

Nguyên lý cơ bản về Trí tuệ Nhân tạo

1. Giới thiệu

Trí tuệ nhân tạo (AI) không phải là một khái niệm mới trong lĩnh vực công nghệ, đặc biệt là trong tự động hóa và cơ giới hóa các quy trình khác nhau được sử dụng trong công việc hàng ngày của chúng ta, cũng như trong các lĩnh vực quốc phòng và các ngành công nghiệp công, tư. Khái niệm cơ bản về AI có từ năm 1940, khi sự kết hợp giữa các máy tự động và các chức năng sinh học được nghiên cứu trong lĩnh vực điều khiển học. Vào thời điểm đó, sự phát triển của các máy móc và các công nghệ tự động hóa cơ học đã đạt đến đỉnh cao do sự cạnh tranh và chiến tranh toàn cầu, đặc biệt là trong Chiến tranh Thế giới thứ II, một trong những yếu tố mạnh mẽ thúc đẩy sự phát triển của các công nghệ điều khiển học tiên tiến dựa trên sự kết hợp giữa sinh vật hữu cơ và tự động hóa quá trình cơ học. Đây được coi là điểm khởi đầu của AI hiện đại.

Theo thời gian, lĩnh vực AI đã trải qua nhiều giai đoạn, vừa kích lệ vừa cản trở sự phát triển của công nghệ sáng tạo này. Ngày nay, một trong những ứng dụng tiên tiến nhất của công nghệ AI chính là nền tảng ChatGPT. Thành tựu của ChatGPT cũng được thúc đẩy bởi các công nghệ tiên tiến trong lĩnh vực AI. Nhiều cơ chế khác nhau, từ các quá trình cơ học được vận hành bởi các máy móc thủy lực, điện tử, cơ khí với dữ liệu và phản hồi từ các quá trình đó để điều khiển và quản lý các quy trình cơ giới hóa, cho đến các thiết bị điều khiển học tiên tiến trong lĩnh vực chăm sóc sức khỏe và các phương tiện bay không người lái (UAV), thậm chí là các hệ thống an ninh mạng tiên tiến đều được hỗ trợ bởi công nghệ AI và các nhánh như học máy (ML) và các lĩnh vực khác.

Khái niệm trí tuệ nhân tạo bắt nguồn từ sự song song với trí tuệ con người, dựa trên bộ não của con người. Khả năng đặc biệt của bộ não con người là quan sát và học hỏi từ kinh nghiệm thế giới thực để đưa ra các quyết định thông minh

dựa trên sự hiểu biết về các sự kiện và môi trường xung quanh. Khả năng này khiến con người trở thành sinh vật mạnh mẽ nhất trong số các loài động vật trên trái đất (theo quan điểm sinh học, con người được xem là động vật). Khái niệm AI khám phá và định nghĩa các quy trình và khả năng của máy móc, đặc biệt là các hệ thống máy tính và phần mềm, để phát triển trí thông minh nhân tạo có thể tương đương với khả năng của trí tuệ con người. Do đó, công nghệ mà trong đó các chức năng, hoạt động và hành vi giống như con người được kiểm soát bởi máy móc được gọi là trí tuệ nhân tạo.

AI là một lĩnh vực công nghệ rộng lớn, tập trung vào các hoạt động tự động được kiểm soát bởi máy móc, đặc biệt là các máy tính hiện đại, với các ứng dụng phần mềm tiên tiến đóng vai trò chủ đạo trong toàn bộ hệ thống. Toàn bộ quá trình phát triển khả năng của AI tập trung vào các thành phần chính giúp các nhà khoa học hoặc kỹ sư xây dựng các khả năng tiên tiến đó. Các thành phần hoặc lĩnh vực cơ bản của AI bao gồm:

- Học tập
- Lập luận
- Giải quyết vấn đề
- Nhận thức
- Hiểu ngôn ngữ

AI được sử dụng rộng rãi trong nhiều lĩnh vực và có thể nói là có mặt trong hầu hết các lĩnh vực của doanh nghiệp và đời sống hiện đại. ChatGPT là một trong những ứng dụng của AI mà cuốn sách này chủ yếu tập trung vào. Các lĩnh vực khác sử dụng công nghệ AI dưới nhiều hình thức và biểu hiện khác nhau bao gồm:

- Tự động hóa công nghiệp
- Ngành sản xuất
- Công nghệ thông tin và truyền thông (ICT)
- Ngành hàng không và quốc phòng

- Ngành an ninh mạng và phát triển phần mềm
- Công nghệ vũ trụ
- Kỹ thuật các loại
- Ngành y tế và dược phẩm

Giá trị thị trường toàn cầu của ngành AI đã tăng mạnh trong vài thập kỷ qua. Theo dự báo của Statista, tổng quy mô thị trường AI trên toàn cầu sẽ đạt giá trị 1887,5 tỷ USD vào năm 2030, từ mức 142,3 tỷ USD vào năm 2022. Quy mô thị trường AI dự kiến sẽ tăng gấp 20 lần so với giá trị thị trường năm 2021 trong khoảng 8 năm tới. Các lĩnh vực quan trọng dự kiến sẽ đạt được mức tăng trưởng cao bao gồm marketing, phân tích dữ liệu, chuỗi cung ứng, robot, y tế, thị giác máy tính (CV), chatbots và nhiều lĩnh vực khác.

Sự tăng trưởng mạnh mẽ trong lĩnh vực này được thúc đẩy bởi một loạt các lợi ích mà AI mang lại cho các ngành công nghiệp. Một số lợi ích quan trọng và mong muốn mà công nghệ AI mang lại cho ngành bao gồm:

- Tiết kiệm thời gian khổng lồ cho các công việc thường ngày
- Cung cấp khả năng chi tiết hóa mọi chức năng và hoạt động
- Xử lý dữ liệu lớn nhanh chóng
- Tăng năng suất và hiệu quả tổ chức một cách đáng kể
- Mở ra các lĩnh vực và cơ hội mới trong kinh doanh và phát triển
- Tạo ra kết quả chính xác và mong muốn nhất
- Hoạt động liên tục 24/7 và sự nhất quán trong năng suất

AI đóng vai trò quan trọng trong nhiều ứng dụng phần mềm tự động hóa hiện đại, các nền tảng và công cụ, mang lại các khả năng nâng cao trong các quy trình của mọi lĩnh vực trong đời sống và kinh doanh. Một số ứng dụng, nền tảng hoặc công cụ nổi bật được hỗ trợ bởi AI bao gồm:

- ChatGPT 4
- Các công cụ tìm kiếm trên web
- Hệ thống đề xuất thương mại điện tử
- Amazon Alexa
- Trợ lý ảo Siri
- Xe tự lái
- Ứng dụng ra quyết định tự động
- Chatbots trên web

Trước khi đi sâu vào các khía cạnh khác nhau của nền tảng ChatGPT, rất quan trọng để hiểu về các khái niệm và công nghệ cơ bản hoạt động ở lõi của nền tảng này. Các công nghệ cốt lõi để hiện thực hóa ý tưởng của nền tảng ChatGPT bao gồm AI và các lĩnh vực con như học máy (ML) và xử lý ngôn ngữ tự nhiên (NLP). Tất cả các khái niệm này sẽ được thảo luận riêng biệt trong các chương tiếp theo. Trong chương giới thiệu này, trọng tâm chính sẽ là công nghệ cơ bản nhất hoạt động ở phía sau nền tảng ChatGPT, được gọi là trí tuệ nhân tạo và sự phát triển, tiến bộ hiện đại, các thành phần, loại hình, khả năng và các khía cạnh khác một cách toàn diện.

2. Lịch sử ngắn gọn về Trí tuệ Nhân tạo (AI)

Lịch sử của AI hiện đại bắt đầu từ những ý tưởng triết học về các máy tự động dưới nhiều hình thức và định dạng khác nhau trong dòng thời gian lịch sử loài người. Nhiều nhà khoa học và nghiên cứu cho rằng một ý tưởng triết học, trừu tượng hoặc huyền thoại luôn là nền tảng của bất kỳ phát minh hay khám phá khoa học nào. Do đó, có thể nói rằng lịch sử AI hiện đại cũng có thể chia thành hai giai đoạn chính:

- **Giai đoạn Triết học**
- **Giai đoạn Khoa học**

Giai đoạn triết học bắt đầu từ rất lâu trước Công nguyên

(trước lịch Dương), truy về những ý tưởng triết học, huyền thoại và các trù tượng khác được viết trong các tác phẩm văn học, truyện và các tài liệu khác của thời cổ đại. Các ý tưởng quan trọng trong thời kỳ này chủ yếu tập trung vào các câu chuyện huyền thoại và các khái niệm trù tượng liên quan đến tôn giáo, truyền thống và văn hóa. Giai đoạn khoa học chủ yếu tập trung vào việc thực hiện và sáng tạo ra các ý tưởng, công thức, máy móc, kỹ thuật hay các thực thể khác có sự hỗ trợ từ các phát minh khoa học hoặc lý luận khoa học.

Lịch sử triết học của AI có thể tìm thấy dấu vết của khái niệm AI dưới hình thức những siêu nhân nhân tạo, người máy cơ học, những thứ siêu nhiên và các khái niệm tương tự mà các nhà triết học đã tạo ra qua các khái niệm tôn giáo và huyền thoại. Những ý tưởng này chủ yếu được tìm thấy trong các sáng tạo của triết học, văn học, nghệ thuật, tâm lý học và các lĩnh vực xã hội khác. Các khái niệm về AI đã xuất hiện từ những sáng tạo trong thời kỳ này, đặc biệt là trong các huyền thoại, nghệ thuật, truyện tưởng tượng, và các phát minh triết học.

Giai đoạn triết học có thể được chia nhỏ thêm thành các mốc thời gian quan trọng để theo dõi sự phát triển các khái niệm về AI:

- **380 TCN–1600:** Nhiều ý tưởng của con người được hình thành dưới dạng người máy cơ học, máy tính và các hệ thống toán học xoay quanh các thiết bị phi con người hoạt động như con người. Những sáng tạo này chủ yếu mang tính triết học mà không có lý luận khoa học cụ thể.
- **1600–1700:** Trong giai đoạn này, ý tưởng về các máy có khả năng “biết tất cả” bắt đầu xuất hiện, như máy "The Engine" trong tiểu thuyết *Gulliver's Travels* của Jonathan Swift.
- **1700-1900:** Tiểu thuyết *Erewhon* của Samuel Butler (1872) mô tả rằng máy móc trong tương lai có thể sở hữu ý thức

giống như bộ não con người.

- **1900–1950:** Đây được coi là thời kỳ giữa của khái niệm AI và là điểm khởi đầu của công nghệ AI hiện đại. Trong thời gian này, nhiều khái niệm giả tưởng xuất hiện, cùng với một số phát triển toán học dựa trên các khái niệm và sáng tạo khoa học thực tế.

Giai đoạn hiện đại của AI:

- **1950:** Alan Turing đưa ra lý thuyết về máy tính và trí thông minh trong cuốn sách của ông, giới thiệu bài kiểm tra Turing để đo lường khả năng trí tuệ của máy móc.
- **1952:** Arthur Samuel phát triển chương trình máy tính chơi cờ Caro tự động học mà không cần sự can thiệp của con người.
- **1955:** John McCarthy đề xuất tổ chức hội thảo đầu tiên về ý tưởng "AI" và lần đầu tiên thuật ngữ "trí tuệ nhân tạo" được sử dụng.
- **1961:** Lần đầu tiên trong lịch sử AI, một robot cơ khí thực tế được tạo ra và triển khai trong dây chuyền lắp ráp của GM.
- **1965:** Joseph Weizenbaum phát triển chương trình máy tính tương tác có tên ELIZA, có thể trò chuyện với con người bằng tiếng Anh.
- **1997:** IBM phát triển máy tính Deep Blue, có thể chơi cờ và đánh bại nhà vô địch cờ thế giới lúc bấy giờ.

Giai đoạn thứ hai của lịch sử AI có thể được bắt đầu từ năm 1950, khi bài kiểm tra Turing được phát triển. Sau thời điểm này, kỷ nguyên AI hiện đại với lý luận khoa học bắt đầu. Lịch sử của kỷ nguyên hiện đại của AI như sau:

- **1950** – Lý thuyết về máy tính và trí thông minh được Alan Turing giới thiệu trong cuốn sách của mình. Cuốn sách này thảo luận về "Trò chơi Khởi đầu", đặt ra câu hỏi liệu máy móc có khả năng học hỏi và suy nghĩ hay không.

Khái niệm này đã phát triển và sau đó trở thành công cụ cơ bản để đo lường khả năng trí tuệ của máy móc, được gọi là bài kiểm tra Turing. Khái niệm về bài kiểm tra Turing sẽ được thảo luận ở phần tiếp theo. Đây cũng được coi là điểm khởi đầu của công nghệ AI hiện đại.

- **1952** – Arthur Samuel phát triển chương trình máy tính chơi cờ Caro tự động học mà không cần sự can thiệp của con người.

- **1955** – John McCarthy đề xuất tổ chức hội thảo về ý tưởng “AI”. Hội thảo này diễn ra vào năm sau và thuật ngữ “trí tuệ nhân tạo” lần đầu tiên được công nhận sau những đóng góp của McCarthy trong lĩnh vực này.

- **1956** – Chương trình AI máy tính đầu tiên được tạo ra bởi ba người: Allen Newell, Cliff Shaw và Herbert Simon.

- **1958** – McCarthy phát triển ngôn ngữ LISP, một trong những ngôn ngữ phổ biến và chủ yếu trong lập trình máy tính AI và các ứng dụng liên quan.

- **1959** – Một lĩnh vực mới của AI được phát hiện và được đặt tên là lĩnh vực học máy (ML), do Arthur Samuel đặt ra.

- **1961** – Lần đầu tiên trong lịch sử AI, một robot cơ khí thực tế được tạo ra và triển khai trong dây chuyền lắp ráp của GM tại nhà máy New Jersey. Robot này được gọi là Ultimate, được phát minh bởi George Devol.

- **1964** – Daniel Bobrow phát triển một chương trình AI có khả năng giải quyết các bài toán đại số. Đây là điểm khởi đầu cho kỷ nguyên ML của AI hiện đại.

- **1965** – Chương trình máy tính tương tác mạnh mẽ mang tên ELIZA được Joseph Weizenbaum phát triển, có thể trò chuyện với con người bằng tiếng Anh. Đây được coi là điểm khởi đầu của giao tiếp giữa con người và trí tuệ nhân tạo trong máy móc.

- **1966** – "Shakey the Robot", người máy điện tử đầu tiên, được phát triển bởi Charles Rosen cùng với đội ngũ hơn 11 người.

- **1968** – Máy tính được lập trình theo thuật toán có tên

HAL xuất hiện trong bộ phim khoa học viễn tưởng 2001: *A Space Odyssey*. Trong phim này, khái niệm kiểm soát tàu vũ trụ và giao tiếp với nó được giới thiệu.

- **1970** – Đại học Waseda Nhật Bản phát triển robot hình người đầu tiên, có khả năng nhìn, giao tiếp hạn chế với con người và di chuyển bằng chi.

- **1973** – Báo cáo "Tình hình nghiên cứu AI" được James Lighthill trình bày tại Hội đồng Khoa học Anh. Báo cáo này không mấy khích lệ cho nghiên cứu AI và dẫn đến việc giảm ngân sách cho nghiên cứu trong lĩnh vực này, giai đoạn này được gọi là "mùa đông của AI".

- **1977** – Bộ phim *Star Wars* được phát hành, giới thiệu các robot giao tiếp với con người và thực hiện các chức năng khác.

- **1979** – Cải tiến lớn về khả năng của robot Stanford Cart, được phát triển lần đầu tiên vào năm 1961, với khả năng di chuyển tự động qua một phòng đầy ghế mà không cần sự can thiệp của con người.

- **1980** – Sinh viên Đại học Waseda phát triển robot WABOT-2, có khả năng chơi nhạc trên các nhạc cụ kỹ thuật số, đọc nốt nhạc và giao tiếp với con người.

- **1981** – Chính phủ Nhật Bản đầu tư 850 triệu USD cho nghiên cứu các hệ thống máy tính thế hệ tiếp theo, có thể thực hiện các hoạt động giống con người như dịch ngôn ngữ, trò chuyện với con người, suy luận như bộ não con người và hiểu hình ảnh.

- **1984** – Bộ phim *Electric Dreams* được phát hành, kể về một máy tính có ý thức trong một câu chuyện tình yêu giữa máy tính, người đàn ông và người phụ nữ, với máy tính tên là Edgar.

- **1986** – Lần đầu tiên, Mercedes Benz phát triển một chiếc xe tự lái. Xe này có thể chạy với tốc độ khoảng 88 km/h trên các con đường không có chướng ngại vật. Đây là một bước tiến lớn trong ngành công nghiệp ô tô hiện đại. Chiếc xe này được phát triển dưới sự giám sát của Ernst Dickmanns.

- **1988** – Báo cáo nghiên cứu khoa học “Lý luận xác suất trong các hệ thống thông minh” được công bố bởi Judea Pearl. Cùng năm đó, chatbot đầu tiên được phát triển bởi lập trình viên Rollo Carpenter, có tên là Jabberwacky. Chatbot này có khả năng mô phỏng cuộc trò chuyện của con người một cách rất thú vị và hài hước. Người sáng tạo ra nó sau đó đã phát triển phiên bản nâng cao của chatbot này vào năm 1990 – Cleverbot.
- **1995** – Chatbot mạnh mẽ khác, ALICE (Artificial Linguistic Internet Computer Entity), được phát triển bởi Richard Wallace, một nhà khoa học máy tính và lập trình viên. Ông lấy cảm hứng từ chatbot trước đó, ELIZA, được phát triển bởi Weizenbaum. ALICE có khả năng thu thập dữ liệu ngôn ngữ tự nhiên.
- **1997** – Một loại mạng nơ-ron có tên là Long Short-Term Memory (LSTM) được phát triển bởi Sepp Hochreiter và Jürgen Schmidhuber, các nhà khoa học máy tính. Cũng trong năm này, IBM phát triển máy tính Deep Blue có thể chơi cờ và sau đó đánh bại nhà vô địch cờ thế giới lúc bấy giờ.
- **1998** – Đồ chơi robot cho trẻ em có khả năng AI, Furby, được phát minh bởi Dave Hampton và Caleb Chung. Đây là đồ chơi AI đầu tiên dành cho trẻ em.
- **1999** – Sony giới thiệu robot mạnh mẽ AIBO – Robot trí tuệ nhân tạo, là một chú chó robot có khả năng hiểu và học từ môi trường, người điều khiển và các robot AIBO khác. Nó có thể thực hiện hơn 100 lệnh mà chủ nhân đưa ra.
- **2000** – Robot KISMET được phát triển bởi Cynthia Breazeal. Robot này có hình dáng như khuôn mặt con người, với đầy đủ các bộ phận như mí mắt, mũi, môi, lông mày, v.v. KISMET có thể nhận dạng và mô phỏng cảm xúc qua diện mạo khuôn mặt của mình. Cùng năm, Honda phát hành robot hình người ASIMO.
- **2001** – Bộ phim khoa học viễn tưởng *A.I. Artificial Intelligence* được phát hành, do Steven Spielberg đạo diễn.
- **2002** – Robot hút bụi tự động Roomba được phát triển

bởi công ty I-Robot. Robot này có khả năng làm sạch bụi và tránh các chướng ngại vật.

- **2004** – Bộ phim khoa học viễn tưởng *I, ROBOT* được phát hành, đạo diễn bởi Alex Proyas. Bộ phim này dự đoán môi trường khoa học của năm 2035.
- **2006** – Thuật ngữ “machine reading” được đặt ra, định nghĩa là “hiểu văn bản tự động không cần giám sát”. Thuật ngữ này được sáng tạo bởi Oren Etzioni, Michele Banko, và Michael Cafarella.
- **2007** – Cơ sở dữ liệu ImageNet được tạo ra, mục đích chính là hỗ trợ nghiên cứu nhận dạng đối tượng qua cơ sở dữ liệu hình ảnh lớn phục vụ cho việc huấn luyện máy tính.
- **2009** – Phát triển xe tự lái của Google là một dự án bí mật, sau đó chiếc xe này đã vượt qua bài kiểm tra tự lái tại bang Nevada vào năm 2014.
- **2010** – Microsoft phát hành trò chơi Kinect qua Xbox 360, đây là ứng dụng đầu tiên có thể nhận diện chuyển động cơ thể thông qua các camera 3D và hồng ngoại. Cùng năm, ImageNet tổ chức cuộc thi nhận dạng hình ảnh quy mô lớn – ImageNet Large Scale Visual Recognition Challenge (ILSVRC).
- **2011** – Hệ thống AI IBM Watson được tạo ra, có khả năng hỗ trợ nền tảng hỏi đáp dựa trên xử lý ngôn ngữ tự nhiên (NLP). Cùng năm, Apple Inc. phát hành trợ lý ảo Siri.
- **2012** – Mạng nơ-ron khổng lồ được huấn luyện bởi các nhà khoa học của Google, Jeff Dean và Andrew Ng, bao gồm 16.000 bộ vi xử lý với hơn 10 triệu hình ảnh mèo không nhãn qua video YouTube.
- **2013** – Hệ thống học máy (ML) có ngữ nghĩa Never Ending Image Learner (NEIL) được phát triển bởi nhóm nghiên cứu tại Đại học Carnegie Mellon.
- **2014** – Cortana được phát hành bởi Microsoft, và cùng năm, Amazon Inc. phát hành trợ lý ảo ALEXA.
- **2015** – Elon Musk, Steve Wozniak và Stephen Hawking ký một lá thư mở kêu gọi cấm AI phát triển vũ khí cho

chiến tranh.

