Một hệ thống thông minh là gì?

Martin Molina

Khoa Trí tuệ nhân tạo, Đại học Politécnica de Madrid

Tháng Hai 2022

Trừu tượng

Khái niệm hệ thống thông minh đã xuất hiện trong công nghệ thông tin như một loại hệ thống bắt nguồn từ các ứng dụng thành công của trí tuệ nhân tạo. Mục tiêu của bài báo này là đưa ra một mô tả chung về một hệ thống thông minh, tích hợp các phương pháp tiếp cận trước đó và tính đến những tiến bộ gần đây trong trí tuệ nhân tạo. Bài báo mô tả một hệ thống thông minh một cách chung chung, xác định các thuộc tính chính và các thành phần chức năng của nó, đồng thời trình bày một số danh mục phổ biến. Mô tả được trình bày tuân theo một cách tiếp cận thực tế được sử dụng bởi các kỹ sư hệ thống. Tính tổng quát và cách sử dụng của nó được minh họa bằng các ví dụ hệ thống trong thế giới thực và liên quan đến các phương pháp trí tuệ nhân tạo.

# Giới thiệu

Nhân loại đã đạt được những tiến bộ đáng kể thông qua việc phát triển các công cụ ngày càng mạnh mẽ và tinh vi. Trong thời đại của cuộc cách mạng công nghiệp, một số lượng lớn các công cụ đã được chế tạo như những cỗ máy tự động hóa nhiệm vụ đòi hỏi nỗ lực thể chất. Trong thời đại kỹ thuật số, các công cụ dựa trên máy tính đang được tạo ra để tự động hóa các nhiệm vụ đòi hỏi nỗ lực tinh thần. Khả năng của các công cụ này đã được tăng dần để thực hiện các tác vụ đòi hỏi ngày càng thông minh hơnence( bằng tiếng Anh). Sự phát triển này đã tạo ra một loại công cụ mà chúng ta gọi là hệ thống thông minh.

arXiv:2009.09083v2 [cs.CY] 12 tháng 2 năm 2022

Hệ thống thông minh giúp chúng tôi thực hiện các nhiệm vụ chuyên môn trong các lĩnh vực chuyên môn như chẩn đoán y tế (ví dụ: nhận dạng khối u trên hình ảnh X-quang) hoặc quản lý sân bay (ví dụ: tạo ra sự phân công mới của cổng sân bay khi có sự cố). Họ cũng có thể thực hiện cho chúng tôi những công việc tẻ nhạt (ví dụ: lái xe tự hành hoặc dọn dẹp nhà cửa) hoặc các nhiệm vụ nguy hiểm như khám phá các khu vực không xác định (ví dụ: khám phá dưới nướcn).

Sự phát triển của một loại hệ thống như vậy hiện là một ngành kỹ thuật của công nghệ thông tin đòi hỏi các phương pháp và công cụ hiệu quả. Đặc tính chính xác của một hệ thống thông minh là không tầm thường bởi vì nó dựa trên các thuật ngữ liên quan đến nhận thức, một lĩnh vực không được hiểu đầy đủ và thừa nhận các cách hiểu khác nhau. Một số thuật ngữ được sử dụng thậm chí có thể thay đổi với đề xuất của các mô hình tính toán mới về trí thông minh và những phát hiện khoa học mới liên quan đến sự hiểu biết của chúng ta về tâm trí.

Mục đích chính của bài báo này là trình bày một đặc tính của một hệ thống thông minh tích hợp và cập nhật các quan niệm trước đây về loại hệ thống đó. Nó tuân theo một cách tiếp cận thực dụng để trở nên hữu ích trong bối cảnh kỹ thuật để giúp các công cụ hệ thốnghình thành, phân tích và xây dựng các hệ thống thông minh bằng cách sử dụng các khái niệm và thuật ngữ thường được chấp nhận trong trí tuệ nhân tạo. Do đó, mô tả được trình bày tập trung vào các đặc điểm chức năng liên quan đến trí thông minh mà máy móc có thể exhibit với trạng thái hiện tại của công nghệ.

