

Is My Phone Reading My Mind? by Dr Matt Agnew
First published by Allen & Unwin, Australia, in the English Language, 2024
Copyright © Text, Matthew Agnew 2024
Copyright © Internal illustrations, Hana Kinoshita Thomson 2024

Xuất bản theo Hợp đồng bản quyền giữa Allen & Unwin và Nhà xuất bản Kim Đồng, 2024
Bản quyền tiếng Việt thuộc về Nhà xuất bản Kim Đồng, 2025

Biên mục trên xuất bản phẩm của Thư viện Quốc gia Việt Nam
Agnew, Matt
"Ây Ai" có gì sai? : Những điều cơ bản về đạo đức AI để sống cùng trí tuệ nhân tạo / Matt Agnew ; Thu Huyền dịch. - H. : Kim Đồng, 2025. - 104 tr. : minh họa ; 25 cm
ISBN 978-604-2-39368-3

1. Trí tuệ nhân tạo 2. Đạo đức
006.301 - dc23

KDF6511p-CIP

"ÂY AI" CÓ GÌ SAI?

NHÀ XUẤT BẢN KIM ĐỒNG
55 Quang Trung, phường Hai Bà Trưng, thành phố Hà Nội
ĐT: 024 3943 4730 - 024 3942 8632
Website: <http://www.nxbkimdong.com.vn> - Email: info@nxbkimdong.com.vn
CHI NHÁNH NXB KIM ĐỒNG TẠI MIỀN TRUNG
102 Ông Ích Khiêm, phường Thanh Khê, thành phố Đà Nẵng
ĐT: (0236) 3 812 335 - Email: cnkimdongmt@nxbkimdong.com.vn
CHI NHÁNH NXB KIM ĐỒNG TẠI TP. HỒ CHÍ MINH
248 Cống Quỳnh, phường Bến Thành, Thành phố Hồ Chí Minh
ĐT: (028) 3 925 1001 - (028) 3 925 0987 - Email: cnkimdong@nxbkimdong.com.vn

Chịu trách nhiệm xuất bản: Giám đốc BÙI TUẤN NGHĨA
Chịu trách nhiệm nội dung: Tổng Biên tập VŨ THỊ QUỲNH LIÊN
Biên tập: NGUYỄN TRUNG HIẾU
Trình bày: NGUYỄN QUỲNH KHUYẾN
Chế bản: TRẦN THỊ TUYẾT
Sửa bài: NGUYỄN ÁNH LY

In và gia công 2.000 bản - Khổ 21 cm x 25 cm - Tại Công ty CP In và Văn Hóa Phẩm Hà Nội
Địa chỉ: Thôn An Hạ, xã An Khánh, thành phố Hà Nội
Số xác nhận đăng kí xuất bản: 3880-2025/CXBIPH/18-322/KĐ cấp ngày 01/10/2025
Quyết định xuất bản số: 3868/QĐKĐ kí ngày 24/11/2025
In xong và nộp lưu chiểu quý 4/2025
ISBN: 978-604-2-39368-3

MỤC LỤC

Giới thiệu 4

PHẦN I: “A BỜ CỜ” VỀ “ÂY AI”

1. Trí tuệ nhân tạo là gì? 10

2. AI ở đâu trong cuộc sống thường ngày? 20

3. Đạo đức là gì? 37

PHẦN II: ĐẠO ĐỨC CỦA AI

4. Thiên lệch không chủ đích là gì? 52

5. AI có thể bị sử dụng sai lệch như thế nào? 61

6. AI có thể ảnh hưởng tới con người như thế nào? 77

7. Vấn đề kiểm soát AI & đạo đức rô-bốt 89

Kết luận 99

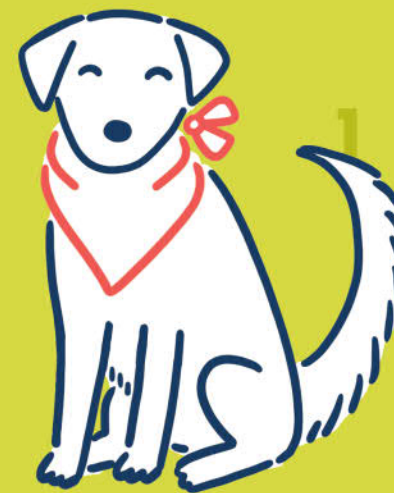
Chú dẫn 103

GIỚI THIỆU

Chào bạn! Tôi là **Tiến sĩ Matt**, và tôi cực kì phấn khích khi thấy bạn đọc cuốn sách này! Bạn chọn cuốn sách này chắc vì bạn cũng giống tôi – có một tình yêu vô bờ bến với việc nghiên cứu, tìm hiểu về rô-bốt – phải không? Hay chỉ đơn giản là bạn thích mọi thứ liên quan tới khoa học, và tình cờ biết tới cuốn sách này? Hay là bìa sách đẹp quá khiến bạn bị cuốn hút? Dù vì bất kì lí do gì thì **tôi cũng rất vui!**