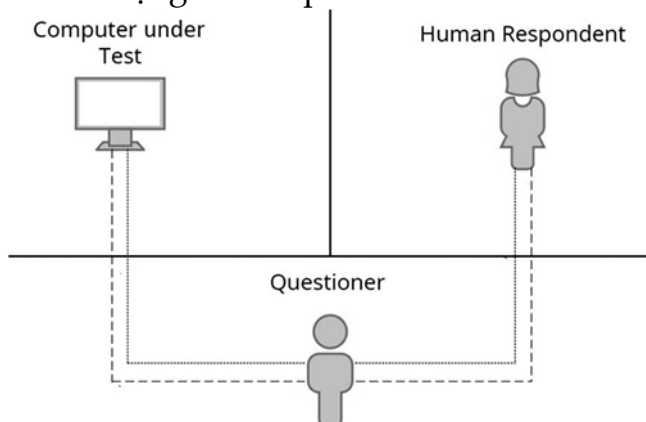
- **2016** – Robot Android SOPHIA được phát triển, được xem là robot công dân đầu tiên do có hình dáng giống con người. Cùng năm, Google phát hành trợ lý thông minh Google Home.
- **2018** – BERT, hệ thống đại diện ngôn ngữ không giám sát hai chiều đầu tiên, được phát triển bởi Google. Nó sử dụng học chuyển giao để thực hiện một loạt các nhiệm vụ NLP tự động. Samsung cũng phát hành trợ lý ảo "Bixby" trong cùng năm.

3. Bài kiểm tra Turing là gì?

Bài kiểm tra Turing là một trong những hệ thống kiểm tra cơ bản và mạnh mẽ nhất, được đề xuất bởi nhà khoa học máy tính người Anh Alan Turing vào năm 1950. Ông đã nghiên cứu về công nghệ AI trong thập niên 1940. Bài kiểm tra này được thiết kế để kiểm tra xem liệu máy tính được thử nghiệm có sở hữu trí tuệ giống như bộ não con người hay không. Bài kiểm tra này còn được gọi là bài kiểm tra bắt chước, trong đó máy tính với một loại trí tuệ nào đó sẽ được kiểm tra qua bộ câu hỏi cụ thể để đưa ra phản hồi từ máy tính (có trí tuệ mong đợi). Nếu máy tính có thể đánh lừa được người chấm thi, máy tính đó sẽ được coi là có AI. Theo định nghĩa của Alan Turing, nếu máy tính có thể bắt chước phản hồi giống như con người, thì máy tính đó được cho là có trí tuệ. Nếu một chương trình máy tính có thể đánh lừa người hỏi trong các điều kiện cụ thể, chương trình máy tính đó sẽ được công nhận là đã vượt qua bài kiểm tra Turing của AI.

Bài kiểm tra Turing bao gồm ba địa điểm riêng biệt, nơi sẽ đặt ba bàn phím và các máy tính đầu cuối. Các máy tính này được kết nối với nhau qua các bàn phím và màn hình. Trong ba máy đầu cuối đó, một máy sẽ được điều khiển bởi chương trình máy tính đang được kiểm tra; máy thứ hai sẽ được điều khiển bởi một con người để trả lời các câu hỏi của người hỏi; máy thứ ba sẽ được điều khiển bởi người chấm thi. Người chấm thi sẽ đặt câu hỏi cho cả hai máy tính

để nhận được phản hồi. Nếu người chấm thi không thể phân biệt được câu trả lời từ hai máy tính ít nhất một nửa số lần, thì máy tính đó sẽ được xem là có trí tuệ. Các câu hỏi được đặt ra trong một phạm vi cụ thể, giới hạn bởi các điều kiện nhất định, và sẽ được hỏi nhiều lần để đánh giá sự khác biệt giữa các phản hồi.



Hình 1.1: Môi trường bài kiểm tra Turing

Trong môi trường được mô tả (Hình 1.1), người hỏi sẽ đặt câu hỏi trong các điều kiện cụ thể, hiển thị trên màn hình của các phòng. Máy tính và người trả lời sẽ cung cấp câu trả lời cho câu hỏi. Khi cả hai câu trả lời được nhận, người hỏi sẽ quyết định câu trả lời nào đến từ máy tính và câu trả lời nào đến từ con người. Quá trình này sẽ được lặp lại nhiều lần. Nếu máy tính có trí tuệ (như đã được định nghĩa), nó sẽ cố gắng đánh lừa người hỏi bằng cách đưa ra các câu trả lời có thể làm người hỏi phân vân không biết câu trả lời đó đến từ con người hay máy tính.

Các khả năng cần thiết để máy tính có thể vượt qua bài kiểm tra Turing bao gồm:

- Khả năng học máy (ML)
- Khả năng xử lý ngôn ngữ tự nhiên (NLP)
- Khả năng điều khiển động cơ
- Khả năng đại diện tri thức
- Khả năng thị giác máy tính (CV)

- Khả năng lý luận tự động

Phiên bản ngược của bài kiểm tra Turing cũng được sử dụng rộng rãi trong các hệ thống bảo mật hiện đại dựa trên AI để kiểm tra xem thực thể tương tác có phải là con người hay không. Bài kiểm tra phổ biến nhất trong các môi trường web hiện đại được gọi là **CAPTCHA** (Bài kiểm tra Turing tự động công cộng để phân biệt máy tính và con người). Đây là bài kiểm tra chống Turing, trong đó người tương tác được xác định qua các hình ảnh, văn bản và các dấu hiệu khác kèm theo câu hỏi đặc thù môi trường để đảm bảo thực thể tương tác không phải là robot.

Các bài kiểm tra phổ biến khác trong lĩnh vực AI hiện đại bao gồm:

- Bài kiểm tra Turing toàn diện
- Bài kiểm tra Turing ngược
- Bài kiểm tra tín hiệu trí tuệ tối thiểu
- Bài kiểm tra Marcus
- Thử thách Winograd Schema
- Bài kiểm tra Lovelace 2.0

Hạn chế của bài kiểm tra Turing:

- Câu hỏi có tính chất hạn chế trong môi trường cụ thể
- Kiến thức hạn hẹp
- Sử dụng nhiều câu hỏi “có” hoặc “không”
- Các câu trả lời trong cuộc trò chuyện có thể tạo ra môi trường phức tạp
- Tương tác giữa con người và máy móc kém hiệu quả
- Giao diện giao tiếp phức tạp và không trực quan

Một số hệ thống AI đã thử vượt qua bài kiểm tra Turing:

- **Parry** – Mô phỏng một người mắc chứng tâm thần phân liệt hoang tưởng.

- **ELIZA** – Là một chatbot sử dụng NLP mô phỏng giao tiếp giữa con người và máy tính.
- **Eugene Goostman** – Ứng dụng AI này đã vượt qua nhiều bài kiểm tra Turing và cũng giành chiến thắng trong một số cuộc thi.

4. Mục tiêu của Trí tuệ Nhân tạo

Từ thời cổ đại của lịch sử loài người, các máy móc đã được tạo ra để làm cho các hoạt động thường ngày trong xã hội trở nên dễ dàng, hiệu quả và năng suất hơn. Theo thời gian, các máy móc mới đã được sáng chế dựa trên những kỹ thuật mới và các phát minh khoa học. Những máy móc này đã nâng cao tốc độ của các quy trình và năng suất tổng thể vì máy móc không bị mệt mỏi như con người. Chúng ta chỉ cần bảo dưỡng máy móc để chúng hoạt động đúng mục đích, thậm chí là liên tục trong thời gian dài.

Ban đầu, các máy cơ học được phát minh để cải thiện hiệu suất của các hoạt động và quy trình trong cuộc sống và kinh doanh đang phát triển. Các máy cơ học đã được bổ sung bằng các máy điện vào một giai đoạn nhất định để tăng năng suất so với các phiên bản trước. Sự khao khát về hiệu quả, đổi mới và năng suất đã dẫn đến sự ra đời của các hệ thống điện tử hiện đại, được sử dụng để điều khiển và tự động hóa các hệ thống mà trước đây phải do con người can thiệp. Hệ thống máy móc ngày nay được vận hành bởi khả năng tính toán tiên tiến và nhiều ngành khoa học hiện đại khác để đạt được sự tự động hóa toàn diện trong các quy trình và hoạt động của doanh nghiệp hiện đại và các hoạt động hàng ngày.

Tuy nhiên, con người vẫn là yếu tố chính trong tất cả các tự động hóa, đặc biệt trong việc ra quyết định và các chức năng khác. Vào một thời điểm nào đó, con người bắt đầu suy nghĩ về việc thay thế khả năng suy nghĩ của con người và tích hợp điều đó vào các máy móc. Ý tưởng sáng tạo này được gọi là

trí tuệ nhân tạo. Ngày càng nhiều nhánh con của công nghệ AI đang dần dần xuất hiện để phát triển trí tuệ trong các máy móc theo cách mà nhiều nhiệm vụ suy nghĩ do con người thực hiện ngày nay cũng được hoàn thành bởi các máy móc hiện đại với khả năng suy nghĩ tương tự con người. Vì vậy, các mục tiêu chính của phương pháp phát triển AI hiện đại là sử dụng hiệu quả sức mạnh của các khoa học và phát minh hiện đại để xây dựng bộ não máy móc có thể kiểm soát tất cả các quy trình và hoạt động mà con người thực hiện trong cuộc sống truyền thống (ít nhất là một số thử nghiệm để đạt được mức độ trí tuệ giống con người!).

Các mục tiêu quan trọng nhất mà các nhà khoa học máy tính và dữ liệu muốn đạt được từ AI bao gồm:

- Giải quyết vấn đề
- Lập luận
- Học hỏi
- Đạt được trí tuệ tổng quát
- Lập kế hoạch có thông tin
- Hiểu biết sâu hơn về dữ liệu

Chi tiết về các mục tiêu chính của công nghệ AI mà các nhà khoa học muốn đạt được sẽ được mô tả trong các phần sau.

Giải quyết vấn đề

Giải quyết vấn đề là một trong những mục tiêu cơ bản nhất của công nghệ AI hiện đại. Việc giải quyết tất cả các loại vấn đề có thể đạt được bằng cách cài đặt khả năng suy nghĩ vào máy móc, để chúng có thể suy nghĩ, học hỏi và hiểu từ kinh nghiệm và môi trường trước đó. Giải pháp vấn đề thông qua trí tuệ máy bao gồm những điểm sau:

- Xây dựng các thuật toán tính toán hiệu quả
- Phát triển khả năng ra quyết định dựa trên thuật toán
- Lập luận logic một cách độc lập
- Mô phỏng quá trình lập luận của con người
- Áp dụng lý thuyết xác suất

- Học hỏi từ kinh nghiệm quá khứ
- Phân tích dữ liệu để đưa ra quyết định

Các giải pháp vấn đề được tập trung giải quyết đến từ tất cả các lĩnh vực trong cuộc sống hiện đại và kinh doanh. Dưới đây là những quy trình quan trọng nhất được hỗ trợ bởi AI, có thể đạt được mục tiêu cơ bản của việc giải quyết vấn đề:

- Tự động hóa quy trình công nghiệp
- Tự động hóa giao tiếp qua chatbot dưới dạng văn bản, giọng nói và video
- Tự động hóa chuỗi cung ứng và hậu cần
- Bảo trì và vận hành các hệ thống tự động
- Xe tự lái
- Máy bay không người lái
- Công nghệ vũ trụ và quốc phòng

Lập luận

Lập luận là một trong những quá trình cơ bản nhất trong bộ não con người để suy ra kết quả từ các tình huống, dữ liệu, sự kiện, điều kiện và niềm tin đã cho. Đây là khả năng của tâm trí, trong đó các quyết định được đưa ra dựa trên quá trình lập luận. Tương tự, lập luận là khái niệm cơ bản trong AI vì nó dựa trên các phương thức mà bộ não con người tuân theo. Các thuật toán AI đưa ra quyết định dựa trên dữ liệu hoặc thông tin có sẵn dưới dạng sự kiện, thống kê, môi trường, điều kiện, và các yếu tố khác. Do đó, có thể nói rằng một trong những mục tiêu chính của AI là lập luận, vì nó đóng vai trò quan trọng trong việc ra quyết định dựa trên thông tin hoặc tri thức có sẵn. Lập luận cũng có thể được định nghĩa như sau: Quá trình suy nghĩ hợp lý chung dẫn đến việc rút ra kết luận hợp lệ.

Các loại lập luận chính thường được áp dụng trong các mô hình và thuật toán AI bao gồm:

- Lập luận suy diễn

- Lập luận theo lẽ thường
- Lập luận quy nạp
- Lập luận đơn điệu
- Lập luận phi đơn điệu
- Lập luận suy diễn ngược

Lập luận suy diễn còn được gọi là loại lập luận hợp lệ, có nghĩa là nếu kết luận của lập luận là đúng, thì các giả thuyết cũng phải đúng. Nó được sử dụng rộng rãi như một loại logic mệnh đề trong lĩnh vực AI. Lập luận suy diễn này cũng được điều khiển bởi nhiều quy tắc, quy định và sự kiện. Lập luận suy diễn còn được gọi là lập luận phản quy nạp hoặc lập luận từ trên xuống. Quá trình này bắt đầu từ các giả thuyết tổng quát và kết thúc bằng kết luận cụ thể. Ví dụ, một giả thuyết là tất cả các loài bò sát đều bò, và một giả thuyết khác là rắn là loài bò sát; kết luận từ hai giả thuyết này có thể dễ dàng rút ra rằng rắn cũng sẽ bò, nhờ vào lập luận suy diễn. Quá trình lập luận suy diễn có thể được chia thành bốn giai đoạn chính:

- Lý thuyết
- Giả thuyết
- Mẫu
- Xác nhận

Lập luận quy nạp là loại lập luận đối lập với lập luận suy diễn. Nó còn được gọi là lập luận từ dưới lên và lập luận nhân quả. Đây cơ bản là một loại logic mệnh đề. Trong loại lập luận này, một chuỗi dữ liệu được sử dụng để đạt được kết luận thông qua quá trình tổng quát hóa. Kết luận cũng được gọi là tuyên bố tổng quát. Sự thật của kết luận không được đảm bảo bởi sự thật của các giả thuyết, do bản chất của xác suất, điều này được sử dụng để tìm ra kết luận trong loại lập luận này. Ví dụ về lập luận này như sau: tất cả các con chim bồ câu đã được

quan sát trong tất cả các vườn thú mà người quan sát đã đến đều có màu trắng. Điều này suy ra rằng tất cả các con chim bồ câu đều có màu trắng dựa trên lập luận quy nạp. Kết luận được đưa ra trong lập luận này dựa trên các quan sát trước đó của người nghiên cứu rằng ở tất cả các vườn thú, chim bồ câu đều có màu trắng; do đó, giả sử rằng tất cả chim bồ câu trên thế giới đều có màu trắng. Tuy nhiên, điều này không đảm bảo rằng chim bồ câu sẽ luôn luôn có màu trắng. Nhưng có khả năng gần như 100% rằng chim bồ câu sẽ có màu trắng dựa trên nghiên cứu và quan sát đạt được qua phương pháp nghiên cứu. Tuy nhiên, cũng có thể chim bồ câu có thể có màu sắc khác. Chúng có thể sống ở các môi trường, cảnh quan, hoặc vị trí địa lý khác, và có các yếu tố khác. Các giai đoạn của lập luận quy nạp có thể được chia thành bốn phần chính:

- Quan sát
- Mẫu
- Giả thuyết
- Lý thuyết

Lập luận suy diễn ngược

Trong lập luận suy diễn ngược, kết luận được tìm ra thông qua những khả năng của điều kiện có khả năng xảy ra nhất. Nói cách khác, kết luận được rút ra từ lập luận suy diễn ngược không được đảm bảo dựa trên các giả thuyết. Đây cũng được xem là một sự mở rộng của lập luận suy diễn. Nó thuộc loại lập luận trong lĩnh vực lý luận logic. Trong loại lập luận này, nhiều quan sát được xem xét để đưa ra kết luận. Ví dụ, một quan sát cho thấy mặt đất sẽ ướt khi trời mưa và một quan sát khác là mặt đất ướt. Kết luận từ hai quan sát này là trời đang mưa. Tuy nhiên, điều này không đảm bảo rằng trời đang mưa; có thể là do một đường ống nước vỡ làm mặt đất ướt hoặc có thể là do mưa

(có thể cả hai hoặc chỉ một trong số đó!). Vì vậy, có thể nói kết luận không được đảm bảo trong loại lập luận này.

Một loại lập luận quan trọng khác là lập luận theo lẽ thường. Trong loại lập luận này, kinh nghiệm đóng vai trò rất quan trọng trong việc rút ra kết luận. Ví dụ, nếu ai đó nhảy xuống nước, họ sẽ bị ướt. Kết luận của bất kỳ giả thuyết nào được xử lý thông qua các quan sát không chính thức từ kinh nghiệm quá khứ. Loại lập luận này dựa trên kiến thức trực giác và các quy tắc để đưa ra phán đoán. Logic chính xác không phải là yếu tố chủ yếu trong loại lập luận này, mà là các quy tắc lẽ thường được sử dụng để đưa ra quyết định. Lẽ thường là khả năng của bộ não con người để nhận thức từ các kinh nghiệm quá khứ và áp dụng chúng vào các giả thuyết hiện có theo cách mà kết luận có khả năng xảy ra nhất.

Lập luận đơn điệu là loại lập luận dựa trên các sự kiện không thay đổi hiện có, có thể được bổ sung thông tin mới hoặc giảm bớt thông tin hiện có, nhưng không làm thay đổi các sự kiện đã được suy ra từ các tuyên bố hoặc giả thuyết. Ví dụ, Trái Đất quay quanh Mặt Trời là một sự thật không thể thay đổi; có thể thêm thông tin như mất bao nhiêu ngày, giờ, phút, giây, nhưng không thể thay đổi sự thật cơ bản này. Tương tự, Mặt Trăng không có ánh sáng riêng, cũng là một sự thật đã được chứng minh và không thể thay đổi. Vì vậy, lập luận đơn điệu sẽ đưa ra quyết định một lần và quyết định đó sẽ luôn giữ nguyên. Các đặc điểm chính của lập luận đơn điệu là:

- Kết luận được đưa ra một lần và luôn luôn sử dụng.
- Thông tin có thể được bổ sung vào kết luận hoặc có thể được sửa đổi nhưng không thể bị bác bỏ.
- Tập hợp các giả thuyết không thay đổi hoặc thay đổi khi thêm thông tin mới vào kết luận chính.

- Được suy ra từ các sự kiện có sẵn và không bị ảnh hưởng bởi thông tin mới thêm vào các sự kiện hiện có.
- Các trường hợp sử dụng phổ biến của lập luận đơn điệu bao gồm các hệ thống lý luận thông thường và các hệ thống logic.

Lập luận phi đơn điệu là loại lập luận liên quan đến các kết luận mà có thể đúng trong lập luận đơn điệu nhưng khi thêm thông tin hoặc sự kiện mới vào các giả thuyết hiện có, tính hợp lệ của kết luận đó sẽ bị thay đổi và trở thành không hợp lệ. Đây là lập luận ngược lại với lập luận đơn điệu, như tên gọi của nó. Các đặc điểm chính của loại lập luận này bao gồm:

- Thêm kiến thức hoặc sự kiện mới vào kết luận có sẵn có thể thay đổi tính hợp lệ của kết luận.
- Các ứng dụng phổ biến của lập luận phi đơn điệu thường là những mô hình logic không chắc chắn hoặc chưa hoàn chỉnh.
- Nhận thức của con người về nhiều vấn đề trong thế giới thực là một ví dụ điển hình về lập luận phi đơn điệu.

Lợi ích của lập luận phi đơn điệu trong nghiên cứu bao gồm:

- Rất hữu ích trong việc xây dựng logic cho robot trong các môi trường thế giới thực.
- Hỗ trợ cả các sự kiện xác suất và giả thuyết để làm cho nó hiệu quả hơn trong nhiều mô hình nghiên cứu và logic khác.

Nhược điểm của lập luận phi đơn điệu bao gồm:

- Tính hợp lệ của kết luận không cố định, do đó nó trở nên phức tạp hơn khi sử dụng làm các thực thể hợp lệ trong các thuật toán hoặc mô hình.
- Không thể sử dụng để chứng minh các định lý.

Ví dụ về lập luận phi đơn điệu:

- Cá có thể bơi.
- Chim không thể bơi.
- Cá ngựa là một loại cá.

Dựa trên các sự kiện trên, chúng ta có thể kết luận rằng cá ngựa có thể bơi vì nó là một loại cá. Tuy nhiên, nếu chúng ta thêm thông tin hoặc sửa đổi thông tin rằng cá ngựa là một con chim, thì kết luận hợp lệ từ bộ dữ liệu này sẽ hoàn toàn bị vô hiệu hóa. Đây chính là lập luận phi đơn điệu trong nghiên cứu và phát triển logic.

Học hỏi

Học hỏi là một trong những khả năng quan trọng nhất của bộ não con người, làm cho nó mạnh mẽ hơn các loài động vật và sinh vật khác trên thế giới. Học hỏi đóng vai trò rất quan trọng trong trí tuệ tự nhiên, vì khả năng này giúp bộ não con người hiểu và học từ kinh nghiệm hoặc môi trường. Sự phát triển của AI dựa trên ý tưởng rằng máy móc có thể bắt chước bộ não con người để thực hiện tất cả các hoạt động và chức năng mà bộ não con người có thể làm. Vì vậy, phát triển khả năng học hỏi trong máy móc là một trong những mục tiêu quan trọng nhất của AI. Quá trình phát triển khả năng học hỏi trong máy móc được gọi là **học máy**.

Theo định nghĩa của **học máy (ML)**, đó là khả năng của máy móc để học từ dữ liệu đầu vào như kinh nghiệm, thông tin và môi trường dưới dạng dữ liệu, và dự đoán kết quả dựa trên kinh nghiệm và dữ liệu đầu vào hiện tại mà không cần lập trình phần mềm cụ thể để dự đoán hoặc đưa ra quyết định.

