế hoạch dài và điều phối (làm xong việc này rồi chuyển sang việc khác). Cách tổ chức ấy giống như con người: nhiều phản xạ hoạt động gần như tự động để bảo vệ an toàn, còn suy nghĩ chậm rãi chỉ xuất hiện khi có thời gian và khi cần cân nhắc. Nhờ vậy, robot có thể "sống sót" trong thế giới thật - nơi không cho phép dừng lại quá lâu để suy luận. Đồng thời, robot là nơi các ý tưởng học bằng thưởng - phạt và tối ưu hóa thể

hiện rõ nét: thay vì lập trình từng động tác, người ta có thể để robot học chính sách hành động qua thử - sai trong mô phỏng, rồi chuyển sang thế giới thật với các cơ chế an toàn. Những hướng khác như mô phỏng tiến hóa hay hành vi bầy đàn cũng thường được dùng để cho thấy một bài học lớn: đôi khi không cần một bộ não siêu phức tạp; nhiều thành phần đơn giản phối hợp đúng cách có thể tạo ra hành vi "thông minh" ở cấp hệ thống, giống như đàn kiến hay bầy cá. Tuy vậy, robot cũng phơi bày mặt trái của việc "học từ kinh nghiệm": kinh nghiệm trong thế giới vật lý đắt đỏ và nguy hiểm. Vì thế, bài toán cốt lõi không chỉ là làm robot học nhanh, mà là làm robot học an toàn - học mà không phá hỏng đồ đạc, không gây hại cho người, không biến các sai số hiếm gặp thành tai nạn nghiêm trọng.

Khi robot đủ tốt để thay thế con người trong các tác vụ vật lý - sản xuất, vận chuyển, kho bãi, chăm sóc, thậm chí hỗ trợ y tế - tác động của AI vượt khỏi phạm vi thông tin và bước thẳng vào nền kinh tế lao động. Lúc này, câu hỏi xã hội không còn là "AI nói gì", mà là "AI làm gì" và "ai phải chịu trách nhiệm khi nó làm sai". Một robot giao hàng đi trên vỉa hè, một cánh tay máy trong nhà máy, hay một robot hỗ trợ người già đều đặt ra cùng một chuỗi vấn đề: tiêu chuẩn an toàn tối thiểu là gì, ranh giới giữa tự chủ và giám sát ở đâu, khi có tai nạn thì trách nhiệm thuộc về nhà sản xuất, đơn vị vận hành hay người dùng, và xã hội cần thích nghi ra sao khi một phần năng lực lao động được tự động hóa. Đồng thời, robot làm rõ hơn những câu hỏi đạo đức vốn dễ bị "mềm hóa" trong môi trường số: khoảng cách giữa một quyết định sai và một hậu quả thật gần hơn, nên các lựa chọn thiết kế - từ tốc độ, mức tự chủ, đến cơ chế dừng khẩn cấp - không chỉ là tối ưu kỹ thuật mà là lựa chọn mang tính đạo đức. Và khi robot được nối với con người qua các giao diện thần kinh - máy hoặc các hệ hỗ trợ tăng cường năng lực, đường ranh "đâu là con người, đâu là máy" bắt đầu trở nên mờ hơn theo cách rất thực tế: nếu một phần cảm nhận, một phần chuyển động, hay một phần quyết định được trung gian bởi hệ thống thông minh, thì quyền kiểm soát nằm ở đâu, sự phụ thuộc tăng tới mức nào, và làm sao để công nghệ phục vụ con người thay vì

âm thầm định hình lại con người. Ở điểm này, robot không còn chỉ là một ứng dụng của AI; nó trở thành phép thử rõ ràng nhất cho câu hỏi lớn của cả cuốn sách: khi trí tuệ đi kèm năng lực hành động trong thế giới thật, chúng ta phải xây dựng nó thế nào để vừa hữu ích, vừa an toàn, vừa không đánh đổi các giá trị xã hội một cách vô thức.