Tôi yêu rô-bốt! Luôn luôn yêu thích rô-bốt. Tôi nghĩ, khi còn nhỏ, chúng ta ai cũng có giai đoạn thích rô-bốt, và đôi khi (nếu may mắn), giai đoạn đó còn kéo dài tới tận khi chúng ta bước vào tuổi trưởng thành. Tôi thấy mình thật may mắn khi được học tập và làm việc trong thế giới của **trí tuệ nhân tạo** (thứ này không hẳn là rô-bốt, mà là “bộ não” có thể điều khiển rô-bốt). Nhờ được tham gia vào thế giới kì thú đó, tôi đã học được rất nhiều điều, không chỉ về cách nó vận hành, mà quan trọng hơn là cả những tác động của nó tới chúng ta, vô cùng thú vị và khó lường trước. Có lẽ tôi dẫn lời giới thiệu đi hơi xa rồi.

Chào mừng bạn đến với cuộc hành trình tìm hiểu về trí tuệ nhân tạo - AI và đạo đức của AI! Câu này có một số từ cần được định nghĩa, nhưng đừng lo – chúng ta sẽ làm việc đó sau. Chúng ta sẽ tìm hiểu cùng nhau dần dần, cho tới khi nào bạn có thể tự nhận mình là **chuyên gia đạo đức trí tuệ nhân tạo** (hãy thử đọc nhanh cụm từ đó ba lần xem sao!)



Trí tuệ nhân tạo có ở quanh chúng ta. Càng ngày nó càng đóng vai trò **lớn hơn và quan trọng hơn** trong cuộc sống thường ngày. Thực tình, bạn chẳng cần biết tất cả những trí tuệ nhân tạo siêu việt quanh mình vận hành ra sao (phức tạp lắm!), nhưng nhất định phải biết rằng chúng có tồn tại thật, và chúng đang ảnh hưởng tới cuộc sống của bạn, theo những cách mà có thể bạn hoàn toàn không nhận biết được!

Với sự gia tăng của trí tuệ nhân tạo trong cuộc sống, chúng ta cần đặt ra một số câu hỏi. **Chúng ta có thể tin tưởng trí tuệ nhân tạo tới mức nào? Nó có “để mắt” tới chúng ta không? Làm sao chúng ta biết nó không điều khiển suy nghĩ của chúng ta theo một hướng nào đó?**

Lời đáp cho những câu hỏi này đặc biệt quan trọng, giúp chúng ta có thể sống hòa hợp với trí tuệ nhân tạo, đồng thời có thể tiếp tục làm chủ cuộc sống cũng như suy nghĩ của mình.

Trước khi tìm hiểu về đạo đức của trí tuệ nhân tạo, chúng ta sẽ bắt đầu với những điều cơ bản. **Điều quan trọng phải hiểu trước tiên!**



BONG BÓNG VUI!

Đây là vòng tròn **Bong Bóng Vui!** Xuyên suốt cuốn sách này, tôi sẽ kể những chuyện thú vị vào trong vòng tròn này. Trong mỗi vòng tròn có thể là một kiến thức lí thú, một câu chuyện hay ho, hoặc đơn giản chỉ là một chi tiết khiến chúng ta bất ngờ về chủ đề đang nhắc tới.

Vậy nên, hãy đọc ngay những thông tin trong các vòng tròn bong bóng, hoặc bỏ qua nếu bạn đang chăm chú dõi theo nội dung khác, rồi quay lại đọc bong bóng khi hết chương sách.

Hãy đọc theo cách của bạn!

HỘP THÔNG THÁI

Thỉnh thoảng bạn có thể bắt gặp một từ mà mình không biết. Đừng lo - đã có **HỘP THÔNG THÁI** ở đây để giúp bạn hiểu các từ khóa.