Thông thường, các chương trình máy tính hoặc ứng dụng phần mềm được sử dụng để thực hiện nhiều chức năng và khả năng tự động qua máy tính. Trong các chương trình đó, máy tính hoạt động như một công cụ mà không có khả năng suy nghĩ. Các chương trình rõ ràng là một tập hợp các hướng dẫn điều khiển máy móc thực hiện các hoạt động nhất định

mà không có sự can thiệp của máy móc để học hỏi và đưa ra quyết định. Quyết định được đưa ra hoàn toàn dựa trên các hướng dẫn, thường ở dạng "nếu thế này thì thế kia". Máy móc không tự đưa ra quyết định. Trong AI, máy móc sử dụng khả năng của chúng để học hỏi thông qua kinh nghiệm về môi trường, thông tin và sự kiện. Vì vậy, một trong những mục tiêu cơ bản của công nghệ AI là làm cho máy móc học hỏi từ kinh nghiệm.

Học máy có hai loại cơ bản:

- **Học có giám sát**
- **Học không giám sát**

Học có giám sát là một loại học máy trong đó dữ liệu đưa vào mô hình học máy được gắn nhãn hoặc gắn thẻ. Điều này được thực hiện thông qua sự can thiệp của con người, bằng cách để máy học từ dữ liệu đầu vào dưới dạng dữ liệu đã được gắn nhãn hoặc gắn thẻ dưới dạng văn bản, giọng nói, video, hình ảnh, v.v. Máy móc sử dụng dữ liệu đó để so sánh và đưa ra quyết định đối với các dữ liệu thử nghiệm mới cần nhận dạng, hiểu biết và ra quyết định để tạo ra kết quả hoặc dự đoán. Học có giám sát sử dụng dữ liệu lịch sử để dự đoán dữ liệu trong tương lai thông qua các thuật toán khác nhau.

Học không giám sát sử dụng dữ liệu huấn luyện mà không cần gắn nhãn hoặc gắn thẻ dưới dạng dữ liệu thô. Các mô hình không nhận thông tin được gắn thẻ, vì vậy chúng cố gắng tìm các đặc điểm tương tự để nhóm các đối tượng giống nhau dựa trên các đặc điểm tương đồng và tách các đối tượng hoặc thông tin dựa trên sự khác biệt. Do đó, chúng cần rất nhiều dữ liệu để phát triển khả năng học hỏi từ dữ liệu huấn luyện đó. Các đặc điểm được sử dụng để thiết lập mối liên hệ có ý nghĩa giữa các đối tượng. Học không giám sát yêu cầu một lượng lớn dữ liệu để phát triển khả năng học hỏi trong máy tính.

Các hạn chế và nhược điểm của các loại học máy:

- Việc gán thẻ và đánh dấu dữ liệu thông qua nhân lực là một thực tiễn tốn kém cho các công ty.
- Học không giám sát cần rất nhiều dữ liệu huấn luyện hơn so với học có giám sát.

Để khắc phục những nhược điểm của các loại học máy chính, một số loại học kết hợp khác được sử dụng trong các thuật toán và mô hình học máy hiện đại:

- **Học bán giám sát**
- **Học củng cố (RL)**

Học củng cố (Reinforcement Learning - RL) là loại học được sử dụng để tăng phần thưởng trong môi trường AI thông qua hành động của một tác nhân trong hệ sinh thái AI. Chương trình máy tính thực hiện hành động ngẫu nhiên trong môi trường và nhận phản hồi dưới dạng phần thưởng hoặc hình phạt của hành động đó. Phản hồi này buộc tác nhân thay đổi trạng thái và hành động của mình để đạt được phần thưởng tối ưu. Phương pháp này được sử dụng rộng rãi trong robot học hiện đại và nhiều ứng dụng khác.

Các đặc điểm chính của học củng cố (RL):

- Các thuật ngữ chính trong học củng cố bao gồm tác nhân AI, môi trường, trạng thái, hành động, phần thưởng, chính sách, giá trị, và Q-value.
- Tác nhân không được lập trình để thực hiện hành động cụ thể, mà hành động của nó là ngẫu nhiên và nhận phản hồi dưới dạng phần thưởng hoặc hình phạt.
- Nguyên lý cơ bản đằng sau hành động của tác nhân là "thử và sai".
- Môi trường trong học củng cố là ngẫu nhiên và không thể dự đoán.

Học củng cố (RL) có thể được chia thành hai loại chính:

- **Củng cố tích cực** (Positive reinforcement)
- **Củng cố tiêu cực** (Negative reinforcement)

Trong học củng cố tích cực, việc bổ sung một yếu tố như giá trị hoặc hành vi có thể được thực hiện để tăng khả năng đạt được kết quả mong muốn.

Đạt được Trí tuệ Tổng quát

Khát vọng về AI bắt đầu với mục tiêu đạt được khả năng nhận thức tương đương hoặc vượt trội hơn so với bộ não con người. Mức độ nhận thức tương đương hoặc tốt hơn con người thuộc về cấp độ **siêu AI**. Có ba loại cấp độ AI chính:

- **Narrow AI** (AI Hẹp)
- **General AI** (AI Tổng quát)
- **Super AI** (AI Siêu)

AI hẹp là cấp độ cơ bản nhất của AI, trong đó một nhiệm vụ hoặc nhóm chức năng cụ thể được thực hiện qua máy móc mà không có sự can thiệp của con người. Hiện nay, AI hẹp là trạng thái hiện tại của AI, trong đó các quyết định cho các nhiệm vụ cụ thể được thực hiện qua máy móc mà không cần sự hỗ trợ của con người. Ví dụ của AI hẹp bao gồm các chatbot hiện có, SIRI, ChatGPT, các công cụ gợi ý khác nhau và các nền tảng khác. Chúng rất khả năng thực hiện một số nhiệm vụ sau khi được huấn luyện với lượng lớn dữ liệu huấn luyện. AI hẹp được gọi là AI yếu vì nó còn rất xa so với mức độ trí tuệ hoặc nhận thức của bộ não con người.

Trí tuệ tổng quát là cấp độ AI tiên tiến thứ hai. Cấp độ này đề cập đến khả năng nhận thức của máy móc mà tương đương với con người trong việc thực hiện các nhiệm vụ đa dạng thông qua sự hiểu biết, học hỏi và quyết định toàn diện. Cấp độ này của máy móc được coi là nhanh hơn, hiệu quả hơn và năng suất hơn so với khả năng của con người. Tuy nhiên, cấp độ này vẫn chưa được đạt được dưới bất kỳ hình thức hoặc thiết kế nào. Nhiều nhà khoa học tin rằng cấp độ này của **học máy (ML)** sẽ không thể đạt được, trong khi một số người vẫn hy vọng (hoặc ít nhất là thúc đẩy ý tưởng này).

Siêu AI là ý tưởng về khả năng nhận thức hoặc trí tuệ của máy móc vượt qua mức độ của con người và nhiều người gọi đây là cấp độ công nghệ ngày tận thế. Ý tưởng này có

nội dung rằng máy móc có thể vượt ra ngoài tầm kiểm soát. Con người có thể cố gắng làm mọi thứ hiệu quả hơn và sử dụng những máy móc này, và cách thức hiện tại mà con người làm việc có thể thay đổi hoàn toàn, dẫn đến một kỷ nguyên mới của sự tồn tại của con người. Có thể, tất cả loài người sẽ thay đổi với trí tuệ tiên tiến được cung cấp bởi máy móc và máy móc có thể thay thế loài người và thậm chí tiêu diệt hoàn toàn cuộc sống tự nhiên truyền thống. Đó sẽ là thời kỳ (dự đoán hoặc tưởng tượng) của trí tuệ siêu việt đạt được thông qua sự kết hợp của các công nghệ AI tiên tiến, sửa đổi bộ não con người và hợp nhất cả hai.

Mục tiêu ban đầu xuất hiện trong suy nghĩ của các nhà khoa học làm việc trong lĩnh vực AI là đạt được mức độ AI tương đương với trí tuệ con người. Mục tiêu đó vẫn chưa được đạt được. Nghiên cứu và phát triển (R&D) tiếp tục để đạt được mục tiêu AI mong muốn, được gọi là **trí tuệ nhân tạo tổng quát (AGI)**. AGI cũng được gọi là **AI mạnh**.

Lập kế hoạch

Lập kế hoạch là một khả năng rất mạnh mẽ của bộ não con người. Lập kế hoạch liên quan đến các quá trình suy nghĩ khác nhau và các chuỗi hoạt động được liệt kê dựa trên các suy nghĩ khác nhau phát sinh trong bộ não con người. Các quá trình suy nghĩ này bị ảnh hưởng bởi việc học hỏi qua kinh nghiệm. Vì vậy, các nhà khoa học và chuyên gia trong lĩnh vực cũng coi lập kế hoạch là một trong những mục tiêu quan trọng nhất của AI. Lập kế hoạch chịu ảnh hưởng lớn từ logic và lập luận của chúng ta. Dưới đây là những thành phần cơ bản nhất của lập kế hoạch do con người thực hiện:

- Xác định mục tiêu
- Mô tả phạm vi
- Xác định các thông số nhiệm vụ
- Mô tả mục tiêu
- Chuỗi hành động cần thực hiện

Tất cả các thành phần trên đều cần nhận thức hoặc trí tuệ để xác định, định nghĩa và thực hiện nhằm đạt được mục tiêu.

Dưới đây là các loại lập kế hoạch phổ biến trong AI:

- Lập kế hoạch không gian trạng thái tiến (FSSP)
- Lập kế hoạch không gian trạng thái lùi (BSSP)

Tổng quan, quá trình lập kế hoạch hiện đại nên được thúc đẩy bởi sức mạnh của AI để làm cho nó trở nên năng suất, hiệu quả và nhanh chóng hơn. Điều này có thể đạt được thông qua phân tích dữ liệu, học hỏi từ các sự kiện và kinh nghiệm trong quá khứ, dự báo, phân tích dự đoán và mô hình tối ưu hóa.

5. Hiểu biết sâu sắc về Dữ liệu

Dự báo tương lai luôn là một vấn đề lớn đối với con người từ thời cổ đại. Mọi người rất lo lắng về việc biết được tương lai của chính mình hoặc tương lai của những thứ xung quanh họ như tương lai của các quốc gia, doanh nghiệp, động vật, chim, và gần như mọi thứ có trong hệ sinh thái tự nhiên. Họ đã sử dụng nhiều cách để dự đoán và tượng tượng tương lai bằng cách sử dụng số học, trực giác, chiêm tinh học hoặc các phương thức huyền bí khác để dự đoán hoặc có cái nhìn về tương lai. Phần lớn những thực hành này là huyền bí vì không có một hồ sơ phân tích chính thức về thông tin ẩn chứa trong kinh nghiệm quá khứ, sự kiện và các mẫu hình.

Ngày nay, việc có cái nhìn sâu sắc về tương lai đã trở thành một mục tiêu quan trọng của AI vì nghiên cứu trong lĩnh vực này đã tiến triển rất nhiều. Mặc dù ban đầu không phải là mục tiêu chính của công nghệ AI, nhưng hiện nay có những người đang nỗ lực làm việc trong hướng này.

Hiện tại, thông tin ẩn chứa trong các sự kiện, mẫu hình và kinh nghiệm đã trở thành yếu tố cơ bản thúc đẩy tất cả các loại hình doanh nghiệp trong thế giới hiện đại. Mọi doanh nghiệp đều sử dụng một khối lượng dữ liệu khổng lồ để phân tích từ nhiều khía cạnh khác nhau nhằm thu thập thông tin quý giá từ nó, từ đó xác định hướng đi cho các kế hoạch tương lai để đạt được các mục tiêu mong muốn một cách hoàn hảo (hoặc gần như hoàn hảo) mà không có sự nghi ngờ hay lệch lạc. Phần lớn dữ liệu trên thế giới hiện nay có ở dạng thô hoặc chưa được tổ chức, thường được gọi là dữ liệu không cấu trúc. Dữ liệu không cấu trúc là loại dữ liệu

lớn nhất được tạo ra hiện nay. Các định dạng chính của dữ liệu không cấu trúc có thể là âm thanh, video, văn bản, số liệu, đọc cảm biến, hình ảnh và các loại khác.

Theo ước tính bảo thủ vào năm 2023, trung bình mỗi ngày, thế giới tạo ra khoảng 329 terabyte dữ liệu dưới cả hai định dạng có cấu trúc và không có cấu trúc. Khối lượng dữ liệu khổng lồ này bao gồm dữ liệu được tạo ra, sao chép và tiêu thụ hàng ngày. Với khối lượng dữ liệu lớn như vậy, nếu được phân tích đúng cách và lọc ra thông tin quý giá từ đó, việc nhìn thấu tương lai và dự đoán tình trạng có thể trở nên khả thi. Phân tích một khối lượng dữ liệu lớn như vậy và tìm ra thông tin quý giá có thể thực hiện được thông qua phân tích dữ liệu lớn, đây là yếu tố đầu vào cơ bản trong AI để dự đoán về tương lai.

Phân tích dữ liệu lớn (big data) có thể giúp dự đoán tương lai của bất kỳ quy trình nào liên quan đến dữ liệu này. Phần lớn sự gia tăng dữ liệu này là do dữ liệu video, tiếp theo là mạng xã hội và ngành công nghiệp trò chơi. Một phần lớn lưu lượng truy cập toàn cầu trên Internet đến từ ba lĩnh vực này.

AI có thể cung cấp cái nhìn sâu sắc về tương lai về các quy trình, hoạt động và dự báo bằng cách sử dụng sức mạnh của phân tích dữ liệu lớn. Sự gia tăng sức mạnh của học máy (ML) hoặc AI chủ yếu là nhờ vào khả năng của công nghệ hiện đại trong việc xử lý dữ liệu lớn được tạo ra bởi các hệ thống hiện đại trên thế giới.

Mối quan hệ giữa ML và dữ liệu lớn có thể được mô tả như sau: Một lượng lớn dữ liệu được tạo ra trong hệ sinh thái hiện đại của các hoạt động kinh doanh hoặc cuộc sống hàng ngày. Quản lý khối dữ liệu lớn này dưới các định dạng và hình thức khác nhau như có cấu trúc, bán cấu trúc và không cấu trúc được gọi là **dữ liệu lớn (big data)**. Dữ liệu lớn này sẽ không thể phát triển mạnh mẽ nếu không có sự ra đời của công nghệ hiện đại. Dữ liệu lớn được xử lý rộng rãi bởi các tài nguyên tính toán lớn, điện toán đám mây và các phương pháp phân tích dữ liệu sáng tạo. Phân tích dữ liệu lớn có nghĩa là thu thập, xử lý, phân tích, lưu trữ và quản lý một lượng lớn dữ liệu với sự trợ giúp của các công nghệ tính

toán, lưu trữ và khả năng vận chuyển dữ liệu.

Quá trình xử lý dữ liệu khổng lồ này được thực hiện qua các mô hình phân tích dữ liệu khác nhau, cung cấp thông tin làm nguồn cho AI hoặc học máy hiện đại. Những mô hình này mô phỏng cách bộ não con người học hỏi, lập luận và dự đoán kết quả. Vì vậy, cả dữ liệu lớn và học máy đều có mối quan hệ chặt chẽ để xây dựng một hệ sinh thái có thể dẫn đến dự đoán tương lai hoặc cái nhìn sâu sắc hơn về dữ liệu để đưa ra các quyết định hiệu quả.

6. Các loại Trí tuệ Nhân tạo

AI là một công nghệ rộng lớn bao gồm nhiều phương pháp, chức năng, khả năng và mô hình để hiện thực hóa khái niệm cốt lõi của công nghệ này. Nó đã được chia thành nhiều thành phần, loại, lớp, danh mục, mô hình và tiểu lĩnh vực trong các ngành công nghiệp hiện đại và hệ sinh thái công nghệ. Về các loại AI, chúng được phân thành hai loại chính. Các loại này tiếp tục được chia thành các tiểu loại. Các danh mục quan trọng nhất để định nghĩa các loại AI dựa trên cả chức năng và khả năng của công nghệ. Chức năng bao gồm các khía cạnh của các chức năng khác nhau mà máy móc thực hiện dựa trên AI. Mặt khác, khả năng định nghĩa loại AI khác dựa trên khả năng của các máy móc sử dụng công nghệ AI.

Vì vậy, dưới đây là hai danh mục chính trong phân loại các loại AI:

- **Các loại AI dựa trên chức năng**
- **Các loại AI dựa trên khả năng**

Trong danh mục đầu tiên, sự phân biệt công nghệ AI được thực hiện dựa trên các chức năng mà các máy móc hỗ trợ công nghệ AI thực hiện. Danh mục thứ hai bao gồm các loại công nghệ AI khác nhau dựa trên các khả năng hoặc tính năng của chúng. Các chi tiết về cả hai danh mục công nghệ AI này sẽ được mô tả riêng trong các phần tiếp theo.

Các loại Trí tuệ Nhân tạo Dựa trên Chức năng

Công nghệ AI đang phát triển và tiến hóa theo nhiều hướng, lớp và danh mục khác nhau. Do đó, việc phân loại các loại công nghệ khác nhau trong một lĩnh vực duy nhất là không khả thi. Hiện nay, các loại công nghệ AI chính được chia dựa trên chức năng và khả năng. Phân loại AI dựa trên chức năng còn được gọi là **Danh mục AI Loại-2**. Các loại AI dựa trên chức năng có tổng cộng bốn loại:

- **Máy móc phản ứng**
- **Bộ nhớ hạn chế**
- **Lý thuyết về tâm trí**
- **Tự nhận thức**

Bây giờ, chúng ta hãy tìm hiểu về tất cả các tiểu loại AI nằm trong danh mục chức năng.

Máy Móc Phản Ứng

Loại AI cơ bản nhất dựa trên chức năng là **máy móc phản ứng**. Như tên gọi, đây là loại trí tuệ của máy móc dựa trên phản ứng của một hành động. Nói cách khác, một máy móc cung cấp kết quả giống nhau dựa trên dữ liệu đầu vào nhất định được gọi là **máy móc phản ứng**. Ví dụ, một người mua sắm trực tuyến hoặc khách truy cập muốn mua sách của mình được sử dụng làm đầu vào cho máy tính để phân tích và cung cấp các đề xuất phản ánh sở thích của họ trong các hoạt động trước đó. Kết quả này luôn được hiển thị cho người dùng mỗi khi họ truy cập một trang web cụ thể để mua sách. Trong loại AI này, máy móc không có khả năng ghi nhớ quá khứ để cho thấy rằng nó đã có trải nghiệm tốt hay xấu và giờ đây thay đổi quyết định dựa trên trải nghiệm đó. Máy chỉ quyết định dựa trên dữ liệu được cung cấp tại thời điểm ra quyết định để tạo ra kết quả.

Các đặc điểm phổ biến của loại máy móc phản ứng trong AI:

- Đây là loại AI cơ bản hoặc nền tảng nhất.

- Nó cũng được gọi là **AI Loại-1** dựa trên chức năng.
- Nó không sở hữu bất kỳ trải nghiệm hoặc ký ức quá khứ nào để tinh chỉnh quyết định với kết quả cụ thể hơn.
- Loại AI này là AI đặc thù nhiệm vụ để thực hiện các công việc lặp đi lặp lại dựa trên một tập hợp thông tin để tạo ra kết quả.
- Các máy móc thuộc loại AI này rất hạn chế về số lượng để thực hiện chức năng hướng tới nhiệm vụ.
- Các hệ thống này không cải thiện hoặc học hỏi từ kinh nghiệm quá khứ của chúng. Nó chỉ là một loại máy móc dựa trên quá trình phản ứng cung cấp kết quả dựa trên thông tin đã cho để đưa ra quyết định dựa trên dữ liệu hoặc thông tin đó.
- Không có bất kỳ kinh nghiệm, thời gian hay thực hành nào làm tăng khả năng của máy móc này.
- Không có khái niệm hoặc sự hiểu biết về quá khứ trong loại máy móc này.
- Các ví dụ của loại AI này bao gồm các công cụ gợi ý, bộ lọc thư rác, trò chơi IBM Deep Blue, Google AlphaGo, v.v.
- Máy không có bộ nhớ nội bộ để ghi nhớ và học hỏi cho các quyết định trong tương lai.
- Thời đại hiện nay của AI đã vượt qua cấp độ chức năng của loại máy móc phản ứng và đã tiến lên cấp độ tiếp theo.
- Loại AI này sử dụng các mô hình học máy tĩnh để dự đoán kết quả dựa trên đầu vào đã cho một cách tĩnh.

Bộ nhớ hạn chế

Đây là cấp độ nâng cao của loại chức năng AI, trong đó công nghệ AI hiện đại đang phát triển (như chúng ta quan sát hiện nay). Công nghệ này đã vượt qua loại AI đầu tiên là máy móc

phản ứng và đang cố gắng hoàn thiện cấp độ AI thứ hai này. Loại này cũng được gọi là **AI Loại-2** dựa trên chức năng. Hai loại đầu tiên hiện nay tồn tại trong lĩnh vực công nghệ hiện đại, trong khi hai loại còn lại của AI dựa trên chức năng chưa thực sự hiện thực hóa. Chúng chỉ tồn tại dưới dạng ý tưởng hoặc khái niệm sẽ được hiện thực hóa trong tương lai. Nói cách khác, ngoài máy móc phản ứng và bộ nhớ hạn chế, tất cả các loại AI dựa trên chức năng khác vẫn chỉ là giả thuyết.

Bộ nhớ hạn chế là khả năng của máy móc để lưu trữ và học hỏi từ kinh nghiệm quá khứ trong một mức độ hạn chế, nhằm đưa ra các quyết định thông minh dựa trên kinh nghiệm hoặc ký ức quá khứ. Các mô hình sử dụng trong loại AI này được triển khai giống như các mô hình của loại máy móc phản ứng, nhưng yêu cầu có bộ nhớ hạn chế để học từ kinh nghiệm quá khứ. Dưới đây là các đặc điểm phổ biến của AI bộ nhớ hạn chế:

- Các loại AI này học từ các dự đoán và ký ức trong quá khứ để đưa ra các dự đoán trong tương lai chính xác hơn.
- Việc sử dụng dữ liệu đã lưu trữ được thực hiện trong một khoảng thời gian ngắn hoặc giới hạn để đưa ra quyết định hoặc dự đoán thông minh.
- Ví dụ điển hình của bộ nhớ hạn chế là xe tự lái, sử dụng công nghệ bộ nhớ hạn chế trong thời gian ngắn để đưa ra quyết định chính xác. Trong mô hình này, tốc độ của các xe xung quanh, khoảng cách của các phương tiện khác, giới hạn tốc độ, và các thông tin điều hướng trên đường được sử dụng để đưa ra quyết định thông minh nhằm tránh các tai nạn hoặc sự cố có thể dẫn đến hậu quả nghiêm trọng.
- Trong loại AI này, máy móc vẫn hoàn toàn là máy móc phản ứng nhưng thêm bộ nhớ ngắn hạn để xử lý sự hiểu biết từ kinh nghiệm quá khứ trong một khoảng thời gian rất ngắn, giúp đưa ra các dự đoán đáng tin cậy.
- Sự hiểu biết từ thông tin ngắn hạn trong quá khứ được

thêm vào chương trình học máy chính để tạo ra các hành động hoặc dự đoán kịp thời, đáng tin cậy hơn và có thông tin hơn, từ đó giúp chương trình thay đổi mô hình trong các mô hình máy móc phản ứng.

- Loại AI này hoạt động theo hai cách khác nhau:
 - Các mô hình được huấn luyện liên tục với dữ liệu huấn luyện bởi một nhóm chuyên gia để hiểu từ kinh nghiệm quá khứ.
 - Các mô hình được huấn luyện tự động bởi các môi trường do các nhà phát triển và chuyên gia tạo ra. Các môi trường này được cập nhật dựa trên việc sử dụng và hành vi của mô hình.
- Mô hình học máy được tích hợp vào cấu trúc máy móc để làm cho chúng thành các mô hình bền vững với bộ nhớ hạn chế.
- Học máy bộ nhớ hạn chế trong máy tính được thực hiện qua sáu bước chính:
 - Dữ liệu huấn luyện là bước đầu tiên để huấn luyện các mô hình học máy bộ nhớ hạn chế.
 - Bước thứ hai là tạo ra mô hình học máy.
 - Bước thứ ba là dự đoán dựa trên việc huấn luyện mô hình.
 - Bước thứ tư là nhận phản hồi từ môi trường học máy hoặc từ sự can thiệp của con người.
 - Bước thứ năm là chuyển phản hồi thành dữ liệu và lưu trữ vào kho học để sử dụng.
 - Bước thứ sáu là khởi động lại toàn bộ quá trình này (lặp đi lặp lại).
- Các mô hình quan trọng được sử dụng để xây dựng máy móc AI loại thứ hai dựa trên chức năng bao gồm:
 - **Mô hình RL** (Học củng cố) – Mô hình này hoạt động dựa trên phương pháp thử và sai. Các lỗi và thử nghiệm được sử dụng để dự đoán quyết định đúng đắn. Mô hình này được sử dụng rộng rãi trong các

ứng dụng trò chơi AI như máy tính chơi cờ, DOTA2, và trò chơi Go.

- **Bộ nhớ ngắn hạn dài hạn (Long Short-Term Memory - LSTM)** – Mô hình này được sử dụng phổ biến để dự đoán các chuỗi. Nó được gọi là mô hình dự đoán chuỗi dựa trên mạng nơ-ron. Mô hình này sử dụng trong môi trường học sâu để dự đoán chuỗi. Trong mô hình này, thông tin được lưu trữ trong một khoảng thời gian tùy ý để dự đoán các thứ tự chuỗi trong mạng nơ-ron tuần hoàn.
- **E-GAN hoặc Mạng đối kháng tạo sinh tiến hóa (Evolutionary Generative Adversarial Networks)** – Mô hình này dựa trên sự tiến hóa của các con đường hoặc dự đoán trước đó. Nó giữ thông tin phát triển qua quá trình tạo sinh. Mỗi lần khi mô hình thay đổi thông tin, con đường hoặc dự đoán thay đổi nhẹ để điều chỉnh hướng đi.

Lý thuyết về Tâm trí

Lý thuyết về tâm trí là một loại AI tiên tiến dựa trên chức năng. Trong loại AI này, máy móc có thể hành xử giống như bộ não con người. Nó được kỳ vọng sở hữu (gần như) tất cả các khả năng mà bộ não con người có như hiểu cảm xúc, cảm nhận hành vi của đối phương hoặc máy móc, và đưa ra các kỳ vọng về cách đối xử và cách xử lý với các đối tượng hoặc máy móc đối lập. Vì vậy, chúng ta có thể nói rằng loại AI này gần như tương đương với mức độ của bộ não con người. Loại AI này được hình thành từ khái niệm tâm lý học gọi là lý thuyết về tâm trí. Trong đó, có ám chỉ rằng bộ não con người có suy nghĩ, cảm xúc và cảm giác. Những điều này có thể ảnh hưởng đến hành vi của con người và cũng có thể định hình các tương tác xã hội với những người khác trong xã hội. Khái niệm về loại hoặc cấp độ lý thuyết về tâm trí của AI dựa trên chức năng giống với lý thuyết tâm trí trong lĩnh vực tâm lý

học cổ biến của loại AI lý thuyết về tâm trí bao gồm:

- Cũng được gọi là **AI Loại-3**.
- Không có máy nào đủ tiêu chuẩn cho loại AI này trong thế giới công nghệ hiện đại ngày nay.
- Đây là loại AI giả tưởng hoặc mang tính khái niệm.
- Nhiều người vẫn hoài nghi về khả năng đạt được cấp độ này của AI, trong khi những người khác rất lạc quan với một số lo ngại và lo sợ về tương lai của loài người và các hệ sinh thái tự nhiên khác.
- Các loại máy móc này sẽ có thể hiểu được cảm xúc của những người và máy móc khác, hoặc thậm chí của các thực thể khác trên thế giới.
- Chúng sẽ cũng có thể điều chỉnh hành vi của mình theo môi trường, tình huống và cảm xúc của các thực thể đối lập.
- Chúng sẽ cảm nhận và kỳ vọng điều gì đó từ các phản hồi mà chúng nhận được từ các thực thể đối lập và cũng phản ứng tương ứng.
- Các dự án hoặc mô hình ban đầu có thể giống với cấp độ ban đầu của AI Loại-3 có thể là xe tự lái. Chúng vẫn chưa có khả năng cảm nhận cảm giác, cảm xúc hoặc tín ngưỡng.
- Vào thời điểm này, một số ứng dụng như SIRI, Alexa và các ứng dụng khác có thể cảm nhận và phản hồi cảm xúc của bạn một cách đơn phương. Nhưng chúng không thể thể hiện cảm xúc hoặc tạo ra cảm giác cảm xúc hoặc tín ngưỡng trong hành vi của mình.
- Sự xuất hiện của trí tuệ cảm xúc nhân tạo (AEI) là bước khởi đầu cho sự phát triển của loại AI này trong tương lai (có thể!).

Tự nhận thức

Tự nhận thức là loại AI tiên tiến và cao cấp nhất. Nó được gọi là AI Loại-4 về mức độ chức năng. Tự nhận thức là máy hoặc hệ thống máy tính có nhận thức về cảm xúc, cảm giác,

tín ngưỡng và các khái niệm khác mà con người có thể cảm nhận, không chỉ ở bản thân mà còn ở các thực thể đối lập. Câu chuyện diễn ra như sau: cấp độ của máy móc này không chỉ có thể cảm nhận và hiểu cảm xúc và các trạng thái cảm xúc của một thực thể đối lập (máy móc hoặc con người) mà còn cảm nhận và quản lý các cảm xúc và khái niệm khác trong chính nó, giống như bộ não con người làm. Đó được gọi là cấp độ tự nhận thức, trong đó hiệu quả và hiệu suất của bộ não nhân tạo sẽ cao hơn rất nhiều so với con người; do đó, có thể, một máy móc cấp độ này có thể bắt đầu thao túng con người thành máy móc và thêm các đặc điểm để biến một bộ não của con người thành một máy siêu thông minh. Vì vậy, sự tồn tại của con người tự nhiên có thể bị đe dọa hoặc biến đổi thành một hệ thống mới và các phẩm chất tự nhiên có thể bị thay đổi xấu hoặc không thể đảo ngược.

Dưới đây là các đặc điểm phổ biến và đặc điểm của loại AI Loại-4 hoặc tự nhận thức:

- Đây là loại AI khi giác ngộ (enlightenment) mà các máy móc sẽ vượt trội hơn con người về mặt suy nghĩ, cảm nhận và khả năng thao túng.
- Đây là một loại AI giả thuyết xa rời với thực tế trong lĩnh vực công nghệ hiện đại (ngay cả với xu hướng và thành tựu hiện nay).
- Nhiều người hoặc chuyên gia trong ngành vẫn hoài nghi về khả năng đạt được cấp độ này của AI. Trong khi đó, nhiều người tin rằng cấp độ này của AI có thể đạt được và kết quả của sự phát triển này sẽ gây hậu quả thảm khốc cho thiên nhiên và sự tồn tại của con người do sức mạnh của sự biến đổi AI tiên tiến.
- Các máy móc tự nhận thức có thể coi bản thân chúng như một thực thể tự nhận thức riêng biệt và bắt đầu sử dụng các thực thể khác, bao gồm cả con người, cho các lợi ích tương ứng của chúng như con người hiện nay với các thực thể và tài nguyên của thiên nhiên.

- Nếu cấp độ trí tuệ này được phát triển nhân tạo, sự tồn tại của loài người có thể bị đe dọa vì các máy móc có ý thức có thể coi con người không phải là thực thể cần thiết cho sự tiến bộ của các doanh nghiệp và hệ sinh thái của chúng.
- Đây cũng được gọi là tương lai của AI.

Các loại Trí tuệ Nhân tạo Dựa trên Khả năng

AI đang phát triển không ngừng về nhiều mặt như khả năng, chức năng, quy mô thị trường, ứng dụng công nghiệp, và nhiều yếu tố khác. Sự tiến hóa liên tục của công nghệ này đang mở ra những chiều kích mới về các loại, lĩnh vực và các khái niệm cũng như tiêu chí khác. Truyền thống, công nghệ AI đã được phân loại thành các lớp chính và các tiểu lớp dưới các danh mục đó. Trong phân loại này, các loại AI đã được chia dựa trên hai tiêu chí chính – chức năng và khả năng. Các loại AI dựa trên chức năng đã được đề cập trong phần trước. Hãy cùng thảo luận về các loại AI khác nhau được định nghĩa dựa trên khả năng.

Dựa trên khả năng, AI được chia thành các loại chính sau:

- **Trí tuệ nhân tạo hẹp (ANI)**
- **Trí tuệ nhân tạo tổng quát (AGI)**
- **Trí tuệ nhân tạo siêu việt (ASI)**

Chi tiết của cả ba lớp chính hoặc loại AI dựa trên khả năng sẽ được mô tả riêng trong các phần tiếp theo.

Trí tuệ nhân tạo hẹp (ANI)

Như tên gọi, loại AI này dựa trên phạm vi khả năng hạn chế của các máy móc được trang bị AI. Đây là một loại AI trong đó các nhiệm vụ cụ thể được thực hiện thông qua các máy AI. Trong hầu hết các trường hợp, các nhiệm vụ đơn lẻ được thực hiện qua các máy này, thuộc phạm vi khả năng của loại AI này.

Các đặc điểm và tính chất chính của loại AI ANI bao gồm:

- **ANI** còn được gọi là **AI yếu** hoặc **AI hẹp**, chính xác

là **AI hẹp**.

- Các nhiệm vụ đơn lẻ là khả năng cơ bản nhất liên quan đến loại AI này, chẳng hạn như nhận dạng khuôn mặt, công cụ tìm kiếm, hệ thống đề xuất, nhận dạng giọng nói, xe tự lái, trợ lý giọng nói, và các ứng dụng khác.
- Cấp độ khả năng AI này đã được đạt được trong các ứng dụng thực tế hiện đại.
- Loại AI này chỉ có khả năng sao chép hoặc mô phỏng hành vi của con người dựa trên một bộ khả năng nhỏ.
- Lĩnh vực AI quan trọng nhất đã đạt được mức độ hẹp của AI là **NLP** (Xử lý ngôn ngữ tự nhiên), xử lý ngôn ngữ trong các ứng dụng hiện đại.
- Cấp độ AI này bao gồm các chức năng của cả **máy móc phản ứng** và **AI bộ nhớ hạn chế** dựa trên chức năng. Do đó, các máy móc sử dụng trong loại AI này không có hoặc chỉ có bộ nhớ hạn chế để học từ dữ liệu quá khứ và kinh nghiệm đã gặp phải trong quá trình học.
- Các ứng dụng phổ biến của các quy trình hoặc máy móc có khả năng tương đương với ANI bao gồm **IBM Watson**, drone công nghiệp, nhận dạng khuôn mặt, **ALEXA** và **SIRI**, bộ lọc thư rác, đề xuất nội dung và các ứng dụng khác.
- Hầu hết tất cả các ứng dụng chính và hệ thống AI đều thuộc loại này hoặc có khả năng cao hơn một chút so với loại này.
- Các máy móc thuộc loại AI này không có khả năng suy nghĩ rộng rãi.

Trí tuệ Nhân tạo Tổng quát (AGI)

AGI là một loại AI tiên tiến với một bộ khả năng rộng hơn. Đây cũng được coi là giai đoạn thứ hai của bộ khả năng AI. Cấp độ khả năng này bao gồm việc suy nghĩ và học hỏi giống như bộ não con người. Tuy nhiên, AGI vẫn chưa được

đạt được. Vì vậy, chúng ta có thể nói rằng nó vẫn đang trong quá trình phát triển và chưa trưởng thành đầy đủ.

Các đặc điểm nổi bật và khả năng của loại AI này bao gồm:

- Khả năng suy nghĩ của loại AI này tương đương với bộ não con người.
- Đây là loại AI giả thuyết hoặc khái niệm, chưa hiện thực hóa trong thế giới công nghệ thực tế.
- Không có ví dụ thực tế nào cho thấy AI đạt được cấp độ này.
- Loại AI này còn được gọi là **AI mạnh**.
- Sự phát triển của cấp độ này có thể kích hoạt cuộc đua tồn tại giữa con người và các máy móc AI.
- Nhiều nhà khoa học cho rằng cấp độ này của AI không thể đạt được trong môi trường thực tế. Tuy nhiên, nhiều người, bao gồm cả Stephen Hawking và nhiều chuyên gia ngành công nghiệp đương đại, tin rằng nó sẽ xảy ra và sẽ gây ra một mối nguy hiểm nghiêm trọng cho nhân loại và các hệ sinh thái tự nhiên với nhiều vấn đề đạo đức.
- Khả năng của con người được truyền tải hoặc thực hiện thông qua quá trình tiến hóa sinh học chậm, lý thuyết cho rằng sẽ mất nhiều thời gian. Nhưng tất cả những điều này có thể xảy ra chỉ trong vài phút đối với các máy móc AI nếu chúng đạt được khả năng tương đương con người. Khi đó, con người sẽ tụt lại phía sau máy móc AI về tốc độ, hiệu suất, năng suất và tiến hóa, điều này có thể dẫn đến hậu quả nghiêm trọng đối với con người (như đã tưởng tượng).
- Cấp độ của những chiếc máy này sẽ có thể suy nghĩ, học hỏi, quyết định và phản ứng với các tình huống một cách độc lập, giống như hoặc thậm chí tốt hơn bộ não con người.
- Mặc dù có những thử thách lớn trên con đường đạt được cấp độ này của nhận thức, các nhà khoa học vẫn

đang theo đuổi mục tiêu đạt được nó.

- Việc sử dụng hiệu quả nhận diện khuôn mặt, nhận dạng cử chỉ, phát hiện chuyển động và nhận diện giọng nói là một số công nghệ tiềm năng có thể thúc đẩy nghiên cứu và phát triển để đạt được cấp độ AGI.

Trí tuệ Nhân tạo Siêu việt (ASI)

ASI là loại AI giả thuyết hoặc khái niệm tiên tiến nhất, trong đó khả năng của máy được kỳ vọng sẽ vượt qua khả năng của bộ não con người về suy nghĩ, học hỏi, quyết định và hành động. Cấp độ AI này sẽ vượt qua sự ưu việt của nhận thức con người thông qua các máy móc được trang bị AI siêu cấp. Các máy của cấp độ ASI chắc chắn sẽ vượt qua con người về tất cả các mặt vì tiến hóa sinh học và tiến hóa số.

Bộ não con người và các khả năng của con người dựa vào tiến hóa sinh học, điều này mất thời gian dài cho các hoạt động khác nhau như học một khái niệm mới, chuyển giao dữ liệu từ người này sang người khác. Trên thực tế, việc phát triển nhận thức ở trẻ em và cải thiện hiệu quả học hỏi qua thực hành là rất tốn thời gian, mất hàng tháng và năm. Nhưng nếu các máy có thể chuyển giao quá trình học hỏi cho máy khác trong vài giây, sự cải thiện sẽ diễn ra nhanh hơn nhiều so với tiến hóa sinh học! Đó chính là những gì đang được hình dung ở đây. Khả năng của AI dưới dạng dữ liệu số có thể được sử dụng và chuyển giao ngay lập tức đến bất kỳ máy móc nào khác. Quá trình huấn luyện các siêu máy có thể phát triển liên tục mà không bị gián đoạn hoặc mệt mỏi.

Trong các trường hợp và điều kiện như vậy, nếu các máy đạt được khả năng của bộ não con người, chúng chắc chắn sẽ vượt qua loài người. Các đặc điểm quan trọng của ASI bao gồm:

- Công nghệ nhận thức thông minh và tuyệt vời nhất trên thế giới.

- Nó sẽ vượt qua khả năng của con người và có thể biến loài người thành nô lệ!
- Đây là loại AI giả thuyết hoặc hư cấu, rất xa với khả năng đạt được trong tương lai gần.
- Nhiều chuyên gia hoài nghi về việc đạt được cấp độ AI này hoặc nhận thức nhân tạo.
- Nếu cấp độ nhận thức này được đạt được, có thể sẽ xuất hiện nhiều mối nguy hiểm và lo ngại cho loài người, hệ sinh thái tự nhiên, đạo đức, tín ngưỡng tôn giáo và nhiều vấn đề khác.
- Cấp độ công nghệ này sẽ có bộ nhớ, lưu trữ dữ liệu, tốc độ xử lý, khả năng ra quyết định, phân tích sâu, kỹ năng giải quyết vấn đề và nhiều khía cạnh khác cao hơn nhiều so với con người, khiến khả năng của con người tụt lại phía sau và sẽ chiếm lĩnh tất cả các lĩnh vực và ngành công nghiệp cùng một lúc.

7. Cấu trúc của Hệ thống Trí tuệ Nhân tạo (AI)

Hệ thống AI có thể xuất hiện dưới nhiều cấu hình khác nhau trong các lĩnh vực và ngành nghề khác nhau. Ví dụ, một hệ thống phần mềm AI hay AI-based chạy dưới dạng chatbot trên một trang web sẽ có hệ sinh thái hoàn toàn khác biệt so với một hệ thống rô-bốt hay xe tự lái. Do đó, về cấu trúc của một hệ thống AI, nó có các thành phần và mục khác nhau. Về mặt khái niệm, một hệ thống AI có thể được chia thành hai thành phần chính:

- **Đại lý thông minh (IA)**
- **Môi trường thông minh (IE)**

Các thành phần khác có vai trò là cầu nối giữa hai thành phần chính của hệ thống AI bao gồm:

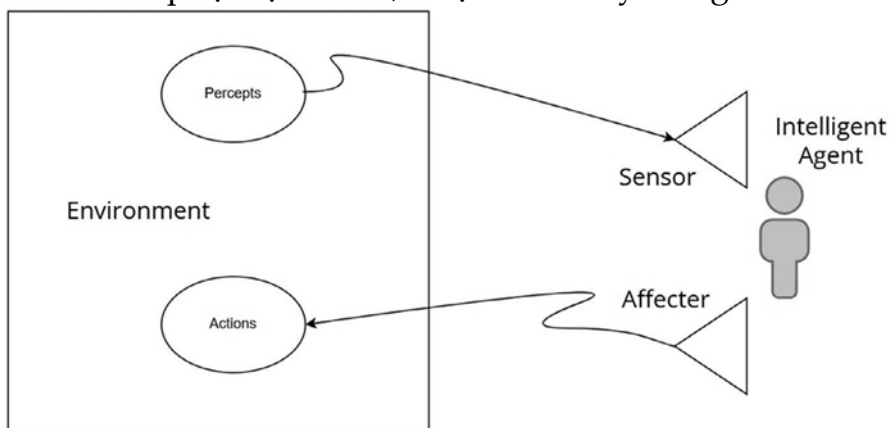
- **Nhận thức (Percepts)**
- **Hành động (Actions)**

Hai thành phần này liên quan trực tiếp đến IA để tương tác với môi trường. Đầu tiên là **nhận thức**, được cảm nhận qua một thiết bị hoặc thành phần có thể cảm nhận đầu vào

từ môi trường AI và truyền tải nó đến IA để hiểu và học từ nhận thức đó. Nhận thức cơ bản là đầu vào mà IA cảm nhận tại một thời điểm cụ thể. **Hành động** là các thành phần do các thực thể vật lý và phần mềm của đại lý điều khiển để thực hiện hành động đối với những nhận thức mà IA đã nhận được. Các nhận thức tương tự như các giác quan của chúng ta qua tai, mũi, mắt, da và các giác quan khác. Sau đó, tay, lưỡi, chân và các bộ phận khác của cơ thể sẽ thực hiện các hành động phù hợp. Nhận thức và hành động cũng được gọi là các thành phần phụ trợ của hệ thống AI, dùng cho các hoạt động phụ trợ để tương tác giữa môi trường và đại lý.

Một nhận thức được nhận qua các thiết bị khác nhau trong AI như camera, microphone ghi âm, các loại sóng điện từ, dòng điện, điện áp và các thiết bị cảm biến khác. Các thiết bị này cung cấp nhận thức cho đại lý để cảm nhận và đưa ra quyết định dựa trên mô hình trí tuệ mà nó sử dụng, từ đó đại lý có thể khởi tạo hành động. Đại lý thực hiện hành động qua các bộ điều khiển hoặc bộ chấp hành trong các loại hệ thống AI khác nhau để đưa ra hành động hợp lý nhất.

Sơ đồ minh họa về hệ thống AI, bao gồm các thành phần cốt lõi và phụ trợ của nó, được trình bày trong Hình 1.2.



Hình 1.2: Sơ đồ minh họa hệ thống AI

Hãy cùng giải thích chi tiết các thành phần của hệ thống AI với các ví dụ tương tự từ trí tuệ con người hay trí tuệ tự nhiên.

Môi trường (Environment)

Môi trường là thế giới rộng lớn hoặc hệ sinh thái của các mục, không gian và dữ liệu dưới các dạng khác nhau tạo thành một hệ sinh thái ngoài đại lý (agent) và cấu trúc toàn bộ của nó trong hệ thống AI. Môi trường này của một hệ thống AI là toàn bộ thế giới ngoài IA.

Nói cách khác, môi trường là một hệ sinh thái trong đó một hoặc nhiều IA có thể sống, hoạt động, và được cung cấp một cấu trúc cho phép các đại lý cảm nhận dữ liệu và phản ứng với hành động đối với dữ liệu đó.

Các đặc điểm và tính chất chung của môi trường AI bao gồm:

- Một số môi trường rất đơn giản và nhỏ gọn, chỉ bao gồm hệ thống máy tính và các phụ kiện với một đại lý trong bộ nhớ máy tính.
- Một số môi trường, đặc biệt là các môi trường lớn và phức tạp dựa trên cả phần cứng và phần mềm, rất phức tạp và chi tiết.
- Một ví dụ về môi trường AI là hệ sinh thái kiểm tra Turing.
- Các môi trường thương mại điện tử và chatbot dựa trên phần mềm hiện đại rất phức tạp, chúng lấy một phạm vi rộng dữ liệu và hành động dưới dạng nhận thức và phản hồi trong môi trường để tạo ra một hệ thống hợp lý.
- Có nhiều loại môi trường dựa trên các đặc tính khác nhau. Một số loại môi trường quan trọng về các chức năng, tính năng và các tiêu chí khác bao gồm:
 - **Môi trường đơn đại lý (Single-Agent) hoặc đa đại lý (Multi-Agent)** – Một môi trường chứa một đại lý gọi là môi trường đơn đại lý và một môi trường chứa nhiều đại lý gọi là môi trường đa đại lý.
 - **Môi trường rời rạc (Discrete) hoặc liên tục (Continuous)** – Nếu các trạng thái của môi trường được xác định rõ ràng với một số đặc tính riêng biệt,

nó gọi là môi trường rời rạc. Khi có nhiều đặc tính và đầu vào thay đổi liên tục, đó là môi trường liên tục.

- **Môi trường tĩnh (Static) hoặc động (Dynamic)** – Trong quá trình hoạt động của đại lý, nếu môi trường thay đổi, nó gọi là môi trường động; nếu không, đó là môi trường tĩnh.
- **Môi trường xác định (Deterministic) hoặc không xác định (Non-Deterministic)** – Nếu với trạng thái hiện tại của môi trường và hành động của đại lý, trạng thái tiếp theo của môi trường có thể được xác định, thì môi trường gọi là xác định; nếu không, môi trường là không xác định.
- **Môi trường có thể quan sát đầy đủ (Fully Observable) hoặc có thể quan sát một phần (Partially Observable)** – Nếu một môi trường hoàn toàn rõ ràng về trạng thái của nó tại bất kỳ thời điểm nào của nhận thức và hành động, nó được gọi là có thể quan sát; nếu không, nó là có thể quan sát một phần.
- **Môi trường có thể tiếp cận (Accessible) hoặc không thể tiếp cận (Inaccessible)** – Nếu tất cả các thiết bị cảm biến liên quan đến đại lý có thể tiếp cận toàn bộ trạng thái của môi trường thông minh, thì gọi là có thể tiếp cận; nếu không, nó không thể tiếp cận.
- **Môi trường theo chu kỳ (Episodic) hoặc không theo chu kỳ (Non-Episodic)** – Những môi trường mà trong đó đại lý có thể thực hiện toàn bộ vòng đời hoạt động trong các tập, chẳng hạn như lấy nhận thức và thực hiện hành động trong một tập, gọi là môi trường theo chu kỳ. Mỗi tập tiếp theo của đại lý không phụ thuộc vào tập trước đó. Nếu không, môi trường được gọi là không theo chu kỳ.

Tác nhân thông minh nhân tạo (IA) là một phần mềm hoặc hệ thống tự động trong một hệ thống AI, có khả năng nhận thông tin qua các cảm biến, đưa ra quyết định dựa trên những thông tin đó và thực hiện hành động một cách tự động hoặc độc lập mà không cần sự can thiệp của con người. Nói cách khác, IA có khả năng nhận thức hệ thống môi trường thông minh thông qua các cảm biến khác nhau. Nó có thể xử lý thông tin dựa trên các khả năng nội tại hoặc thuật toán và thực hiện các hành động phù hợp nhờ vào quá trình ra quyết định hiệu quả thông qua mô hình hoặc thuật toán thông minh bên trong.

Dưới đây là các đặc điểm và tính chất chính của một IA:

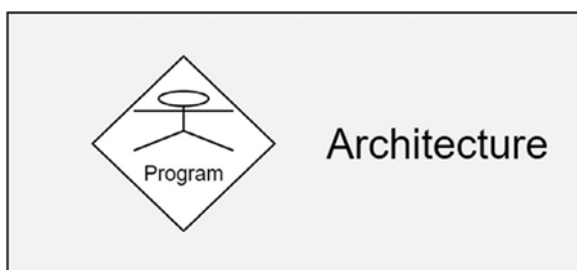
- Là một thực thể tự động hoặc độc lập, tồn tại dưới dạng phần mềm, phần cứng hoặc kết hợp cả hai, để thực hiện các nhiệm vụ khác nhau dựa trên đầu vào nhận được từ môi trường thông minh thông qua quá trình tự đưa ra quyết định.
- Có khả năng học hỏi liên tục từ môi trường.
- IA có thể thực hiện hai loại chức năng – **cảm nhận (perception)** và **hành động (action)**.
 - **Cảm nhận** được đạt thông qua các cảm biến tương tác với môi trường và tác nhân trong hệ thống thông minh.
 - **Hành động** được thực hiện thông qua các thiết bị kích hoạt (actuators).
- IA có thể là một thực thể đơn lẻ hoặc bao gồm nhiều tác nhân phụ trong một hệ thống AI.
- Các **tác nhân phụ và tác nhân** được tổ chức theo cấu trúc phân cấp để thực hiện một chuỗi nhiệm vụ nhằm tạo ra một hành động hoặc phản hồi thông minh hoàn chỉnh.
 - **Tác nhân phụ** chủ yếu thực hiện các nhiệm vụ cấp thấp.
 - **Tác nhân chính** được sử dụng để thực hiện các nhiệm vụ cấp cao trong hệ thống AI.

- IA có khả năng tương tác với nhiều thực thể khác như con người, các tác nhân khác, môi trường, tác nhân phụ, v.v., một cách tự động.
- Hành vi **định hướng mục tiêu** là đặc tính cốt lõi của IA.
- IA tăng cường khả năng của mình thông qua việc học liên tục từ môi trường và kinh nghiệm.
- IA dựa trên tri thức, vì nó sử dụng kiến thức về giao tiếp, các thực thể, quy trình và thông tin môi trường.
- Chất lượng của IA được đo lường thông qua một thuật ngữ gọi là **tính hợp lý của tác nhân (rationality)**.
 - **Tính hợp lý** là thước đo về các phán đoán chính xác, tính nhạy bén, lý lẽ hợp lý và trách nhiệm.
- **Thực hiện các phản hồi hoặc hành động phù hợp nhất** nhờ vào quá trình ra quyết định hiệu quả được xem là đặc tính nổi bật nhất của tác nhân.
- Dựa trên cơ sở tri thức nội tại và chuỗi cảm nhận, IA có thể tạo ra các kết quả hợp lý nếu đáp ứng được các phản hồi mong đợi và đạt hiệu suất tối đa.
- IA giải quyết các vấn đề trong một môi trường nhân tạo được đặc trưng bởi thuật ngữ **PEAS**, viết tắt của:
 - **Performance measure** (thước đo hiệu suất)
 - **Environment** (môi trường)
 - **Actuators** (bộ truyền động)
 - **Sensors** (cảm biến)
- Các yếu tố chính quyết định tính hợp lý của một tác nhân bao gồm:
 - Thước đo hiệu suất
 - Kiến thức ban đầu về môi trường
 - Chuỗi cảm nhận
 - Hành động

IA là một thực thể khái niệm, có thể bao gồm phần cứng, phần mềm hoặc cả hai. Cấu trúc của IA được định nghĩa bởi công thức sau:

- **Tác nhân thông minh = Kiến trúc + Chương trình tác nhân.**
 - **Chương trình tác nhân:** Việc triển khai chức năng của một tác nhân thông qua chương trình phần mềm hoặc thuật toán được gọi là chương trình tác nhân.
 - **Kiến trúc:** Là cấu trúc phần cứng hoặc phần mềm mà IA thực hiện các chức năng của tác nhân.

Sơ đồ cấu trúc của IA được minh họa trong **Hình 1.3**.



Hình 1.4 minh họa sơ đồ khối cấu trúc của tác nhân.

Sơ đồ khối của một tác nhân thông minh nhân tạo (IA) dựa trên cấu trúc được minh họa trong Hình 1.4.

Các loại Tác nhân Thông minh

Tác nhân là thành phần rất cơ bản trong mọi loại hệ thống AI, được sử dụng trong nhiều ngành công nghiệp và lĩnh vực công nghệ khác nhau. Mỗi tác nhân có các khả năng, chức năng, mức độ tri thức, và quyền ra quyết định khác nhau. Tác nhân thông minh (IA) được phân loại thành nhiều nhóm dựa trên khả năng và mức độ thông minh của chúng. Trong số các loại chính, dưới đây là một số loại quan trọng:

1. Tác nhân Phản xạ Đơn giản (Simple Reflex Agents)

Đây là một trong những loại IA đơn giản nhất. Loại tác nhân này thường được sử dụng trong môi trường thông minh có thể quan sát được vì trong các môi trường quan sát một phần, có khả năng xảy ra các vòng lặp liên tục và vô hạn.

Đặc điểm chính:

- Hoạt động dựa trên thông tin cảm nhận hiện tại và không ghi nhớ các thông tin cảm nhận trong quá khứ mà hệ thống thông minh đã quan sát được.
- Có thể thoát khỏi vòng lặp vô hạn bằng các hành động ngẫu nhiên.
- Loại tác nhân này có mức độ thông minh rất thấp hoặc hạn chế.
- Được xem là quá lớn để có thể tạo ra và lưu trữ đầy đủ.
- Sử dụng quy tắc "điều kiện–hành động", nghĩa là nếu điều kiện hoặc trạng thái đúng, hành động sẽ được thực hiện. Điều kiện hoặc trạng thái được ánh xạ tới hành động trong quá trình ra quyết định.
- Không có kiến thức về các thành phần không cảm nhận được của trạng thái hoặc điều kiện.
- Sự thay đổi trong môi trường thông minh sẽ dẫn đến việc cập nhật các quy tắc thu thập.

2. Tác nhân Phản xạ Dựa trên Mô hình (Model-Based Reflex Agents)

Tác nhân phản xạ dựa trên mô hình được định nghĩa đơn giản là tìm một quy tắc có điều kiện khớp với trạng thái hiện tại. Tác nhân này phát triển mô hình của thế giới để xử lý các môi trường quan sát một phần.

Đặc điểm chính:

- Dựa vào lịch sử cảm nhận để cập nhật trạng thái.
- Liên tục theo dõi trạng thái nội tại của nó bằng cách phân tích và cập nhật mỗi thông tin cảm nhận mới.
- Duy trì trạng thái hiện tại bên trong tác nhân, với cấu trúc nhất định mô tả các phần không thể quan sát được của thế giới.
- Việc cập nhật trạng thái yêu cầu hai yếu tố chính:
 1. Cách hành động của tác nhân ảnh hưởng đến thế giới.
 2. Cách thế giới thay đổi một cách độc lập khỏi tác

nhân.

3. Tác nhân Dựa trên Tiện ích (Utility-Based Agents)

Như tên gọi, loại tác nhân này được xây dựng dựa trên nhu cầu của người dùng hoặc tiện ích cuối cùng.

Đặc điểm chính:

- Các hành động được lựa chọn dựa trên sự ưu tiên của tiện ích hoặc trường hợp sử dụng cuối cùng trong mỗi trạng thái.
- Thường được sử dụng trong hệ thống AI để quyết định lựa chọn tốt nhất trong số nhiều tùy chọn.
- Giúp đạt được cách tiếp cận hiệu quả nhất, dễ dàng nhất, nhanh nhất và tiết kiệm nhất để đạt đến mục tiêu.
- "Mức độ hạnh phúc" của tác nhân được coi là thành phần cốt lõi để đánh giá hành động hoặc kết quả hiệu quả của tác nhân.

4. Tác nhân Dựa trên Mục tiêu (Goal-Based Agents)

Trong loại tác nhân này, mục tiêu của hệ thống thông minh là yếu tố cơ bản nhất cần xem xét. Tác nhân sử dụng khoảng cách từ mục tiêu chính hoặc mô tả tình huống mong muốn để đưa ra quyết định trong khái niệm học máy (ML) của hệ thống thông minh.

Đặc điểm chính:

- Hành động được khởi xướng bởi tác nhân được thiết kế để giảm khoảng cách đến mục tiêu.
- Trong tình huống như vậy, tác nhân phải đánh giá và chọn một trong những khả năng hiệu quả nhất để đạt trạng thái mục tiêu của hệ thống.
- Kiến thức được biểu diễn rõ ràng và được sử dụng để hỗ trợ quá trình ra quyết định, giúp tác nhân trở nên linh hoạt và hiệu quả hơn.
- Kiến thức được sử dụng để ra quyết định cũng dễ dàng điều chỉnh trong các tình huống khác nhau, điều này dẫn đến sự thay đổi hành vi của tác nhân một cách dễ dàng.

5. Tác nhân Phân cấp (Hierarchical Agents)

Như tên gọi, loại tác nhân này được tổ chức theo cấu trúc phân cấp từ cấp cao đến cấp thấp. Các tác nhân cấp cao giám sát hành vi của các tác nhân cấp thấp, vốn được thiết kế để thực hiện các hành động hoặc chức năng cấp thấp trong một hệ thống thông minh.

Đặc điểm chính:

- Các tác nhân được tổ chức theo mô hình phân cấp.
- Các mục tiêu và ràng buộc được đặt ra bởi các tác nhân cấp cao, trong khi các tác nhân cấp thấp thực hiện các nhiệm vụ cụ thể.
- Cấu trúc phân cấp này được sử dụng trong các hệ thống thông minh có độ phức tạp cao, yêu cầu thực hiện nhiều nhiệm vụ và chức năng đồng thời.
- Các lĩnh vực ứng dụng phổ biến bao gồm robot, hệ thống giao thông, sản xuất công nghiệp, và các lĩnh vực khác.
- Giải pháp AI tự động được hỗ trợ bởi tác nhân phân cấp phù hợp khi cần ưu tiên và phối hợp hoàn hảo nhiều nhiệm vụ và đầu vào liên quan.
- Thứ tự của các tác nhân trong hệ thống phân cấp phụ thuộc vào mức độ phức tạp của hệ thống.
- Một hệ thống thông minh đơn giản có thể chỉ có tác nhân cấp cao và cấp thấp, trong khi trong môi trường phức tạp, có thể có nhiều cấp tác nhân.
- Trong các hệ thống phức tạp, tác nhân cấp cao đặt mục tiêu, các tác nhân trung gian hướng dẫn và giám sát nhiệm vụ cụ thể của các tác nhân cấp thấp.
- Hệ thống tác nhân phân cấp được coi là hiệu quả và đáng tin cậy nhất so với nhiều loại khác.
- Nó cung cấp khả năng ra quyết định hiệu quả và cải thiện hiệu suất.

6. Tác nhân Học tập (Learning Agents)

Tác nhân học tập, như tên gọi, là một IA có khả năng học hỏi từ kinh nghiệm trong quá khứ khi xử lý lịch sử cảm nhận.

Đặc điểm chính:

- Tác nhân học tập ban đầu được khởi tạo với kiến thức rất cơ bản về môi trường và bắt đầu học từ môi trường thông qua kinh nghiệm.
- Các mô hình học máy (ML) khác nhau được triển khai để liên tục cải thiện quá trình học của tác nhân.
- Về mặt khái niệm, tác nhân học tập được chia thành bốn thành phần chính:
 - **Thành phần Học tập (Learning Element):** Chịu trách nhiệm cho khả năng học hỏi từ môi trường thông qua kinh nghiệm.
 - **Thành phần Phản hồi (Critic Element):** Cung cấp phản hồi cho thành phần học tập về hiệu suất của khả năng học của tác nhân.
 - **Thành phần Hiệu suất (Performance Element):** Đảm nhận việc lựa chọn hành động bên ngoài.
 - **Bộ Tạo Vấn đề (Problem Generator):** Đề xuất các hành động mới để đạt được kinh nghiệm mới và thông tin hữu ích.

7. Hệ thống Tác nhân Đa nhiệm (Multi-Agent Systems – MAS)

Hệ thống tác nhân đa nhiệm bao gồm nhiều tác nhân có khả năng tương tác với môi trường của chúng và phối hợp với các tác nhân khác để đạt được mục tiêu chung hoặc giải pháp tập thể.

Đặc điểm chính:

- Hệ thống này được viết tắt là **MAS** (Multi-Agent System).
- Các hành động của từng tác nhân được phối hợp để góp

phần đạt được một hệ thống tích hợp.

- Mỗi tác nhân có thể đưa ra quyết định, thực hiện hành động, và cảm nhận môi trường trong trạng thái tự trị hoặc bán tự trị.
- Các lĩnh vực ứng dụng phổ biến bao gồm hệ thống giao thông, mạng xã hội, robot, và nhiều lĩnh vực khác.
- Các đặc điểm khác bao gồm:
 - Hiệu suất cải thiện.
 - Tính linh hoạt cao trong môi trường phức tạp.
 - Giảm chi phí.
 - Tăng hiệu quả.

Hệ thống MAS có thể được phân loại thêm thành các danh mục như:

- **Hệ thống đồng nhất (Homogeneous):** Các tác nhân có khả năng, hành vi, và mục tiêu tương tự.
- **Hệ thống không đồng nhất (Heterogeneous):** Các tác nhân có khả năng, mục tiêu, và hành vi khác nhau.
 - Hệ thống không đồng nhất khó điều phối hơn nhưng mang lại tính linh hoạt và mạnh mẽ hơn cho các giải pháp phức tạp.
- **Hệ thống tác nhân hợp tác (Cooperative):** Các tác nhân phối hợp và hợp tác với nhau để đạt mục tiêu chung.
- **Hệ thống tác nhân cạnh tranh (Competitive):** Các tác nhân cạnh tranh để đạt được mục tiêu riêng.

Trong một hệ thống thông minh, hệ thống MAS có thể vừa mang tính cạnh tranh vừa mang tính hợp tác.

Ứng dụng trong AI:

- **Lý thuyết trò chơi (Game Theory):** Phân tích tương tác chiến lược giữa các tác nhân, dự đoán hành vi của chúng.
- **Học máy (ML):** Nâng cao khả năng ra quyết định của tác nhân thông qua huấn luyện với dữ liệu.
- **Mô hình hóa dựa trên tác nhân (Agent-Based Modeling):** Giám sát sự tương tác giữa các tác nhân và giải quyết các

vấn đề phức tạp trong hệ thống thông minh.

Cảm biến (Sensor)

Cảm biến là một thiết bị điện tử, cơ học hoặc loại thiết bị khác có khả năng nhận biết sự thay đổi của các thông số cụ thể trong môi trường AI như nhiệt độ, độ ẩm, ánh sáng, âm thanh, lực đẩy và nhiều thông số khác, sau đó gửi thông tin này đến tác nhân thông minh (IA) trong hệ thống AI. Cảm biến đóng vai trò rất quan trọng trong công nghệ AI hiện đại vì tất cả các loại dữ liệu trong môi trường đều được đo lường và truyền tới tác nhân dưới dạng các thông tin cảm nhận (percepts). Nói cách khác, IA giám sát môi trường với sự hỗ trợ của các cảm biến bằng cách nhận các chỉ số tham số dưới dạng thông tin cảm nhận, được phân loại thành nhiều loại dựa trên các vật liệu, ngành kỹ thuật và khả năng khác nhau. Một số loại chính bao gồm:

- Cảm biến điện tử
- Cảm biến điện từ
- Cảm biến ánh sáng
- Cảm biến nhiệt
- Cảm biến cơ học
- Cảm biến cảm xúc
- Cảm biến quán tính
- Cảm biến mô-men quang điện tử
- Cảm biến lực mô-men cơ học
- Cảm biến sinh học
- Cảm biến khoảng cách

Các loại cảm biến phổ biến nhất được sử dụng rộng rãi trong lĩnh vực AI hiện đại bao gồm:

- Camera
- Micro
- Nhiệt kế
- Cảm biến RFID (nhận dạng tần số vô tuyến)

- Hệ thống radar
- LIDAR
- Sonar
- Hệ thống siêu âm
- Cảm biến mã

Một loạt các mạch điện mới đã được phát triển để nhận biết nhiều tham số khác nhau có thể ứng dụng trong AI hiện đại như luồng khí, áp suất, vị giác, mùi, mưa, độ ẩm, và nhiều thông số khác.