6. CẢM BIẾN VÀ VIỆC "NHÌN_NGHE_CHẠM" THẾ GIỚI

Trong bức tranh về AI, phần "trí tuệ" thường được hiểu như thuật toán suy luận hay mô hình học sâu, còn phần "cảm biến" bị xem như lớp phụ trợ kỹ thuật. Nhưng trong thực tế, cảm biến mới là cửa ngõ duy nhất để một hệ thống chạm vào thế giới: AI không trực tiếp biết sự vật, nó chỉ biết những tín hiệu được đo và được số hóa. Vì vậy, nếu dữ liệu đầu vào sai, méo, thiếu hoặc lệch ngữ cảnh, mọi tầng xử lý phía sau - dù tinh vi đến đâu - cũng chỉ đang suy luận trên một bức tranh đã bị biến dạng. Đây là lý do nhận thức (perception) không phải "một mô-đun", mà là nền móng. Một chiếc xe tự lái có thể có bộ lập kế hoạch tuyệt hảo, nhưng chỉ cần camera lóa nắng, lidar nhiễu, hoặc radar phản xạ sai bởi mặt đường ướt là đủ để biến quyết định hợp lý thành quyết định nguy hiểm. Nói cách khác, trí tuệ của hệ thống tự trị bị giới hạn bởi chất lượng của đôi mắt, đôi tai và làn da nhân tạo của nó - những thứ quyết định nó "thấy gì" trước khi nó "nghĩ gì".

Nhìn theo cách này, cảm biến không chỉ là phần cứng; nó là một chuỗi các lựa chọn: chọn loại cảm biến nào, đặt ở đâu, lấy mẫu nhanh chậm ra sao, lọc nhiễu thế nào, đồng bộ thời gian giữa các nguồn dữ liệu ra sao, và quan trọng nhất là hệ thống tin vào dữ liệu ấy đến mức nào. Thế giới thật luôn có nhiễu: ánh sáng đổi liên tục, vật thể bị che khuất, âm thanh bị dội, bề mặt phản chiếu gây ảo ảnh, rung động làm sai lệch gia tốc kế, bụi bẩn và mưa làm giảm chất lượng hình ảnh, và vô số "sự kiện hiếm" mà bộ dữ liệu huấn luyện thường không bao phủ đủ. Chính vì vậy, một hệ thống nhận thức tốt không phải là hệ thống "nhìn rõ nhất trong điều kiện lý tưởng", mà là hệ thống biết tự nghi ngờ khi điều kiện xấu: biết phát hiện lúc nào cảm biến đang bất thường, biết ước lượng độ tin cậy của từng nguồn, biết hạ mức tự chủ khi tín hiệu kém, và biết giữ an toàn trong trạng thái không chắc chắn. Năng lực này nghe có vẻ đơn giản, nhưng nó là ranh giới giữa một hệ thống "thông minh" trên giấy và một hệ thống "đáng tin" trong đời sống.

Thị giác máy là ví dụ điển hình cho việc nhận thức phức tạp hơn ta

tưởng. Nó không dừng ở chuyện "nhận ra đồ vật", mà bắt đầu từ những bước rất nền tảng: phát hiện cạnh, nhận biết vùng, phân đoạn đối tượng khỏi nền, theo dõi chuyển động, ước lượng độ sâu, rồi mới tiến tới nhận dạng và hiểu cảnh. Mỗi bước đều có thể sai theo cách tinh vi: một cái bóng có thể bị hiểu thành vật cản; một biển báo bị dán sticker có thể bị hiểu nhầm; một khuôn mặt bị ánh đèn hắt ngang có thể làm mô hình nhận dạng lệch đi. Những bài toán như stereo vision - dùng hai góc nhìn để suy ra độ sâu - cho thấy bản chất của khó khăn: cùng một vật thể, chỉ cần đổi góc nhìn, đổi ánh sáng, bị che một phần, hoặc có bề mặt trơn phản chiếu là việc "đối sánh" giữa hai ảnh đã trở nên mong manh. Và trong thế giới thật, các điều kiện này không phải ngoại lệ; chúng là trạng thái bình thường. Vì vậy, cảm biến tốt không chỉ là cảm biến "độ phân giải cao", mà là cả một hệ thống nhận thức biết kết hợp nhiều nguồn để bù cho điểm yếu của nhau: camera mạnh về ngữ nghĩa, lidar mạnh về hình học khoảng cách, radar chịu thời tiết tốt hơn, IMU cho biết chuyển động tức thời, còn xúc giác cho robot biết lực tiếp xúc thật sự khi cầm nắm.

Cảm biến chạm (xúc giác) và lực mô men thường là phần bị đánh giá thấp, nhưng lại quyết định robot có thể thao tác tinh tế đến đâu. Nhìn một vật không đồng nghĩa với hiểu cách cầm nó; cầm được cũng không đồng nghĩa với cầm an toàn. Muốn nhặt một chiếc ly mỏng, robot phải biết lực vừa đủ để không trượt mà cũng không bóp vỡ; muốn mở một cánh cửa, nó phải cảm nhận được lúc nào khóa kẹt, lúc nào cần đổi góc xoay. Những hiểu biết này không thể suy ra hoàn toàn từ thị giác, vì chúng nằm ở tương tác vật lý tức thời. Cảm biến lực và xúc giác biến thao tác thành "đối thoại" giữa robot và vật thể: robot thử, cảm nhận, điều chỉnh, rồi mới hoàn thành hành động. Khi thiếu lớp này, robot dễ trở thành một cỗ máy mạnh nhưng thô, phù hợp trong môi trường chuẩn hóa nhưng dễ gây lỗi trong đời thực.

Từ góc nhìn hệ thống, phần cảm biến còn gắn với một vấn đề sâu hơn: khoảng cách giữa dữ liệu huấn luyện và dữ liệu vận hành. Trong phòng thí nghiệm, dữ liệu thường sạch, góc nhìn ổn định, ánh sáng kiểm soát, vật thể ít bị che khuất. Nhưng ngoài đời, thế giới luôn

"lệch chuẩn": camera bẩn, micro bị gió, cảm biến nhiệt bị ảnh hưởng bởi nguồn nóng gần đó, và quan trọng nhất là bối cảnh thay đổi theo mùa, theo địa điểm, theo hành vi con người. Vì vậy, năng lực cảm biến quyết định giới hạn thực tế của robot và hệ thống tự trị không chỉ vì nó đo "đúng hay sai", mà vì nó định nghĩa cái mà hệ thống có thể biết một cách đáng tin. Khi cảm biến yếu, hệ thống buộc phải hành động trong mù mờ; khi cảm biến mạnh nhưng không được thiết kế để đánh giá độ tin cậy, hệ thống lại mắc một sai lầm nguy hiểm khác: tự tin sai. Chính ở điểm này, "nhận thức" trở thành một phần đạo đức của kỹ thuật: thiết kế một hệ thống biết khi nào nó không chắc, và biết giảm tham vọng để giữ an toàn, đôi khi quan trọng hơn thiết kế một hệ thống đạt điểm benchmark cao trong điều kiện lý tưởng.