Nhân tiện, tôi xin nhắc nhẹ: Không ai có thể biết tất tần tật mọi thứ, thế nên, đừng bao giờ ngại hỏi. Nếu không rõ một điều gì đó, tôi luôn hỏi cho bằng được!

Chúng ta sẽ tìm hiểu xem trí tuệ nhân tạo là gì, và có thể tìm thấy nó ở đâu trong cuộc sống thường ngày. Sau đó, chúng ta sẽ đào sâu vào một vấn đề quan trọng: đạo đức của trí tuệ nhân tạo. Chúng ta sẽ cùng xem quan điểm của con người về thế giới có thể ảnh hưởng tới cách thức vận hành của trí tuệ nhân tạo như thế nào, trí tuệ nhân tạo có thể bị sử dụng sai lệch ra sao, ảnh hưởng của trí tuệ nhân tạo tới con người khi nó tiếp tục phát triển, và cuối cùng là làm sao có thể kiểm soát tốt nhất trí tuệ nhân tạo trong tương lai.

Sau khi đọc xong cuốn sách này, tôi hi vọng các độc giả sẽ được trang bị nhiều kiến thức mới về trí tuệ nhân tạo và đạo đức, về những việc cần làm để **giữ an toàn** cho tất cả, cũng như về cách trí tuệ nhân tạo có thể **định hình** và **làm phong phú thêm** cuộc sống con người.

0 1 0 1 0 0 0 0
0 1 1 0 1 1 0 0
0 1 1 1 0 1 0 1
0 1 1 1 0 1 0 0
0 1 1 0 1 1 1 1



“A BỜ CỜ”
VỀ
“ÂY AI”

I

N

‘A

E

P

CHƯƠNG 1:

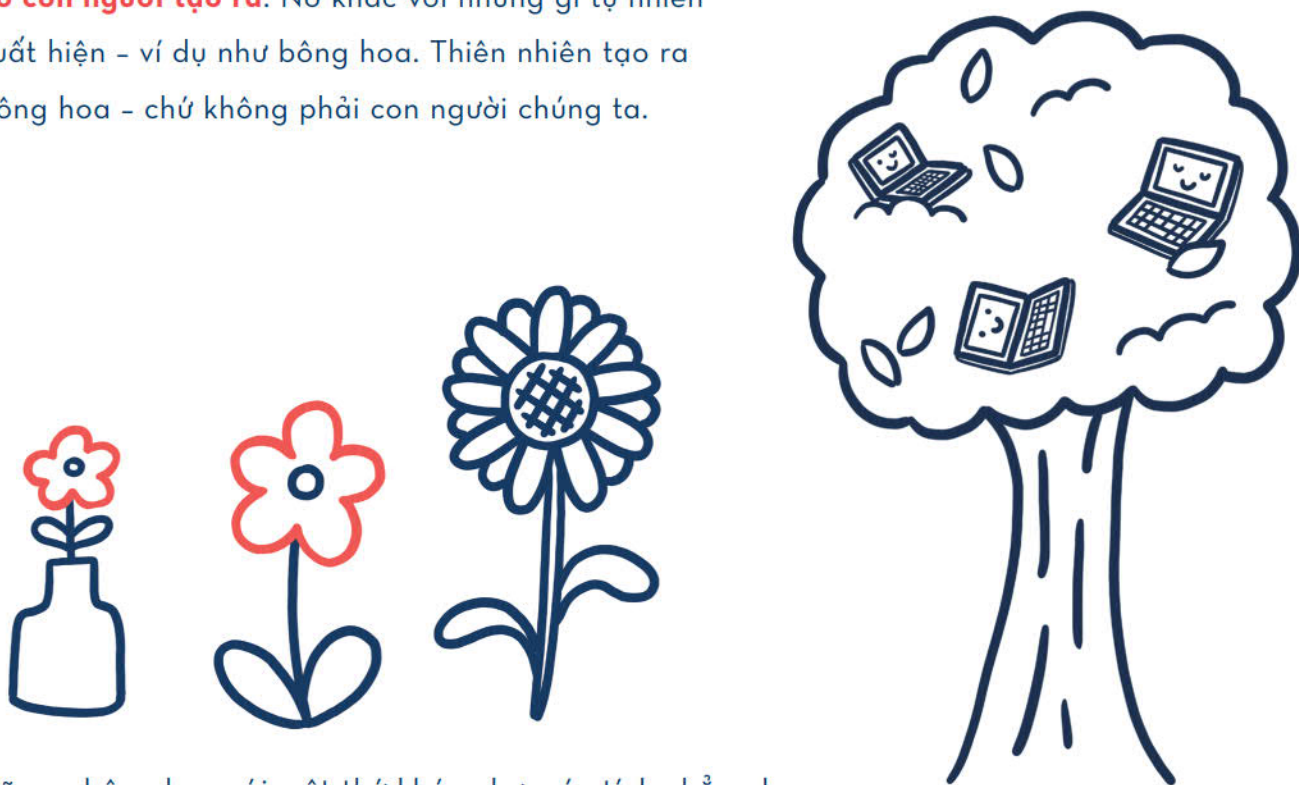
TRÍ TUỆ NHÂN TẠO LÀ GÌ?