Bộ tác động (Affecters)

Bộ tác động là các thiết bị hoặc thành phần được sử dụng trong hệ thống thông minh để tác động đến môi trường thông minh. Các bộ tác động được kích hoạt thông qua hành động hoặc phản hồi của IA nhằm ảnh hưởng đến các thành phần của môi trường. **Các bộ tác động phổ biến nhất trong AI hiện đại bao gồm:**

- Chân giả
- Tay giả
- Ngón tay giả
- Bánh xe giả
- Bánh răng giả
- Máy thổi khí
- Mũi điện tử

Bộ chấp hành (Actuator)

Bộ chấp hành là các thiết bị cơ học hoặc điện cơ học có chức năng chuyển đổi năng lượng thành chuyển động. Nói cách khác, bộ chấp hành thường là các thiết bị cơ học hoặc điện được sử dụng để truyền các hành động vật lý thông qua một hệ thống tự động, đặc biệt là các thực thể robot, đến môi trường do IA điều khiển. Bộ chấp hành thường được sử dụng trong lĩnh vực robot và các quy trình tự động hóa công nghiệp trong sản xuất,

logistics, và các lĩnh vực khác .

8. Các lĩnh vực khác nhau của Trí tuệ nhân tạo (AI)

AI là một lĩnh vực đang phát triển mạnh mẽ, đặc biệt trong vài thập kỷ qua; hơn nữa, tương lai của công nghệ AI rất tươi sáng. Sự phát triển vượt bậc này cùng những triển vọng trong tương lai liên tục mở ra các lĩnh vực và miền mới trong không gian công nghệ mở rộng này. Nhìn chung, khái niệm AI bao hàm một phạm vi rộng lớn nhằm mô phỏng tất cả các chức năng, hành vi, hoạt động, khả năng nhận thức, lập luận, ra quyết định và các đặc điểm khác của con người với sự hỗ trợ từ các máy móc hoặc máy tính thông minh.

Để đạt được mức độ thông minh như vậy, nhiều dự án và nghiên cứu đang được thực hiện đồng thời tại các quốc gia, viện nghiên cứu, công ty và trường đại học. Với những tiến bộ lớn trong công nghệ xử lý ngôn ngữ, các ứng dụng tạo nội dung và chatbot đã trở thành các thành phần chính trong lĩnh vực AI. Theo thời gian, các lĩnh vực và miền mới tiếp tục xuất hiện trong miền công nghệ rộng lớn này. Một số lĩnh vực quan trọng bao gồm:

- Điện toán nhận thức (Cognitive Computing - CC)
- Học máy (Machine Learning)
- Học sâu (Deep Learning)
- Mạng nơ-ron nhân tạo (Artificial Neural Networks - ANNs)
- Thị giác máy tính (Computer Vision)
- Xử lý ngôn ngữ tự nhiên (Natural Language Processing)
- Hệ chuyên gia (Expert Systems - ES)
- Học tăng cường (Reinforcement Learning)
- Robot (Robotics)

Hãy cùng tìm hiểu các lĩnh vực chính này của AI trong các phần tiếp theo.

Điện toán nhận thức (Cognitive Computing - CC)

CC là một miền con của lĩnh vực AI rộng lớn. Mục đích chính của CC là hỗ trợ con người trong việc ra quyết định và tư duy thông qua việc mô phỏng các quá trình ra quyết định và tư duy của não bộ con người. Đây là một công nghệ hỗ trợ giúp não bộ con người đưa ra các quyết định có cơ sở hơn. Những đặc điểm và tính năng quan trọng nhất của CC bao gồm:

- CC và AI thường được sử dụng thay thế cho nhau, nhưng thực tế chúng khác nhau.
- CC là một miền con của công nghệ AI.
- Đây là một công nghệ hỗ trợ dựa trên khả năng AI của máy móc nhằm giúp não bộ con người đưa ra quyết định dựa trên thông tin và nâng cao quá trình tư duy của con người bằng các kỹ thuật mô phỏng.
- Mục tiêu chính của AI là thay thế tư duy và ra quyết định của con người bằng sức mạnh của máy móc, trong khi mục tiêu cốt lõi của CC là nâng cao khả năng tư duy và ra quyết định của não bộ con người.
- CC cung cấp cái nhìn sâu sắc vào các vấn đề phức tạp mà não bộ con người không thể phân tích để tìm ra giải pháp hiệu quả và có cơ sở.
- Các quá trình tư duy khác nhau được mô phỏng trong CC để khám phá các giá trị ẩn trong dữ liệu vấn đề và đưa ra giải pháp phù hợp nhất.
- Một ví dụ về CC là ứng dụng trong chăm sóc sức khỏe hoặc điều trị, có thể xem xét tất cả các yếu tố như tuổi tác, lịch sử y tế, đặc điểm di truyền, kiểu cơ thể, cân nặng, chiều cao, dị ứng và các thông tin khác của bệnh nhân để đưa ra một kết quả thông tin cao giúp bác sĩ quyết định điều trị chính xác hơn.
- So với AI, có thể nói rằng CC tập trung vào việc tăng cường tư duy của con người, trong khi AI tập trung vào tự động hóa tư duy của con người thông qua máy móc.
- Để chọn cách tốt nhất trong nhiều lựa chọn, CC được

sử dụng thay vì công nghệ AI.

- CC cho phép con người làm việc hiệu quả với khối lượng lớn dữ liệu phi cấu trúc nhằm tìm ra mối quan hệ tốt nhất để hỗ trợ quyết định của con người.
- Quyết định cuối cùng không được thực hiện bởi các nền tảng CC, khác với công nghệ AI nơi quyết định cuối cùng được đưa ra bởi công nghệ AI.
- Các đặc điểm khác của CC bao gồm tính tương tác, khả năng thích nghi, trạng thái liên tục, lặp lại và tính ngữ cảnh.
- Độ chính xác trong phân tích, trải nghiệm người dùng, hiệu quả quy trình, năng suất và chất lượng dịch vụ cao hơn là một số lợi ích chính của CC.

Học Máy (Machine Learning - ML)

ML là một lĩnh vực rất quan trọng và đầy triển vọng của công nghệ AI. Nó được định nghĩa là một lĩnh vực công nghệ trong đó máy móc được đào tạo bằng dữ liệu huấn luyện để tự học và đưa ra quyết định một cách tự động mà không cần lập trình rõ ràng hoặc bất kỳ sự can thiệp nào từ con người. Máy tính được đào tạo thông qua các mô hình ML, được hỗ trợ bởi các thuật toán ML viết bằng mã ngôn ngữ máy. Công nghệ ML đã phát triển thành nhiều loại và phân lớp, tiếp tục mở rộng đáng kể trên nhiều ngành công nghiệp và lĩnh vực. Các ứng dụng phổ biến nhất trong đời sống và kinh doanh hàng ngày bao gồm:

- Công cụ tìm kiếm
- Công cụ gợi ý
- Xe tự lái
- Nền tảng mạng xã hội
- Môi trường an ninh mạng
- Ứng dụng thương mại điện tử
- Chatbots

- Chăm sóc sức khỏe dự phòng
- Ứng dụng nhận dạng hình ảnh
- Ứng dụng nhận dạng giọng nói
- Ứng dụng phân tích cảm xúc
- Ứng dụng dự đoán/phân tích xu hướng

Các lĩnh vực con của công nghệ Học Máy

Công nghệ ML bao gồm nhiều lĩnh vực con, dựa trên nguyên tắc huấn luyện máy móc bằng các loại thuật toán nhất định. Các loại chính của ML bao gồm:

- **Học có giám sát (Supervised ML):**

Đây là loại ML mà máy móc được đào tạo bằng cách sử dụng các tập dữ liệu huấn luyện có nhãn hoặc không nhãn. Trong loại ML này, đầu vào và đầu ra của một thuật toán được xác định trước. Thuật toán được thiết kế để ánh xạ hoặc đánh giá dữ liệu gắn nhãn đầu vào với kết quả hoặc đầu ra đã được định nghĩa trước.

- **Học không giám sát (Unsupervised ML):**

Đây là phương pháp trong đó các nhà khoa học dữ liệu sử dụng dữ liệu không gắn nhãn hoặc không đánh dấu làm đầu vào cho các thuật toán để chúng tự học từ các điểm tương đồng hoặc khác biệt giữa các đặc điểm hoặc thực thể. Thuật toán được thiết kế để tìm ra bất kỳ liên kết nào giữa hai hoặc nhiều thực thể để nhóm chúng lại hoặc tách chúng ra. Trong phương pháp này, dữ liệu huấn luyện và đầu ra cũng không được định nghĩa trước.

- **Học bán giám sát (Semi-supervised ML):**

Đây là phương pháp trung gian giữa học có giám sát và học không giám sát. Học có giám sát có độ chính xác và độ tin cậy cao, nhưng chi phí gắn nhãn hoặc đánh dấu khối lượng lớn dữ liệu là rất cao. Học

không giám sát có chi phí thấp hơn, nhưng hiệu suất thấp hơn. Do đó, mô hình trung gian, được gọi là học bán giám sát, là sự cân bằng tốt nhất giữa hai phương pháp này.

Ngoài ba loại chính này, còn có một loại ML khác được gọi là **Học tăng cường (Reinforcement ML)**, với các đặc điểm độc đáo và được coi là một lĩnh vực riêng trong AI.

Các đặc điểm và chức năng chính của Học Máy

- ML được thực hiện thông qua các thuật toán, xác định đầu vào và đầu ra hoặc kết quả, và đánh giá kết quả dựa trên các định nghĩa đó.
- Không sử dụng lập trình rõ ràng để thực hiện các hành động cụ thể. Chương trình được định nghĩa thông qua thuật toán, viết bằng ngôn ngữ máy và được cài đặt như phần mềm cơ bản để đạt được mức độ học và ra quyết định mong muốn dựa trên các tập dữ liệu huấn luyện.
- **Các hoạt động hoặc chức năng toán học quan trọng nhất của thuật toán ML bao gồm:**
 - Mô hình hồi quy
 - Các loại phân loại khác nhau
 - Hoạt động tổ hợp
 - Chức năng phân cụm
 - Khai thác mối quan hệ
 - Giảm chiều dữ liệu
 - Phát hiện bất thường
- ML cung cấp cho máy móc khả năng dịch, hiểu, thực thi và phân tích dữ liệu để giải quyết các vấn đề phức tạp nhất một cách tự động mà không cần sự can thiệp của con người trong suốt quá trình hoặc chu kỳ hoạt động.
- Cung cấp dự đoán và phân tích hành vi chính xác và hiệu quả trong môi trường thời gian thực.

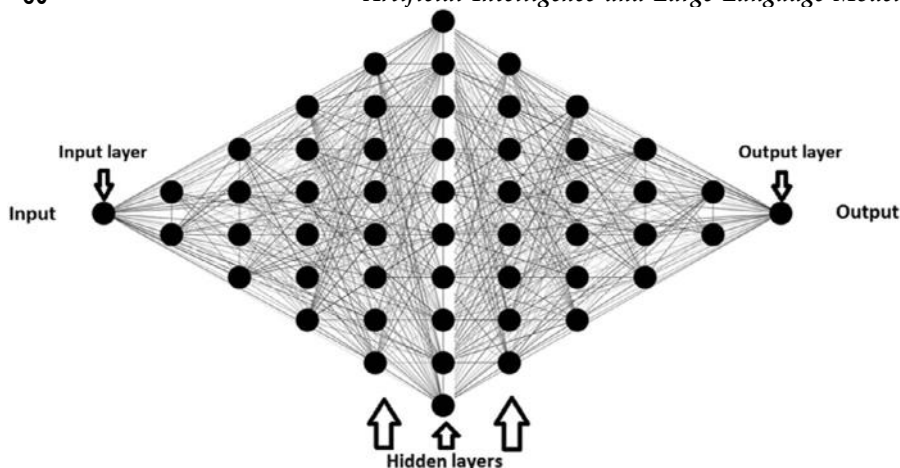
Ứng dụng chính của Học Máy trong kinh doanh

- ML đã hoạt động trong các lĩnh vực như Quản lý quan hệ khách hàng (CRM), Trí tuệ kinh doanh (BI) và Khai phá dữ liệu (DM), Hệ thống thông tin nhân sự (HRIS), trợ lý ảo, chatbot trên trang web và nhiều ứng dụng tự động hóa quy trình khác.
- ML cung cấp phân tích dữ liệu chính xác để tạo ra thông tin kinh doanh giá trị, giúp tăng cường sự hài lòng của khách hàng và năng suất kinh doanh.

Học Sâu (Deep Learning)

DL là một loại AI quan trọng hiện nay. Cụ thể, nó là một dạng chuyên biệt của ML được sử dụng để huấn luyện máy tính từ dữ liệu, giống như cách não người tiếp nhận để hiểu và học từ kinh nghiệm hoặc ví dụ. DL sử dụng các mạng nơ-ron lớn với nhiều lớp ẩn và khối lượng dữ liệu khổng lồ dưới dạng hình ảnh, văn bản, giọng nói và các dạng khác.

DL cung cấp mức độ chính xác cao nhất mà ngay cả não người cũng không thể đạt được trong một số trường hợp. Cấu trúc của DL khác với ML thông thường ở nhiều khía cạnh, chẳng hạn như số lượng lớp trong mạng nơ-ron, khối lượng dữ liệu, sự cải thiện liên tục về khả năng của máy móc, công suất xử lý cần thiết và các yếu tố khác. Sơ đồ nguyên lý của một mạng nơ-ron DL được hiển thị trong Hình 1.5.



HÌNH 1.5: Sơ đồ nguyên lý của một mạng nơ-ron được sử dụng trong học sâu. (Nguồn: Pixabay.)

Dưới đây là các đặc điểm và tính năng quan trọng nhất của công nghệ học sâu (DL)³⁸:

- Trong mô hình ML này, máy móc học bằng cách tiếp nhận khối lượng dữ liệu gần như khổng lồ, tương tự như cách não người học từ các ví dụ hoặc kinh nghiệm với dữ liệu thực tế dưới dạng hình ảnh, cảm giác, văn bản, giọng nói và các dạng khác.
- Các mô hình DL học trực tiếp từ tập hợp hình ảnh, giọng nói và văn bản thông qua nhiệm vụ phân loại.
- Nó cung cấp hiệu suất và độ chính xác cực kỳ đáng tin cậy, thậm chí vượt qua cả khả năng của não người trong một số trường hợp.
- So với ML thông thường, kỹ thuật có giám sát và không giám sát, các mô hình DL sử dụng mạng nơ-ron với số lượng lớp ẩn lớn hơn nhiều lần – trong một số trường hợp, số lớp có thể vượt quá 100.
- DL sử dụng sức mạnh tính toán khổng lồ thông qua GPU (bộ xử lý đồ họa) hiệu suất cao với cấu trúc tính toán song song để đáp ứng yêu cầu xử lý.
- Với sự phát triển của điện toán đám mây và kỹ thuật phân cụm (kết hợp cả hai), thời gian huấn luyện các mô hình DL đã được giảm đáng kể.

- DL sử dụng hàng triệu hình ảnh và các dạng dữ liệu khác cho mục đích huấn luyện. Tập dữ liệu huấn luyện thường được gán nhãn.
- Trong một số trường hợp, các mô hình DL với số lượng lớn lớp ẩn vượt trội hơn con người về độ tin cậy, độ chính xác và hiệu quả.
- Các ứng dụng quan trọng nhất của các mô hình DL trong ngành công nghiệp và kinh doanh hiện đại bao gồm:
 - Lái xe tự động và hàng không vũ trụ
 - Tự động hóa công nghiệp và nghiên cứu y học
 - Quốc phòng và điện tử
- Trong DL, các mạng nơ-ron tích chập (CNN) được sử dụng rộng rãi.
- Việc tự động trích xuất đặc trưng qua mạng nơ-ron tích chập thay thế cho việc trích xuất đặc trưng thủ công trong các mô hình ML thông thường.
- Nó đặc biệt phù hợp cho các ứng dụng thị giác máy tính (CV) nhờ khả năng tự động trích xuất đặc trưng, mang lại độ chính xác và độ tin cậy cao hơn.
- Công nghệ DL cũng được coi là một mô hình học tập đầu-cuối.
- Trái ngược với các kỹ thuật ML truyền thống với số lớp ẩn giới hạn, công nghệ DL không ngừng mở rộng và cải thiện khả năng của máy móc mà không gặp phải hiện tượng hội tụ nông, thường xảy ra ở các mô hình ML truyền thống.
- Quá trình huấn luyện mô hình DL thường bao gồm ba bước hoặc giai đoạn chính:
 - Huấn luyện từ đầu với lượng lớn dữ liệu (được gán nhãn/đánh dấu)
 - Học chuyên giao, trong đó mô hình đã được huấn luyện tiếp tục được tinh chỉnh

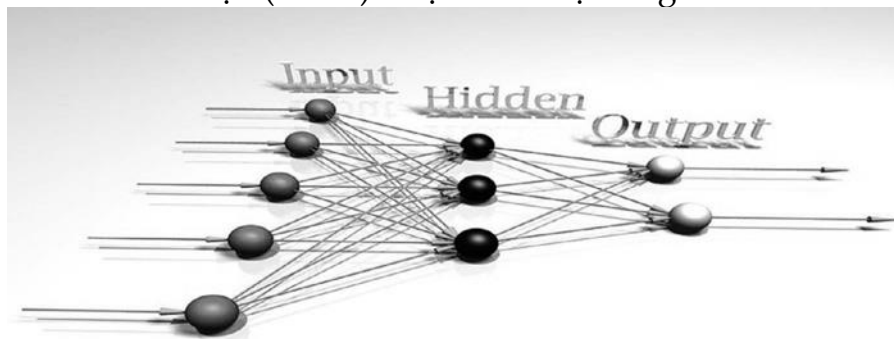
- Tự động trích xuất đặc trưng qua mạng nơ-ron tích chập

Các mô hình DL được sử dụng rộng rãi trong nhiều ứng dụng AI nhờ tính linh hoạt, khả năng thích nghi, hiệu quả chi phí, độ bền, độ chính xác, khả năng mở rộng, tính giải thích được và nhiều khía cạnh khác. Nó được coi là lĩnh vực AI tiên tiến và mạnh mẽ nhất, với tiềm năng phát triển cực kỳ sáng lạn trong tương lai.

Mạng Nơ-ron Nhân Tạo

Cụm từ "mạng nơ-ron nhân tạo" được lấy cảm hứng từ mạng nơ-ron tự nhiên, cấu trúc cốt lõi của các nơ-ron sinh học được kết nối với nhau để xử lý tín hiệu trong não người. Mạng lưới sinh học đó được gọi là mạng nơ-ron tự nhiên, và việc mô phỏng mạng nơ-ron tự nhiên đó cho máy móc được gọi là mạng nơ-ron nhân tạo (ANN).

Tương tự như các nơ-ron được kết nối trong não người, các nơ-ron nhân tạo, được gọi là "nút" (nodes), được kết nối với nhau thông qua các dây dẫn, được định nghĩa bởi trọng số (weightage), tương đương với khớp thần kinh (synapse) trong não tự nhiên. Dữ liệu đầu vào của AI tương tự như sợi nhánh (dendrites) trong não tự nhiên, và đầu ra của AI tương đương với sợi trục (axon) trong não người. Nút (node) được ví như nhân tế bào (cell nucleus). Kiến trúc của mạng nơ-ron nhân tạo (ANN) được hiển thị trong **Hình 1.6**.



Hình 1.6: Kiến trúc ANN. (Nguồn: Flickr.)

Trong mạng nơ-ron nhân tạo (ANN), dữ liệu đầu vào được đưa vào kiến trúc dưới dạng dữ liệu huấn luyện hoặc dữ liệu

kiểm tra. Dữ liệu huấn luyện được gán nhãn và nạp vào lớp đầu vào dưới dạng tập dữ liệu huấn luyện, trong khi dữ liệu kiểm tra được đưa vào lớp đầu vào dưới dạng tập dữ liệu kiểm tra. Một giá trị ngưỡng được thiết lập trên trọng số của dữ liệu đầu vào. Nếu trọng số của đầu vào cho một nút cao hơn giá trị ngưỡng, tín hiệu sẽ được xử lý tiếp đến nút tiếp theo; nếu không, tín hiệu sẽ bị mất tại nút đó. Trọng số được xác định dựa trên mức độ quan trọng và liên quan của thông tin.

Các lớp ẩn được sử dụng để tiếp tục xử lý và tinh chỉnh các tín hiệu đầu vào từ các nút của lớp ẩn trước đó, sau đó chuyển đến cấp độ tiếp theo của các lớp, có thể là lớp ẩn hoặc lớp đầu ra. Đầu ra của một nơ-ron sẽ thay đổi liên tục với việc học mở rộng của dữ liệu đầu vào và môi trường thông qua các ví dụ. Mạng nơ-ron càng được huấn luyện kỹ càng với dữ liệu, nó càng hoạt động mạnh mẽ và đáng tin cậy hơn.