KẾT LUẬN

Có một câu hỏi luôn xuất hiện sớm hay muộn khi ta đọc về AI: rốt cuộc "thông minh" nghĩa là gì, và vì sao máy đôi lúc làm được những việc tưởng như rất người, nhưng cũng có thể vấp ở những chỗ ngớ ngẩn đến khó tin? Câu hỏi ấy tưởng đơn giản, nhưng càng đào sâu càng mở ra một bwusc tranh rộng hơn nhiều: trí tuệ không phải một chiếc công tắc bật - tắt, cũng không phải một thức đo duy nhất. Nó là khả năng học và thích nghi để đạt mục tiêu trong một thế giới vừa thiếu chắc chắn vừa giới hạn nguồn lực. Khi hiểu vậy, ta bắt đầu nhìn thấy lý do AI có thể tồn tại dưới nhiều hình dạng, và vì sao việc đánh giá AI chỉ bằng chuẩn "giống con người" thường dẫn ta lạc hướng. Một con chim không cần ngôn ngữ vẫn bay qua bão gió; một hệ thống không "hiểu" theo cách ta hiểu vẫn có thể dự đoán, tối ưu, và hành động hiệu quả. Từ đây, người đọc được dẫn đến một kết luận quan trọng: AI không chỉ có một con đường đúng, vì "trí tuệ" vốn đã đa dạng ngày từ gốc.

Rồi ta bước vào hai lối làm AI đã định hình cả ngành. Một bên là tư duy ký hiệu: viết thế giới thành luật, biến vấn đề thành tìm kiếm trong không gian trạng thái, dùng heuristic để đi nhanh hơn thay vì mò mẫm, xây hệ chuyên gia để suy luận trong miền hẹp. Lối này đem lại cảm giác chắc chắn vì ta lần theo được "tại sao" của kết luận. Nhưng đời sống thật hiếm khi sạch sẽ như một bộ quy tắc; chỉ cần thêm ngoại lệ, mơ hồ, hoặc một tình huống "na ná" khác đi ở chi tiết nhỏ, kho luật phình lên, hệ thống nặng nề và mong manh. Lối còn lại - AI hiện đại - đổi hẳn câu hỏi: không cố viết tri thức thành câu chữ, mà để mô hình tự rút ra quy luật từ dữ liệu. Mạng nơ-ron học bằng cách điều chỉnh vô số tham số để giảm sai số; càng nhiều dữ liệu và năng lực tính toán, mô hình càng nắm được những mẫu hình phức tạp của ngôn ngữ, hình ảnh, âm thanh, chuyển động. Nhưng chính vì "học từ những gì đã thấy", AI hiện đại luôn mang theo nỗi lo cổ điển của thống kê: khi bối cảnh đổi khác, nó còn đáng tin không? Và vì nó là một hệ sinh thái - dữ liệu, cách huấn luyện, cách đánh giá, cách triển khai - chỉ cần một mắt xích lệch (dữ liệu thiên lệch, thước đo sai,

kiểm thử thiếu) là thứ tưởng tốt có thể trở thành nguồn lỗi.

Từ đây, câu chuyện chuyển sang nơi AI phải trả lời bằng hành động chứ không phải bằng văn bản: robot. Khi có thân thể, trí tuệ không còn là "đúng trên giấy", mà là "đúng kịp lúc" trước ma sát, nhiễu, sai số cảm biến và rủi ro thật. Robot nhắc ta rằng nhiều khi phản xạ nhanh, điều khiển theo hành vi và học bằng thử - sai lại hiệu quả hơn suy luận dài dòng; rằng những hệ đơn giản phối hợp có thể tạo ra hành vi phức tạp; và rằng AI, khi bước ra thế giới vật lý, kéo theo các câu hỏi xã hội không thể né tránh: an toàn, trách nhiệm, thay thế lao động, và cả ranh giới người - máy khi các giao diện ngày càng "dính" vào đời sống. Nhưng dù là robot hay phần mềm, mọi trí tuệ đều bắt đầu từ cửa ngõ bị đánh giá thấp: cảm biến và nhận thức. Máy "nhìn - nghe - chạm" bằng dữ liệu thô, và dữ liệu ấy luôn thiếu hoàn hảo. Nếu cửa ngõ sai, mọi tầng suy luận phía trên dù tinh vi đến đâu cũng xây trên nền nghiêng. Bởi vậy, năng lực cảm biến không chỉ là phần kỹ thuật phụ trợ; nó đặt ra giới hạn thực tế cho những gì hệ thống tự trị có thể làm an toàn trong đời thật.