Rõ ràng, điều đầu tiên là: Trí tuệ nhân tạo (hay AI) là gì?

Hãy tìm hiểu từng yếu tố trong cụm từ này, sau đó là việc kết hợp chúng lại với nhau.

NHÂN TẠO có nghĩa là gì?

Nhân tạo được định nghĩa đơn giản và dễ hiểu là: thứ đó **do con người tạo ra**. Nó khác với những gì tự nhiên xuất hiện – ví dụ như bông hoa. Thiên nhiên tạo ra bông hoa – chứ không phải con người chúng ta.

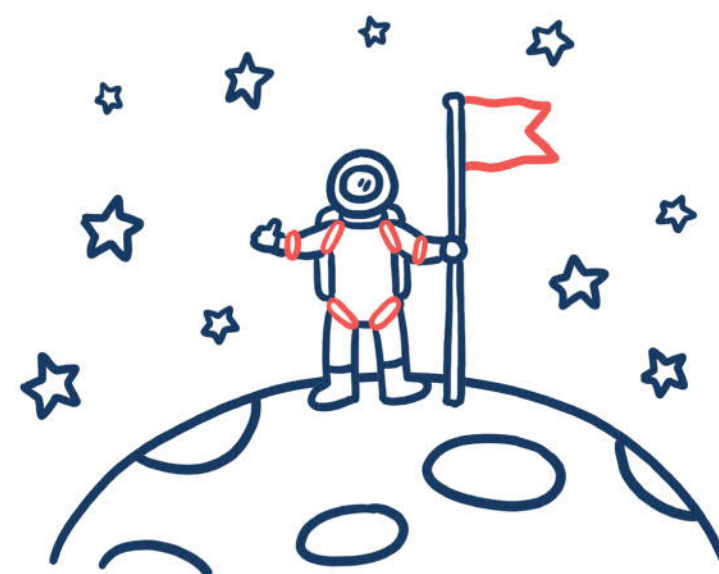


Hãy so bông hoa với một thứ khác như máy tính chẳng hạn.

Máy tính chính là thứ nhân tạo. Nó không thể nào mọc ra từ dưới đất hay treo lủng lẳng trên cây. Chúng ta (con người) đã tạo ra nó. Thế nên, vật **nhân tạo** có nghĩa là **con người chúng ta đã làm ra vật đó**.

TRÍ TUỆ có nghĩa là gì?

Trí tuệ thì khó định nghĩa hơn một chút. Tôi và bạn thì rõ là thông minh rồi. Tôi viết ra cuốn sách này, còn bạn đang đọc và hiểu được những gì tôi viết. Rõ ràng, **con người đều thông minh**. Chúng ta đã chế tạo ra xe ô tô, thiết kế ra những chiếc điện thoại thông minh, khám phá thế giới, đưa con người lên Mặt Trăng, phát hiện ra hố đen, và đã xây dựng nên những nền văn minh phức tạp và kết nối rộng lớn. Tất cả những việc đó đều đòi hỏi phải có trí tuệ.



Vậy chúng ta đã biết con người đều thông minh. So với chúng ta, cây cối không tạo ra được bất cứ thứ gì, cũng không trao đổi, thảo luận được với nhau. Các loại hoa, các loại rau và củ quả cũng thế. Rõ ràng, bất cứ thứ gì được xem là thực vật thì không thể có hành vi nào thể hiện trí thông minh được.

Bạn sẽ hỏi: “Thế hoa hướng dương thì sao? Chúng có thông minh không, tại sao chúng biết quay về hướng Mặt Trời? Cây bầy kẹp còn biết bắt ruồi nữa?” Có một điểm khác biệt nho nhỏ ở đây: