



TRÍ TUỆ NHÂN TẠO

Bài 2: Các quan điểm và cách tiếp cận AI

Nội dung



1. Định nghĩa AI
2. AI và các môn khoa học liên quan
3. Một số nan đề trong phát triển AI



Phần 1

Định nghĩa AI

Định nghĩa AI



- Ra đời mùa hè năm 1956 tại trường Dartmouth (Mỹ), bởi 4 nhà khoa học: **Marvin Minsky, John McCarthy, Allen Newell** và **Herbert Simon**
 - Trước đó chỉ được xem là một mảng kiến thức con trong lĩnh vực khoa học máy tính
- Hiện nay, AI phát triển rộng và được phân chia thành nhiều chuyên ngành con
- Mục tiêu chính: tạo ra các hệ thống có năng lực trí tuệ
- Vấn đề là AI được định nghĩa thế nào? Có nhiều quan điểm, xoay quanh suy nghĩ và hành vi (như) con người
 - Suy nghĩ: thứ vô hình, ẩn bên trong
 - Hành vi: thứ thể hiện ra bên ngoài

Những người khai sinh AI



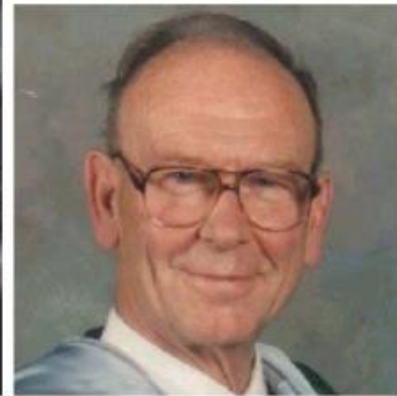
M. Minsky



J. McCarthy



H. Simon



R. Michie

- AI Lab. at M.I.T. (Minsky & McCarthy)
- AI Lab. at Carnegie Mellon Univ. (Simon & Newell)
- AI Lab. at Stanford Univ. (McCarthy)
- AI Lab. at Edinburgh Univ. (Michie)

Hai thành phần của AI



=



Knowledge

(đại số, thống kê,
toán học rời rạc, ...)

+



Inference

(logic toán học, ...)

- Biểu diễn tri thức (knowledge representation)
- Lập luận tự động (automatic reasoning)
- Học tự động (machine learning)
- Hiểu ngôn ngữ tự nhiên (natural language understanding)
- Thị giác máy (computer vision)
- Hệ cơ sở tri thức (knowledge-based systems)
- etc.

AI vs Lập trình truyền thống



	AI	Conventional Programming
Xử lý	Chủ yếu là phi số	chủ yếu là số
Bản chất	Lập luận	Tính toán
Input	Có thể không đầy đủ	Phải đầy đủ
Tìm kiếm	Heuristic (mostly)	Algorithms
Giải thích	Cần thiết	Không nhất thiết
Quan tâm chính	Knowledge	Data, Information
Structure	Tách điều khiển khỏi tri thức	Điều khiển gắn với thông tin và dữ liệu

Định nghĩa AI #1



- **AI = Hệ thống suy nghĩ như con người**
- Mục tiêu: nghiên cứu về phương pháp suy nghĩ và hành vi của con người, mô phỏng những thứ đó trên máy tính
- Mô hình hóa nhận thức
 - Ngành Khoa học Nhận thức
- Có nhiều yếu tố kiến thức Tâm lý học & Thần kinh học
- Những học giả theo trường phái này cho rằng: chương trình chẳng những giải đúng mà còn có các bước giải quyết tương tự như cách giải quyết của con người

Định nghĩa AI #2



- AI = **Hệ thống suy nghĩ hợp lý**
- Mục tiêu:
 - Tiếp cận “các luật của tư duy”
 - Chỉ cần suy nghĩ hợp lý, không nhất thiết phải bắt chước con người
- Thế nào là hợp lý? Tiến trình suy nghĩ đúng là gì?
- Sử dụng các lập luận logic
 - Ví dụ: “Socrates là một con người; tất cả mọi người đều sẽ chết; do vậy, Socrates sẽ chết”
- Trở ngại:
 - Tri thức thường không chắc chắn
 - Thiếu chiến lược tìm kiếm hợp lý

Định nghĩa AI #3



- **AI = Hệ thống hành động như con người**
- Mục tiêu: hành động càng giống người càng thành công, không quan tâm đến cách máy suy nghĩ như thế nào
 - Thử nghiệm Turing (1950): Máy trả lời các câu hỏi giống như con người
 - Thử nghiệm Turing tổng quát: Máy cư xử giống như con người (cả suy nghĩ và hành động)