Và đến đây, một kết luận lớn dần hiện rõ - không cần tuyên bố ồn ào, nhưng đủ mạnh để định hình cách ta đọc phần còn lại của cuốn sách: AI không phải "một công nghệ", mà là một chuỗi quyết định. Quyết định về cách ta định nghĩa trí tuệ và đo lường nó. Quyết định chọn lối viết luật hay lối học từ dữ liệu, và chấp nhận cái giá đi kèm. Quyết định đưa AI ra thế giới vật lý hay giữ nó trong môi trường kiểm soát. Quyết định tin vào dữ liệu đến mức nào khi cảm biến luôn nhiễu. Mỗi quyết định mở ra năng lực mới, đồng thời mở ra rủi ro mới - rủi ro không chỉ nằm trong thuật toán, mà nằm trong cách con người xây, triển khai, và dựa vào nó. Chính vì vậy, mục tiêu của cuốn sách không phải là khiến người đọc "sợ AI", mà là giúp người đọc nhìn AI như nó vốn là: một hệ thống khuếch đại. Nó khuếch đại năng lực, khuếch đại tốc độ, khuếch đại lợi ích - và nếu ta thiếu hiểu biết về nền tảng, nó cũng khuếch đại sai lệch, rủi ro và hậu quả.

ĐỌC THÊM

1. Computer Vision: Algorithms and Applications by R. Szeliski, published by Springer, 2010. This title discusses the current state-ofthe-art in computer vision, looking at the problems encountered and the types of algorithms employed. It is certainly a good, modern next step in the field.
2. Making Robots Smarter: Combining Sensing and Action through Robot Learning by K. Morik, M. Kaiser and V. Klingspor, published by Springer, 1999. This is good as a general follow-on from this chapter or the previous one. It considers learning in robots as a part of a sensor/action feedback loop.
3. Autonomous Mobile Robots: Sensing, Control, Decision Making and Applications by S.S. Ge, published by CRC Press, 2006. This is a comprehensive reference looking at the theoretical, technical and practical aspects of the field. It examines in detail the key components involved, including sensors and sensor fusion.
4. Robotic Tactile Sensing: Technologies and System by R. Dahiya and M. Valle, published by Springer, 2011. An in-depth look at robot touch, including such issues as shape, texture, hardness and material type. It offers comprehensive coverage, including rolling between fingers and sensing arrays.

Khi khép lại những trang cuối cùng của AI Fundamentals - Nền tảng Trí tuệ Nhân tạo, điều quan trọng nhất không nằm ở việc bạn đã ghi nhớ được bao nhiêu thuật ngữ hay mô hình, mà ở chỗ bạn đã sở hữu một "bản đồ tư duy" đủ rõ để bước vào thế giới AI với sự tự tin và tỉnh táo: hiểu AI đang làm gì, vì sao nó làm được điều đó, nó có thể sai theo cách nào, và ta phải đặt câu hỏi gì để dẫn dắt công nghệ phục vụ đúng mục tiêu của con người.

Cuốn sách này được viết như một nền móng - nền móng để bạn tiếp tục đi xa hơn, học sâu hơn, và ứng dụng mạnh mẽ hơn; từ phòng học đến phòng lab, từ doanh nghiệp đến cơ quan nhà nước, từ những bài toán nhỏ đến những hệ thống lớn tạo ra tác động thật. Nếu ở thời điểm bắt đầu, AI từng là một vùng mờ đầy tò mò và e ngại, thì sau hành trình này, hy vọng trong bạn đã hình thành một niềm hứng khởi có định hướng: hứng khởi để tìm hiểu sâu hơn về học máy, học sâu, dữ liệu, tối ưu, đạo đức và quản trị AI - và quan trọng nhất, hứng khởi để biến tri thức thành giải pháp, biến ý tưởng thành giá trị, biến công nghệ thành sức mạnh bền vững cho cộng đồng và tương lai số của Việt Nam.