Các đặc điểm và tính năng chính của ANN bao gồm:

- Một lớp nút đơn hoặc nhiều nút được lập trình để nhận dữ liệu đầu vào được gọi là lớp đầu vào của ANN.
- Các nút đầu vào kết nối với các lớp ẩn của mạng, bao gồm nhiều nút song song với các lớp khác, cũng được gọi là các lớp ẩn, trừ nút cuối cùng được kết nối với nút đầu ra.
- Tất cả các nút giữa lớp đầu vào và đầu ra được gọi là các lớp ẩn.
- Trọng số được tính bằng tổng của các đầu vào và một giá trị bias. Toàn bộ phép tính này được điều chỉnh bởi một công thức được gọi là *hàm truyền*. Đầu ra tính toán bởi hàm truyền được chuyển đến *hàm kích hoạt* làm đầu vào để tạo ra một đầu ra tương ứng với đầu vào nhận được từ hàm truyền trước đó.
- Các tính năng chính khác của ANN bao gồm: khả năng xử lý song song, khả năng phân phối bộ nhớ, khả năng làm việc với kiến thức không đầy đủ, lưu trữ dữ liệu

trên toàn bộ mạng, khả năng chịu lỗi hợp lý, và nhiều đặc điểm khác.

- Các loại hàm kích hoạt quan trọng nhất được sử dụng trong ANN bao gồm hàm nhị phân, hàm tuyến tính, hàm tan hyperbolic, và hàm sigmoidal.
- ANN được chia thành hai loại chính:
 - ANN truyền tiến (*Feed-forward ANN*)
 - ANN truyền ngược (*Feed-backward ANN*)
- ANN cũng có các chức năng điều khiển tự động để kiểm soát việc tính toán trong mạng nơ-ron.
- Việc lưu trữ dữ liệu trong ANN có khả năng chịu lỗi, nghĩa là nếu mạng nơ-ron bị xóa, dữ liệu lưu trữ chỉ bị ảnh hưởng rất ít.
- ANN có thể được nội suy hoặc ngoại suy dễ dàng từ dữ liệu lưu trữ trên mạng.
- Deep Learning (DL) được triển khai hoàn hảo thông qua ANN.
- ANN có thể là mạng nơ-ron dựa trên phần cứng hoặc phần mềm.

Thị giác máy tính Thị giác máy tính (CV) là một lĩnh vực rất quan trọng của AI. Lĩnh vực này xử lý các ảnh số, video và các hình ảnh khác để huấn luyện máy tính học từ những hình ảnh đó giống như não người. Máy tính được cung cấp các hình ảnh và video được ghi lại bởi camera, sau đó các mô hình học máy (ML) chủ yếu dựa trên học sâu (DL) hoặc mạng nơ-ron sẽ trích xuất thông tin có ý nghĩa từ những hình ảnh này, giúp máy tính hiểu nội dung của chúng và đưa ra các quyết định thông minh. Mục đích chính của việc sử dụng hình ảnh và video trong CV là giúp máy tính nhìn thế giới giống như con người, học từ những hình ảnh đó và đưa ra các quyết định phù hợp.

Dưới đây là các đặc điểm và tính năng phổ biến nhất của lĩnh vực phụ thuộc vào CV:

- CV cho phép máy tính nhìn vào các hình ảnh để học và phản ứng với thông tin có trong những hình ảnh đó.
- Thông thường, ở phía sau các hệ thống học máy dựa trên CV là các mô hình mạng nơ-ron học sâu (DL).
- Thị giác máy tính bắt đầu từ những năm 1950 nhưng tăng tốc vào những năm 1970 và phát triển mạnh mẽ trong vài thập kỷ gần đây.
- Độ chính xác của các ứng dụng dựa trên CV đã đạt gần 99% hoặc thậm chí hơn trong những năm 2020.
- Các mô hình học máy nhận thức hình ảnh giống như cách con người giải câu đố. Nó phân biệt và tách các đường viền, cạnh, và các phần của hình ảnh rồi đánh dấu chúng là các thành phần phụ để hiểu toàn bộ hình ảnh được đưa vào mô hình CV.
- Những ứng dụng quan trọng nhất của CV trong công nghệ AI hiện đại bao gồm nhận diện khuôn mặt, ứng dụng an ninh, phân tích hình ảnh và nhiều ứng dụng khác.
- Các mô hình DL phụ thuộc rất lớn vào các ứng dụng CV trong thế giới công nghệ hiện đại.
- Các lĩnh vực quan trọng sử dụng ứng dụng CV bao gồm chăm sóc sức khỏe, thương mại điện tử, quốc phòng, an ninh nội địa, sản xuất, chính phủ và nhiều lĩnh vực khác.

Dưới đây là các quy trình quan trọng nhất được sử dụng trong các chương trình CV:

- Phân đoạn hình ảnh và phân loại hình ảnh
- Phát hiện cạnh và phát hiện đối tượng
- Nhận diện khuôn mặt và so khớp đặc trưng
- Phát hiện mẫu và theo dõi đối tượng

Công nghệ quan trọng nhất làm nên lĩnh vực CV là mạng nơ-ron tích chập (CNN) và học sâu (DL).

Xử lý ngôn ngữ tự nhiên Chúng ta phần nào quen thuộc với NLP (Natural Language Processing), nhưng hãy tìm hiểu thêm về lĩnh vực này. NLP là một lĩnh vực con của AI, xử lý ngôn ngữ tự nhiên mà con người sử dụng dưới dạng văn bản và giọng nói. NLP cho phép máy tính hiểu, học và phản hồi ngôn ngữ tự nhiên mà con người sử dụng dưới dạng văn bản hoặc giọng nói. Nó nhận ngôn ngữ tự nhiên dưới dạng văn bản và giọng nói, sau đó thông qua các hoạt động hoặc quy trình khác nhau để hiểu và quyết định cách phản ứng với thông điệp cụ thể đó, dù là dưới dạng viết hay nói. Hiểu tất cả các thuộc tính gắn liền với giao tiếp nói và viết bằng ngôn ngữ tự nhiên là rất quan trọng trong trường hợp này.

Toàn bộ quy trình công nghệ này có nhiều điểm tương đồng với các quá trình liên quan đến giao tiếp ngôn ngữ tự nhiên của con người. Ví dụ, con người có tai và mắt để tiếp nhận giao tiếp ngôn ngữ tự nhiên dưới dạng giọng nói và văn bản. Máy tính tiếp nhận thông tin qua micro, bàn phím, thiết bị quang học, camera, v.v. Bộ não con người xử lý giao tiếp thông qua khả năng mà nó đã học từ kinh nghiệm và tạo ra phản hồi phù hợp dưới dạng văn bản viết hoặc giọng nói qua lời nói. Máy tính sử dụng các kỹ thuật AI để xử lý thông điệp đầu vào dưới dạng văn bản và giọng nói, hiểu ý nghĩa và cảm xúc trong giao tiếp và phản hồi phù hợp thông qua cùng mô hình AI bằng màn hình hoặc loa.

Dưới đây là các đặc điểm và tính năng quan trọng nhất của công nghệ NLP:

- Hiểu và phản hồi ngôn ngữ của con người dưới dạng văn bản và giọng nói được gọi là NLP.
- Sử dụng các cảm biến khác nhau để đầu vào như

camera, bàn phím, thiết bị quang học, micro, v.v., và sử dụng màn hình số và loa để đầu ra.

- Gồm hai giai đoạn chính: xử lý dữ liệu và xây dựng thuật toán.
- Xử lý dữ liệu được thực hiện thông qua nhiều quy trình như phân tách từ (tokenization), gắn thẻ từ loại (part-of-speech tagging), rút gọn từ (stemming), và chuẩn hóa từ (lemmatization).
- Các thuật toán được sử dụng trong mô hình xử lý ngôn ngữ được phân thành hai loại chính:
 - **Hệ thống dựa trên quy tắc:** Phương pháp này sử dụng các quy tắc đơn giản dựa trên ngôn ngữ học và cấu trúc ngữ pháp để xử lý ngôn ngữ ở cả dạng viết và nói.
 - **Hệ thống dựa trên học máy (ML):** Phương pháp này không định nghĩa quy tắc một cách rõ ràng, mà các mô hình ML do các nhà khoa học và nhà phát triển xây dựng sẽ tự phát triển quy tắc dựa trên việc hiểu và học từ dữ liệu đầu vào dưới dạng văn bản và giọng nói. Các kỹ thuật phổ biến nhất trong xây dựng hệ thống ML là học sâu (DL) và mạng nơ-ron.
- Về khía cạnh kỹ thuật, có hai phương pháp chính được sử dụng trong NLP:
 - **Phân tích cú pháp:** Sử dụng thứ tự và cách sắp xếp các từ phù hợp với cấu trúc ngữ pháp để suy ra ý nghĩa của câu.
 - **Phân tích ngữ nghĩa:** Áp dụng ý nghĩa và cách sử dụng của các từ trong câu để học ý nghĩa của cả câu.
- Các quy trình quan trọng nhất trong phân tích cú pháp của phương pháp NLP bao gồm phân tích cú pháp

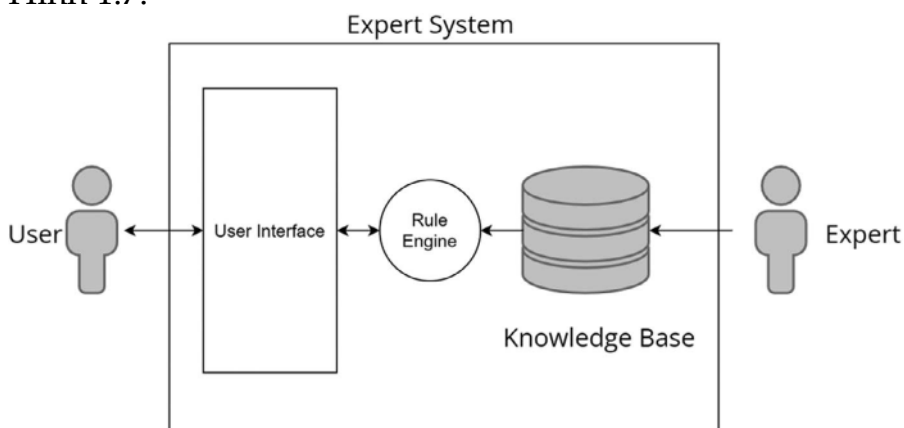
(parsing), phân đoạn từ, chia câu, phân đoạn hình thái, rút gọn từ và chuẩn hóa từ.

- Các kỹ thuật nổi bật nhất trong phân tích ngữ nghĩa của phương pháp NLP bao gồm:
 - Nhận dạng thực thể được đặt tên (Named Entity Recognition)
 - Phân giải ý nghĩa từ (Word Sense Disambiguation)
 - Tạo ngôn ngữ tự nhiên (Natural Language Generation)
- NLP đã xuất hiện trên thị trường dưới nhiều hình thức và đã tồn tại hơn 50 năm.
- Một số bộ công cụ quan trọng nhất được sử dụng để phát triển các mô hình NLP bao gồm:
 - Bộ công cụ xử lý ngôn ngữ tự nhiên (Natural Language Toolkit - NLTK)
 - Bộ công cụ NLP Gensim
 - Kiến trúc NLP Intel
- Các trường hợp sử dụng phổ biến nhất của công nghệ NLP bao gồm phân loại văn bản, trích xuất văn bản, dịch máy, phát hiện thư rác, tác nhân ảo và chatbot, phân tích cảm xúc trên mạng xã hội, tóm tắt văn bản, tạo ngôn ngữ tự nhiên thông qua các ứng dụng GPT, v.v.
- Các lĩnh vực hiện đại sử dụng rộng rãi công nghệ NLP bao gồm:
 - Phân tích phản hồi khách hàng trong ngành chăm sóc khách hàng
 - Tự động hóa dịch vụ khách hàng
 - Hệ thống dịch ngôn ngữ tự động
 - Lĩnh vực nghiên cứu và phân tích học thuật
 - Phân loại và phân tích hồ sơ y tế

- Hệ thống phát hiện đạo văn và lỗi chính tả
- Dự báo và giao dịch thị trường tự động
- Tự động hóa các nhiệm vụ pháp lý
- Phân tích cảm xúc chuyên sâu.

Hệ chuyên gia

Hệ chuyên gia (Expert Systems - ES) là một nhánh con của AI. Đây là một loại hệ thống AI chuyên biệt theo lĩnh vực, xử lý một quy trình cụ thể nhằm dự đoán kết quả dựa trên kinh nghiệm và kiến thức trước đó. Kiến thức hoặc kinh nghiệm mà hệ thống thu được được gọi là cơ sở tri thức (knowledge base). Cơ sở tri thức được phát triển bởi các chuyên gia trong ngành theo cách mà hệ chuyên gia có thể dễ dàng học từ đó bằng cách sử dụng các quy tắc học tập cụ thể được thiết kế trong một mô-đun gọi là động cơ suy luận (inference engine). Loại hệ thống này được phát triển bởi nhà khoa học máy tính Edward Feigenbaum, giáo sư khoa học máy tính tại Đại học Stanford, đồng thời là người sáng lập Phòng thí nghiệm Hệ tri thức của Stanford. Đây là hệ thống đầu tiên trong lịch sử. Hiện nay, có rất nhiều hệ chuyên gia có sẵn trên thị trường. Sơ đồ phác thảo của một hệ chuyên gia được thể hiện trong Hình 1.7.



HÌNH 1.7: Sơ đồ phác thảo của hệ chuyên gia.

Các đặc điểm và tính năng quan trọng nhất của hệ chuyên gia:

- Đây là một quy trình chuyên biệt theo lĩnh vực (một hệ thống thông minh) xử lý một hoạt động hoặc lĩnh vực cụ thể để dự đoán kết quả.
- Hệ thống này có ba thành phần chính:
 - **Cơ sở tri thức (Knowledge Base):** Đúng như tên gọi, đây là nơi lưu trữ dữ liệu hoặc thông tin mà các chuyên gia trong ngành lưu trữ để sử dụng cho việc dự đoán kết quả có căn cứ. Cơ sở tri thức này không ngừng mở rộng theo thời gian nhờ kinh nghiệm và việc bổ sung thêm thông tin. Các thông tin chi tiết về lĩnh vực liên quan được liên tục thêm vào bởi các chuyên gia. Cơ sở tri thức này chỉ có hai giao diện – một giao diện được các chuyên gia sử dụng và giao diện còn lại dành cho động cơ quy tắc, là một thành phần chính khác của hệ chuyên gia.
 - **Động cơ quy tắc (Rule Engine):** Thành phần này bao gồm hai mô-đun.
 - **Mô-đun suy luận (Inference Module):** Là một hệ thống dựa trên quy tắc để ánh xạ kiến thức từ cơ sở tri thức với các quy tắc được xác định trước nhằm đưa ra quyết định có cơ sở.
 - **Mô-đun giải thích (Explanation Module):** Được sử dụng để giải thích với người dùng cách mà một quyết định cụ thể được đưa ra, dựa trên dữ liệu từ cơ sở tri thức.
 - **Giao diện người dùng (User Interface):** Giao diện này được sử dụng để tương tác hoặc giao

tiếp với người dùng. Người dùng nhập các câu hỏi vào hệ thống để nhận được câu trả lời phù hợp. Hệ thống cung cấp phản hồi có căn cứ bằng cách ánh xạ dữ liệu câu hỏi với kiến thức đã có thông qua các quy tắc được xác định trước.

- Việc thu thập thông tin từ cơ sở tri thức được thực hiện qua hai phương pháp:
 - **Chuỗi suy luận ngược (Backward Chaining):**
Trong phương pháp này, các mảnh thông tin hiện có dưới nhiều dạng khác nhau được thu thập và sử dụng trong cây quyết định để tìm ra nguyên nhân tại sao một sự kiện cụ thể đã xảy ra trong quá khứ. Bằng phương pháp này, nguyên nhân gốc rễ của một hoạt động hoặc sự cố được tìm ra bằng cách lần ngược lại các sự kiện hoặc mảnh thông tin, bắt đầu từ những dữ liệu hiện có. Ví dụ về chuỗi suy luận ngược được sử dụng rộng rãi trong các hệ thống chăm sóc sức khỏe, nơi bệnh tật được truy vết bằng cách lần theo các sự kiện và tìm ra nguyên nhân gây bệnh.
 - **Chuỗi suy luận xuôi (Forward Chaining):**
Phương pháp này sử dụng các thông tin và kinh nghiệm quá khứ để dự đoán các sự kiện trong tương lai. Nói cách khác, phương pháp này dự đoán điều gì sẽ xảy ra trong tương lai dựa trên kiến thức hiện có và học hỏi trước đây. Ví dụ về phương pháp này là dự đoán hành vi thị trường trong giao dịch, dịch vụ tài chính và các lĩnh vực khác.
- Ngoài ba thành phần chính đã nêu, còn có một số phần bổ sung không được xem là các phần độc lập của hệ

chuyên gia. Các thành phần này bao gồm giao diện chuyên gia với cơ sở tri thức, các chuyên gia lĩnh vực cung cấp thông tin cho cơ sở tri thức, mô-đun thu thập tri thức, và kết nối nội bộ giữa các thành phần chính của hệ chuyên gia để thiết lập giao tiếp.

- Hệ chuyên gia thường được áp dụng trong các lĩnh vực cụ thể để dự đoán hoặc đưa ra các quyết định tương lai. Ví dụ ứng dụng bao gồm dự đoán hành vi thị trường, dự đoán tình trạng sức khỏe, truy ngược nguyên nhân từ các triệu chứng hiện có của bệnh nhân, khắc phục sự cố trong quy trình sản xuất, dự đoán thị trường tài chính, và nhiều hoạt động khác.
- Các ngành công nghiệp quan trọng mà hệ chuyên gia được sử dụng rộng rãi bao gồm dịch vụ tài chính, giao dịch, tiếp thị, viễn thông, chăm sóc sức khỏe, dịch vụ khách hàng, logistics và vận tải, v.v.
- Các tính năng chính khác của hệ chuyên gia bao gồm độ chính xác cao trong kết quả, suy luận hiệu quả và logic, đưa ra quyết định có cơ sở, cải thiện liên tục cơ sở tri thức để nâng cao chất lượng quyết định, tiết kiệm chi phí, tăng năng suất, v.v.

Học tăng cường (Reinforcement Learning - RL)

RL là một loại học máy (ML), nhưng nó khác biệt với các loại học khác ở nhiều khía cạnh. Do đó, nó được mô tả như một lĩnh vực con riêng biệt của AI. Trong loại ML này, không có sự hỗ trợ từ dữ liệu bên ngoài hay các câu trả lời được gán nhãn để học tập. RL dựa trên nguyên tắc thử-sai, trong đó tác nhân học (learning agent) sẽ cố gắng đoán kết quả trong một bước duy nhất, sau đó kiểm tra xem kết quả đó đúng, sai hay trung lập (không có gì). Một lần thử thành công sẽ được đánh dấu bằng kết quả tích cực hoặc được thưởng, trong khi lần thử không thành công sẽ nhận

phản hồi tiêu cực hoặc hình phạt dưới dạng điểm số. Quá trình này tiếp tục diễn ra, học từ các bước trước, từ đó hình thành hiểu biết dựa trên các kết quả và hệ thống phản hồi liên tục cập nhật đầu ra.

Các đặc điểm và tính năng quan trọng nhất của RL:

- Đây là một loại ML dựa trên mô hình thưởng và phạt cho kết quả hoặc dự đoán. Nếu dự đoán là mong muốn, nó sẽ được thưởng, và nếu không mong muốn, nó sẽ bị phạt.
- Mục tiêu chính của mô hình RL là tối đa hóa phần thưởng bằng cách thực hiện hành động phù hợp nhất để đạt được kết quả mong muốn nhất trong các điều kiện môi trường cụ thể.
- Tác nhân học (agent) không được huấn luyện rõ ràng thông qua các tập dữ liệu huấn luyện như các loại ML khác. Trong RL, tác nhân tương tác với môi trường xung quanh để quyết định hành động tiếp theo.
- Mô hình RL bao gồm bốn thành phần quan trọng:
 - **Chính sách (Policy):** Chính sách là chức năng hoặc hành vi cốt lõi của tác nhân trong các điều kiện nhất định để hành động. Chính sách này có thể được quản lý bởi một hàm đơn giản hoặc một tính toán toán học phức tạp. Hành động tiếp theo được xác định bằng cách tuân theo chính sách của mô hình.
 - **Phần thưởng (Reward):** Đây là giá trị tích cực hoặc tiêu cực phản hồi cho một hành động mà tác nhân thực hiện trong môi trường. Phần thưởng có thể là tích cực nếu kết quả là mong muốn hoặc tiêu cực nếu kết quả là không mong muốn. Mục đích chính của mô hình này là tối đa hóa phần thưởng.
 - **Hàm giá trị (Value Function):** Đây là trọng số tổng thể của các phần thưởng mà tác nhân

ML nhận được trong môi trường này. Hàm phần thưởng có thể thay đổi từ hành động này sang hành động khác trong môi trường ML.