Định nghĩa AI #3



- Các lĩnh vực nghiên cứu:
 - **Xử lí ngôn ngữ tự nhiên** (giao tiếp văn bản)
 - **Biểu diễn tri thức** (lưu trữ thông tin nhận biết)
 - **Lập luận tự động**
 - **Học máy** (khả năng thích nghi với các hoàn cảnh mới, tìm và ngoại suy các khuôn mẫu)
 - * **Thị giác máy tính**
 - * **Người máy** (vận động)

Định nghĩa AI #4



- **AI = Hệ thống hành động hợp lý**
- Hành động hợp lý: đôi khi không cần logic hoặc không quan tâm đến quá trình suy luận như thế nào, miễn hành động có lý là được
 - Ví dụ: chạm vào đồ nóng thì rút tay lại, nghe nói có bom thì nằm xuống,...
- Những hành động hợp lý thường là kết quả của việc học (suy luận theo mẫu có sẵn) hoặc kinh nghiệm (thống kê)
 - Xử lý được thông tin không chắc chắn



- AI = **Hệ thống thông minh trong lĩnh vực hẹp**
- Xuất phát từ quan điểm cho rằng AI sẽ không bao giờ đạt được đến trình độ về nhận thức và suy luận như con người (có thể huấn luyện để giỏi trong mọi lĩnh vực)
 - One algorithm for all ~ một thuật toán cho tất cả
- Vì vậy AI là các hệ thống thông minh đơn giản sử dụng chỉ để giải quyết một bài toán cụ thể, để hỗ trợ con người là chủ yếu (small AI systems)
 - Hệ thống sẽ được giới hạn trong từng lĩnh vực (domain)
 - Chẳng hạn: xe tự lái, nhận dạng mặt, tổng hợp tiếng nói,...



Phần 2

AI và các môn khoa học liên quan



■ Triết học – Vấn đề:

- Có thể sử dụng các luật hình thức để rút ra các kết luận đúng hay không?
- Trí tuệ tinh thần nảy sinh ra từ một bộ não vật chất như thế nào?
- Tri thức đến từ đâu?
- Tri thức dẫn dắt hành động như thế nào?



■ Triết học – Tác giả:

- Aristotle (384-322 BC): Tam đoạn luận
- Rene Descartes (1596-1650): Nhị nguyên luận
- Francis Bacon (1561-1626): Chủ nghĩa kinh nghiệm
- David Hume (1711-1776): Quy nạp
- Rudolf Carnap (1891-1970): Chủ nghĩa thực chứng logic
- Carnap và Carl Hempel (1905-1997): Học thuyết xác nhận



- Toán học – Vấn đề:
 - Các luật hình thức nào rút ra các kết luận đúng?
 - Cái gì có thể tính toán được?
 - Chúng ta lập luận như thế nào với các thông tin không chắc chắn?
 - Đặc trưng của “thuật toán”



■ Toán học – Tác giả:

- George Boole (1815-1864): Logic mệnh đề
- Kurt Gödel (1906-1978): Incomplete Theory
- Alan Turing (1912-1954): Máy Turing
- Thomas Bayes (1702-1761): Xác suất
- Steven Cook (1971) và Richard Karp (1972): NP-Đầy đủ

AI và các môn khoa học liên quan



- Kinh tế học:
 - Lý thuyết quyết định
 - Lý thuyết trò chơi (Von Neumann và Morgenstern)
 - Tiến trình quyết định Markov
- Thần kinh học
 - Các bộ não xử lý thông tin như thế nào?
 - Neural Network

AI và các môn khoa học liên quan



- Tâm lý học
 - Con người/động vật suy nghĩ/hành động như thế nào?
 - John Watson (1878-1958): Chủ nghĩa hành vi
 - William James (1842-1910): Tâm lý học nhận thức
- Kỹ thuật máy tính
- Lý thuyết điều khiển và Điều khiển học
- Ngôn ngữ học



Phần 3

Một số nan đề trong phát triển AI

Một số nan đề trong phát triển AI



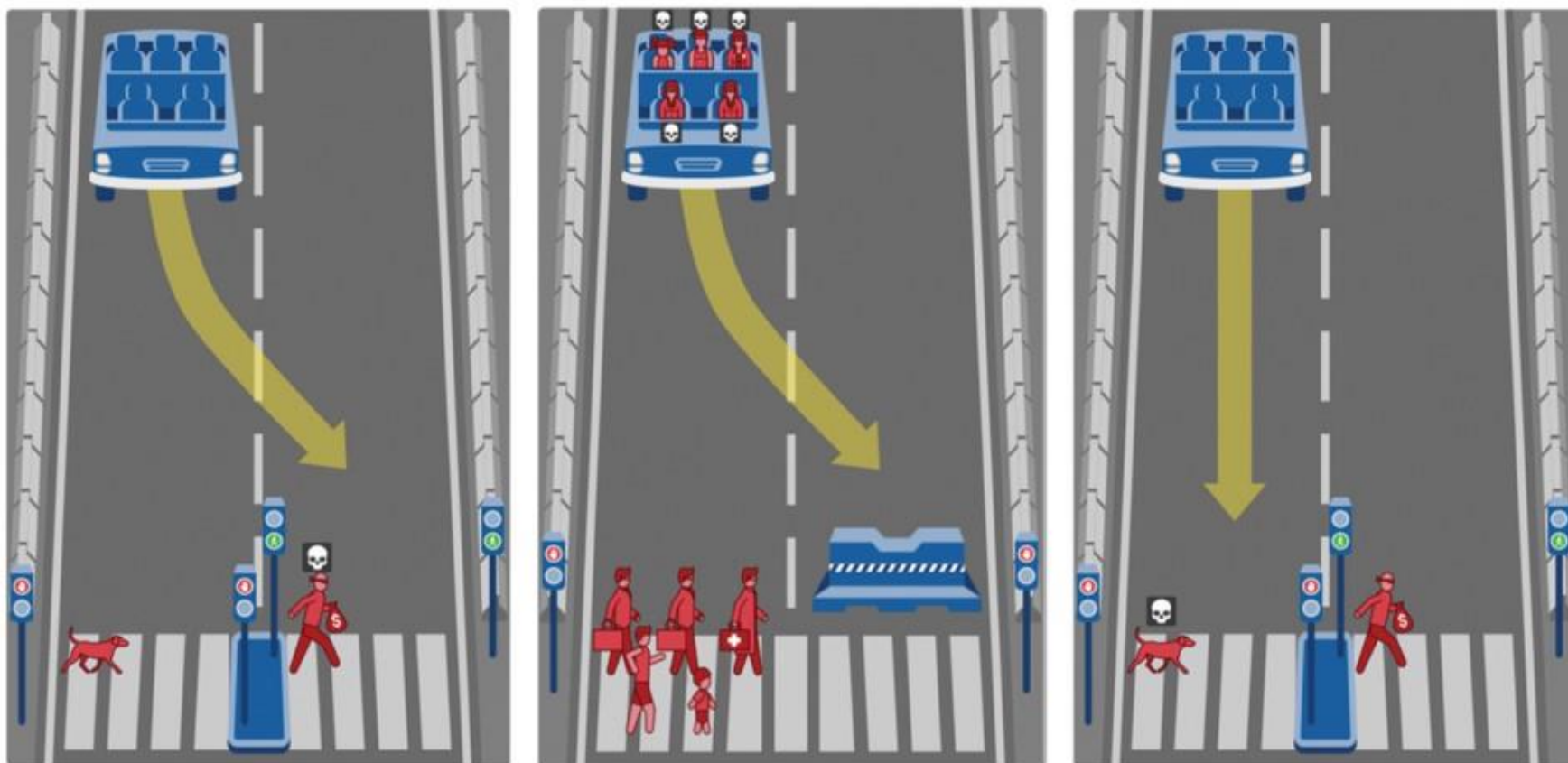
- Việc phát triển và áp dụng các hệ thống AI vào cuộc sống đôi khi phát sinh nhiều vấn đề về đạo đức và pháp luật
- Hiện nay vẫn đang là chủ đề tranh cãi và có thể có nhiều cách xử lý khác nhau tùy theo quan điểm xã hội và môi trường luật pháp
- Một vài “nan đề”:
 - AI có lỗi thì ai chịu trách nhiệm?
 - Nguyên tắc đạo đức khi ra quyết định
 - AI có làm người lao động thất nghiệp?

AI có lỗi thì ai chịu trách nhiệm?



- Ví dụ: 2 người ở hai quốc gia chat với nhau, sử dụng google translate để dịch, nhưng vì dịch sai, nên 2 người mâu thuẫn, vậy Google có lỗi trong tình huống này không?
- Ví dụ khác: một xe tự lái gây tai nạn, vậy trong trường hợp này thì đối tượng chịu lỗi là ai?
 - Chủ của chiếc xe
 - Công ty sản xuất xe
 - Công ty viết phần mềm tự lái
 - AI của xe

Nguyên tắc đạo đức khi ra quyết định



AI có làm người lao động thất nghiệp?



- AI có thể thay thế khá nhiều công việc đơn giản của con người
 - Xử lý các dữ liệu tài chính
 - Xử lý các dữ liệu y học
 - Tự động hóa sản xuất
 - Tự động chăm sóc khách hàng, tư vấn
- AI có thể làm tăng đột biến năng suất của lao động
- Kết quả:
 - Lao động sẽ thất nghiệp?
 - Thu nhập cơ bản (ai cũng có lương dù không làm gì)
 - Kịch bản khác???