- **Mô hình môi trường (Environment Model):** Môi trường là toàn bộ hệ sinh thái mà trong đó tác nhân thực hiện các hành động và nhận kết quả dưới dạng phần thưởng hoặc hình phạt. Môi trường giúp tác nhân quyết định hành động nào là có lợi và hành động nào sẽ gây bất lợi cho mục tiêu chính của việc học, đó là tối đa hóa phần thưởng.
- Tác nhân luôn học từ kết quả của các hành động trước đó, bao gồm cả phần thưởng và hình phạt.
- Quy trình làm việc của RL bao gồm hiểu môi trường, thực hiện một hành động thử nghiệm, nhận kết quả tích cực hoặc tiêu cực, học từ lỗi, và điều chỉnh bản thân trong môi trường thay đổi cho bước tiếp theo.
- Ứng dụng của RL bị giới hạn trong một số lĩnh vực như trò chơi, gợi ý cá nhân hóa, quản lý tài nguyên, quản lý khách hàng, robot, và một số lĩnh vực khác.
- Việc tối ưu hóa sử dụng tài nguyên trong nhiều nền tảng phần mềm quản lý tài nguyên doanh nghiệp (ERM) được thực hiện thông qua RL. Tương tự, trong lĩnh vực robot, nhiều nhiệm vụ không thể được dạy bằng các hình thức huấn luyện khác được thực hiện thông qua mô hình thử-sai của RL.
- Các lĩnh vực chính khác áp dụng RL bao gồm lý thuyết thông tin, lý thuyết điều khiển, nghiên cứu vận hành, lý thuyết trò chơi, trí tuệ bầy đàn, hệ thống đa tác nhân, tối ưu hóa dựa trên mô phỏng, thuật toán di truyền, thống kê, v.v.
- Các thuật toán phổ biến nhất được sử dụng trong RL bao gồm:
 - **Q-Learning:** Đây là một trong những thuật

toán được sử dụng phổ biến nhất trong các hệ thống RL. Trong phương pháp này, tác nhân không được cung cấp chính sách để học trước; tác nhân phải tự học bằng cách thực hiện các hành động.

- **State-action-reward-state-action (SARSA):** Trong thuật toán tuần tự này, tác nhân được cung cấp chính sách để bắt đầu. Chính sách của hệ thống RL là tập hợp các xác suất trong đó các khả năng nhất định được dự đoán dựa trên các hành động cụ thể trong môi trường.
- **Deep Q-Network:** Đây là một thuật toán rất quan trọng khác được sử dụng trong lĩnh vực RL hiện đại. Sự khác biệt cơ bản của thuật toán này là việc sử dụng mạng nơ-ron. Hành động mà tác nhân thực hiện trong môi trường này được dựa trên các mẫu ngẫu nhiên của các hành động có lợi đã thực hiện trong quá khứ.
- Các điểm chính và thuật ngữ trong RL:
 - **Đầu vào (Input):** Đây là trạng thái của mô hình RL nơi hoạt động học bắt đầu. Nó còn được gọi là trạng thái ban đầu của mô hình.
 - **Đầu ra (Output):** Đầu ra phụ thuộc vào nhiều tham số được xem xét trong mô hình như tích cực và tiêu cực, đúng và sai, mong muốn và không mong muốn, v.v. Đầu ra phụ thuộc vào loại giải pháp được chọn.
 - **Huấn luyện (Training):** Quá trình huấn luyện dựa trên chính sách đầu vào và kết quả của các hành động liệu chúng mang lại phần thưởng tích cực hay hình phạt tiêu cực. Huấn luyện tiếp tục với các lỗi hoặc kết quả của hành động trước đó để điều chỉnh hành động tiếp theo nhằm tối đa hóa kết quả tích cực.

Quá trình này được gọi là huấn luyện tác nhân trong RL.

- **Mở rộng (Expansion):** Khả năng học được tăng lên nhờ kết quả của các hành động của hệ thống, được gọi là sự mở rộng khả năng. Mô hình liên tục học và mở rộng khả năng của nó.
- RL có hai loại: tích cực và tiêu cực, được chia dựa trên giá trị kết quả của chúng.

Ngành Robot (Robotics)

Robot vừa là vừa không phải một lĩnh vực con của AI. Đây có thể là một tuyên bố lạ lùng, nhưng khi xét đến các robot hiện đại được hỗ trợ bởi AI, chúng là một phần của AI hoặc một lĩnh vực của AI cho phép các máy móc robot hành xử giống như con người thông qua khả năng tự học từ môi trường nhờ ML (học máy), mà không cần lập trình tường minh. Tuy nhiên, robot truyền thống, đã xuất hiện trên thị trường trong nhiều thập kỷ qua, không phải là một phần của AI.

Trước khi tìm hiểu sâu hơn về robot truyền thống và hiện đại, cần định nghĩa các thuật ngữ cơ bản về ngành robot.

Ngành kỹ thuật kết hợp giữa cơ khí, điện, điện tử, khoa học máy tính và lập trình phần mềm được gọi là robot học (robotics). Lĩnh vực này liên quan đến việc thiết kế các thiết bị cơ khí hoặc điện có thể tự động thực hiện các nhiệm vụ cụ thể mà không cần sự can thiệp trực tiếp của con người. Quá trình tự động hóa này được quản lý thông qua lập trình máy tính, được cài đặt dưới dạng hệ thống điều khiển.

Robot truyền thống

Các dạng robot truyền thống đã xuất hiện trên thị

trường hàng thập kỷ trong nhiều ngành công nghiệp như logistics, sản xuất, kho bãi, nông nghiệp, và các lĩnh vực khác, nơi chúng được sử dụng để thực hiện các nhiệm vụ khó khăn hoặc lặp đi lặp lại, hoặc cả hai. Robot truyền thống không có "bộ não" để suy nghĩ và đưa ra quyết định trong các môi trường thay đổi hoặc trong điều kiện không xác định. Chúng chỉ là các máy móc thực hiện nhiệm vụ cụ thể, có hình dạng khác nhau như bánh xe, cánh tay, chân, v.v.

Robot hiện đại và AI

Với sự tiến bộ của các công nghệ AI hiện đại như thị giác máy tính (CV), nhận dạng giọng nói, và nhận dạng ký tự quang học (OCR), hình thức, hình dáng và định nghĩa về robot hiện đại đã thay đổi đáng kể. Trong bối cảnh này, nhiều người nghĩ rằng robot và AI là các công nghệ hoặc lĩnh vực có thể thay thế lẫn nhau. Tuy nhiên, thực tế, chúng là hai lĩnh vực riêng biệt.

Nếu xét đến robot hiện đại, chúng không thể tồn tại nếu thiếu AI. Các robot hiện đại có thể dựa vào phần mềm, phần cứng, hoặc cả hai. Đôi khi, robot hiện đại được định nghĩa là "lĩnh vực của trí tuệ nhân tạo tập trung vào nghiên cứu và phát triển các công nghệ để hiện thực hóa khái niệm robot thông minh".

Đặc điểm nổi bật của công nghệ robot hiện đại:

- Robot học hiện đại là sự kết hợp giữa kỹ thuật cơ khí/điện tử và AI. Truyền thống, robot chỉ đơn thuần kết hợp giữa kỹ thuật và khoa học máy tính để lập trình hướng dẫn.
- Robot là sản phẩm của công nghệ robot, có thể dựa trên phần cứng, phần mềm, hoặc cả hai.
- AI cung cấp "bộ não" cho các máy móc robot để chúng có thể học tập và đưa ra quyết định giống như não

người.

- Tương lai của robot liên kết chặt chẽ với AI, vì các phát minh robot trong tương lai sẽ tiên tiến hơn so với các robot cơ khí hoặc điện tử truyền thống chỉ có chương trình sẵn để hướng dẫn thực hiện nhiệm vụ mà không cần đưa ra quyết định.
- Ví dụ về các robot hiện đại sử dụng sức mạnh của ML và AI bao gồm: iRobot Roomba, nhà bếp robot, Moxi, robot Iron Ox, xe tự hành trên sao Hỏa (Mars Rovers), v.v.
- Ngoài các robot phần cứng, còn có nhiều robot dựa trên phần mềm thông minh hơn trong nhiều lĩnh vực.
 - Ví dụ: ChatGPT (chatbot), các tác nhân trên website, chatbot hỗ trợ khách hàng, trợ lý ảo, trình thu thập dữ liệu (crawlers), và các công cụ khác.
 - Những robot này hoàn toàn phụ thuộc vào công nghệ AI và kỹ thuật phần mềm.
- Robot phần mềm không được coi là một phần của ngành robot theo cách truyền thống, nhưng chúng được gọi là robot trong lĩnh vực phần mềm hiện đại.
- Trong tương lai, sự khác biệt giữa AI và robot sẽ ngày càng giảm mạnh.

Các thành phần của robot hiện đại:

- Nguồn cung cấp năng lượng và bộ truyền động (actuators)
- Động cơ điện và cảm biến
- Dây cơ (muscle wires) và cơ khí khí nén (pneumatic air muscles)
- Động cơ piezo và động cơ siêu âm

Các ngành công nghiệp sử dụng robot hiện đại:

- Quốc phòng và an ninh

- Hàng không và sản xuất
- Giải trí và nông nghiệp
- Công nghiệp ô tô và máy bay không người lái
- Logistics và vận tải
- Ngành vũ trụ và tên lửa

Công nghệ AI phổ biến trong robot hiện đại:

- Thị giác máy tính (Computer Vision)
- Xử lý ngôn ngữ tự nhiên (NLP)
- Học máy (Machine Learning)
- Xử lý sự kiện phức tạp (CEP)
- Điện toán biên (Edge Computing)
- Điện toán cảm xúc (Effective Computing)
- Thực tế tăng cường (AR)/thực tế ảo (VR)
- Thực tế hỗn hợp (MR)

Các nhiệm vụ phổ biến của robot hiện đại dựa trên AI:

- Nhận diện khuôn mặt và nhận dạng đối tượng
- Ước tính vị trí và phát hiện chuyển động
- Nhận dạng ký tự quang học (OCR) và nhận dạng cử chỉ

9. Logic được sử dụng trong Trí tuệ nhân tạo và các loại hình logic

Như đã đề cập trước đó, lập luận là một trong những kỹ thuật quan trọng nhất đứng sau công nghệ AI. Lập luận được thực hiện bằng cách sử dụng logic, yếu tố chính để đưa ra bất kỳ lý do nào cho một hành động hoặc sự kiện. Do đó, logic được định nghĩa như sau: **“Sự chứng thực hoặc bằng chứng đằng sau bất kỳ lý do nào của hành động hoặc sự kiện.”**

Quá trình ra quyết định bị ảnh hưởng nặng nề bởi tình

huống hiện có, trong đó có nhiều tùy chọn đầu ra để lựa chọn. Cuối cùng, ra quyết định dựa trên việc chọn tùy chọn phù hợp nhất trong các tùy chọn có sẵn trong một tình huống cụ thể. Việc ra quyết định được thực hiện bởi IA (Tác nhân thông minh - Intelligent Agent) hoặc đơn giản bởi một tác nhân trong hệ thống AI. Nếu IA có khả năng hiểu logic, thì nó có thể dễ dàng đưa ra quyết định với sự hỗ trợ của quá trình lập luận.

Trong AI, có hai loại logic chính:

- **Logic suy diễn (Deductive Logic)**
- **Logic quy nạp (Inductive Logic)**

Logic suy diễn (Deductive Logic)

Sự cung cấp bằng chứng hoặc xác thực đầy đủ để đưa ra bất kỳ quyết định nào được gọi là logic suy diễn. Loại logic này đảm bảo rằng tất cả các khía cạnh của các điều kiện được phân tích và đo lường trong quá trình ra quyết định để không có khả năng kết luận mang tính ước lượng hoặc gần đúng.

Ví dụ: Sử dụng các hệ thống thông minh để đề xuất thuốc cho bệnh nhân từ xa hoặc tại chỗ.

Logic quy nạp (Inductive Logic)

Như tên gọi, logic quy nạp là bằng chứng hoặc sự xác thực của lý do trong đó quyết định được đưa ra bằng cách phân tích các thành phần từ dưới lên của môi trường để đạt được quyết định.

Trong logic này, IA phân tích tất cả các phần của môi trường từ góc độ thuộc tính cá nhân của các thành phần, sau đó tổng quát hóa kết quả và cuối cùng đưa ra quyết định dựa trên sự hiểu biết tổng quát đó.

Ví dụ: Lý luận đằng sau các hệ thống Xử lý ngôn ngữ tự

nhien (NLP).

10. Các Loại Logic Chính Khác trong Trí Tuệ Nhân Tạo

Nhánh chính của AI cho phép máy móc học hỏi từ môi trường hoặc từ tri thức chung có sẵn về các hệ sinh thái trong thế giới thực là học máy (ML). ML là quá trình mà máy móc được cung cấp dữ liệu biểu diễn tri thức, sau đó máy móc hiểu tri thức này và đưa ra quyết định dựa trên hệ thống logic, cuối cùng tạo ra quyết định dựa trên thông tin học được từ kinh nghiệm. Việc trình bày tri thức dưới dạng mà máy có thể hiểu được được gọi là logic.

Nói cách khác, điều này tương tự như ngôn ngữ máy được sử dụng trong lập trình máy tính để cung cấp các chỉ thị cho máy tính thực hiện các tác vụ hoặc hành động cụ thể. Các loại logic tổng quát nhất, trong đó có nhiều dạng logic khác nhau, đã được thảo luận trong phần trước. Các loại này bao gồm logic quy nạp và logic suy diễn.

Những dạng logic chính khác, dựa trên hai loại logic đã đề cập ở trên, được sử dụng trong ML như sau:

- Logic mệnh đề
- Logic bậc nhất
- Logic bậc hai
- Logic bậc ba
- Logic bậc cao
- Logic ngữ nghĩa

Tất cả các loại logic này sẽ được thảo luận riêng trong các phần tiếp theo.

Logic Mệnh Đề

Logic mệnh đề là loại logic cơ bản nhất trong AI để biểu

diễn dữ liệu. Nó sử dụng đại số Boolean để biểu diễn tri thức được lấy từ môi trường thế giới thực dưới dạng ngôn ngữ tự nhiên. Logic mệnh đề còn được gọi là logic bậc 0. Logic này dựa trên hai trạng thái, hoặc **ĐÚNG** hoặc **SAI**. Các trạng thái khác nhau được biểu diễn với sự trợ giúp của các toán tử được sử dụng trong đại số Boolean hoặc cách trình bày khác.

Logic mệnh đề có thể được chia thành hai loại chính:

- **Logic Mệnh Đề Nguyên Tử** – Dựa trên tri thức được biểu diễn trong câu hoặc công thức đơn giản, dễ biểu diễn.
- **Logic Mệnh Đề Phức Tạp** – Dựa trên tri thức được biểu diễn trong câu hoặc công thức phức tạp, khó biểu diễn hơn.

Logic Bậc Nhất

Logic bậc nhất là sự mở rộng của logic bậc 0 hoặc có thể được xem là logic mệnh đề phức tạp, là phần thứ hai của logic mệnh đề. Trong loại logic này, các câu hoặc công thức phức tạp hơn được phân tách dựa trên các biến khác nhau để trình bày đầy đủ ý nghĩa của một mệnh đề.

Logic Bậc Hai

Trên thực tế, logic bậc hai là sự mở rộng của logic bậc nhất, trong đó bổ sung các biến được xem xét từ yếu tố cơ bản, được coi là thuộc logic bậc 0 hoặc logic mệnh đề đơn giản. Một đặc điểm hoặc thuộc tính bổ sung của đối tượng trong tri thức môi trường được xem xét trong logic bậc hai để biểu diễn tri thức cho máy móc một cách chính xác hơn, giúp máy móc hiểu đúng.

Logic Bậc Ba

Việc biểu diễn tri thức dưới dạng các câu tự nhiên hoặc

công thức toán học được thực hiện với các ký hiệu logic để trình bày tri thức ở định dạng mà máy có thể hiểu được. Logic bậc 0 xử lý các ký hiệu nguyên tử để biểu diễn toàn bộ câu mà không có biến bổ sung. Logic bậc nhất dịch các biến khác nhau trong câu, và logic bậc hai xem xét các đặc điểm của một đối tượng trong câu hoặc công thức. Logic bậc ba xem xét thêm các thuộc tính bổ sung của câu để chuyển tri thức sang định dạng dễ hiểu hơn cho máy trong môi trường học máy.

Logic Bậc Cao Việc thêm siêu thuộc tính (super-predicate) vào một công thức hoặc câu trong quá trình biểu diễn tri thức theo định dạng logic được thực hiện thông qua logic bậc cao. Điều này có thể tăng cấp độ logic khi số lượng thuộc tính và điều kiện như hằng số và các điều kiện cố định trong việc trình bày thông tin thế giới thực tăng lên.

Logic Ngữ Nghĩa

Logic ngữ nghĩa được sử dụng rộng rãi để mô tả tri thức trong một hệ thống liên quan, nơi các thuộc tính khác nhau được liên kết với các hệ thống hoặc quan hệ khác trong hệ thống chứng minh logic ngữ nghĩa. Đây là một loại logic phức tạp và tiên tiến hơn để biểu diễn các kịch bản tri thức phức tạp, giúp máy móc dễ dàng hiểu được.

Ứng Dụng của Logic trong Trí Tuệ Nhân Tạo

AI là một lĩnh vực đang phát triển mạnh mẽ với nhiều ngành đang nổi lên. Logic là thành phần cốt lõi và nền tảng của công nghệ AI. Vì vậy, logic mở ra cơ hội cho rất nhiều ứng dụng trong các lĩnh vực và ngành nghề khác nhau. Những ứng dụng quan trọng nhất của logic trong hệ thống AI bao gồm:

- **Lý luận Toán học Tự động** – Phát triển các hệ thống tự động để thực hiện các tác vụ cụ thể dựa trên lý luận

toán học.

- **Khám phá Khoa học Tự động** – Việc khám phá các khái niệm khoa học hoàn toàn dựa trên lý luận và logic. Đây là một trong những ứng dụng quan trọng nhất của logic trong AI.
- **Lập trình Quy nạp** – Một lĩnh vực ứng dụng khác của logic là lập trình quy nạp, trong đó người học cố gắng hiểu các chương trình máy tính, cách chúng được phát triển và logic được sử dụng để giúp máy móc hiểu tri thức được biểu diễn dưới các định dạng khác nhau.
- **Xây dựng Hệ Thống Logic Học Máy** – Phát triển các hệ thống logic dựa trên các công nghệ hiện đại như học máy (ML) để mô phỏng hoạt động của bộ não con người, học hỏi và đưa ra quyết định từ các điều kiện môi trường mà không cần sự can thiệp rõ ràng từ con người.
- **Xác minh Hệ thống Máy tính** – Việc xác minh hoặc kiểm định hệ thống máy tính cũng dựa trên việc sử dụng logic ở mức độ rộng lớn. Đây là một ứng dụng chính khác của logic trong việc xác minh và kiểm định các hệ thống máy tính hiện đại.

11. Các Tiếp Cận trong Trí Tuệ Nhân Tạo

AI đang trở thành lĩnh vực quan trọng nhất trong mọi loại công nghệ, ngành công nghiệp, và lĩnh vực của cuộc sống hàng ngày. Mỗi người liên quan đến bất kỳ ngành nghề hoặc lĩnh vực kinh doanh nào, đặc biệt là những người đam mê công nghệ, đều muốn tìm hiểu về công nghệ hiện đại này. Để học các khái niệm và khía cạnh công nghệ của AI, cần phải thực hiện một số nghiên cứu. Trong quá trình nghiên cứu AI, bạn có thể gặp hai cách tiếp cận chính:

- **Tiếp cận Biểu tượng**
- **Tiếp cận Kết nối**

Tiếp Cận Biểu Tượng

Tiếp cận biểu tượng trong nghiên cứu AI còn được gọi là cách tiếp cận AI cổ điển, cách tiếp cận AI truyền thống, hoặc cách tiếp cận AI dựa trên quy tắc. Đây là hình thức tiếp cận truyền thống để nghiên cứu AI. Trong cách tiếp cận biểu tượng, các quy tắc của con người dựa trên hành vi và tri thức được mã hóa vào chương trình máy tính để giúp máy móc học cách xử lý các môi trường mà chúng gặp phải. Các biểu tượng, ký tự, và các cách biểu diễn khác nhau được sử dụng để mô tả tri thức hoặc điều kiện cho máy tính.

Cách tiếp cận này đã được sử dụng rộng rãi trong quá khứ để đạt được tự động hóa dựa trên quy tắc. Tuy nhiên, nó đang dần suy giảm và một cách tiếp cận nghiên cứu mới đang nổi lên, được gọi là cách tiếp cận kết nối.

Tiếp Cận Kết Nối Đây là cách tiếp cận nghiên cứu tiên tiến trong lĩnh vực AI. Nó phổ biến hơn trong các dự án nghiên cứu hiện đại về AI và được sử dụng rộng rãi trong các lĩnh vực học máy (ML) và AI ngày nay. Trong mô hình nghiên cứu này, cách tiếp cận mô phỏng mô hình của bộ não con người để thực hiện các hoạt động thông minh bởi máy móc. Các khái niệm về bộ não, tế bào thần kinh và mạng nơ-ron được áp dụng trong cách tiếp cận nghiên cứu hiện đại này.

Cách tiếp cận biểu tượng trước đây sử dụng các chỉ thị được mã hóa thủ công để máy tính thực hiện các nhiệm vụ tự động. Nhưng cách tiếp cận kết nối cho phép máy

tính tự học từ môi trường được cung cấp và đưa ra các quyết định của riêng chúng trong các tình huống phức tạp và đầy thách thức một cách hoàn hảo.



CÂU HỎI

1. Trí tuệ nhân tạo (AI) là gì, và nó đóng góp như thế nào vào các lĩnh vực khác nhau?
2. Các mục tiêu của AI là gì, và chúng thúc đẩy sự phát triển công nghệ và giải quyết vấn đề ra sao?
3. Các loại AI là gì, và chúng khác nhau như thế nào về ứng dụng và khả năng?
4. Bạn có thể giải thích cấu trúc của một hệ thống AI và các thành phần của nó không?
5. Các lĩnh vực khác nhau của AI là gì, và chúng ảnh hưởng như thế nào đến các ngành công nghiệp và các lĩnh vực nghiên cứu?
6. Logic được sử dụng trong AI như thế nào, và nó đóng vai trò gì trong quá trình ra quyết định trong các hệ thống AI?