Phần còn lại của bài báo trình bày các nội dung sau. Phần 2 trình bày một định nghĩa về hệ thống thông minh với mô tả các thuộc tính của nó và các thành phần chức năng chính của nó. Phần 3 mô tả các khía cạnh trong quá trình phát triển của một hệ thống thông minh như một loại phương pháp trí tuệ nhân tạo được sử dụng để phát triển các thành phần khác nhau của các hệ thống thông minh. Phần 4 mô tả các lớp hệ thống thông minh sử dụng đặc tính được trình bày, bao gồm cả những thành tựu gần đây và cách tiếp cận cổ điển của công nghệ này .

|  |
| --- |
| Định nghĩa 1: Tác nhân thông minh là các hệ thống tính toán sống trong một số môi trường động phức tạp, cảm nhận và hành động tự chủ trong môi trường này và bằng cách đó, hãy nhận ra một tập hợp các mục tiêu hoặc nhiệm vụ mà chúng được thiết kế (Maes 1995) |
| Định nghĩa 2: Các tác nhân thông minh liên tục thực hiện ba chức năng: nhận thức về các điều kiện động trong môi trường; hành động ảnh hưởng đến các điều kiện trong môi trường; và lý doing để giải thích nhận thức, giải quyết vấn đề, rút ra suy luận và xác định hành động (Hayes-Roth 1995) |
| Định nghĩa 3: Các tác nhân thông minh hoạt động tự chủ, nhận thức môi trường của chúng, tồn tại trong một khoảng thời gian dài, thích ứng với sự thay đổi, tạo ra và theo đuổi các mục tiêu (Russell và Norvig 2014) |
| Định nghĩa 4: Tác nhân thông minh là các hệ thống phần cứng tự trị hoặc bán tự trịo r thực hiện các tác vụ trong môi trường phức tạp, thay đổi linh hoạt. Các tác nhân giao tiếp với môi trường của họ và ảnh hưởng đến những thay đổi trong môi trường của họ bằng cách thực hiện các hành động (Muller 1996) |

Bảng 1: Mẫu định nghĩa tác nhân thông minh.

# Định nghĩa hệ thống thông minh

Các hệ thống thông minh đã được đặc trưng trong tài liệu về AI sử dụng khái niệm *tác nhân thông minh* (Wooldridge và Jennings 1995) (Franklin và Graesser 1996) (Russell và Norvig 2014). Khái niệm này nhấn mạnh sựcao quý để hành động[[1]](#footnote-1) trên một môi trường. Bảng 1 cho thấy một mẫu các định nghĩa của hệ thống thông minh được coi là một tác nhân. Những định nghĩa này nhấn mạnh rằng hệ thống hoạt động trong một môi trường, có một tập hợp các khả năng (nhận thức, học tập, v.v.) và đưa ra quyết định về cách hành động.

Hệ *thống* từ, thay vì tác nhân, được sử dụng bởi một số tác giả (Meystel và Albus 2002) và trong các lĩnh vực học thuật của công nghệ thông tin như tên của các khóa học đại học hoặc tạp chí học thuật (ví dụ: Hệ thống thông minh IEEE và Tạp chí quốc tế về hệ thống thông minh và robotic). Trong trường hợp này, hệ thống từ nhấn mạnh sự hiện diện của nhiều thành phần phải được kết hợp đầy đủ để tạo ra trí thông minh. Biết cách thực hiện sự kết hợp này một cách hiệu quả là một trong những khía cạnh quan trọng trong sự phát triển của loại hệ thống như vậy.

Đặc tính của một hệ thống thông minh được sử dụng trong bài báo này tích hợp các phần của các định nghĩa dựa trên tác nhân trước đó. Ví dụ, việc xác định một môi trường riêng biệt nơi tác nhân hoạt động là rất quan trọng để cung cấp một ngữ cảnh operationa l đầy đủ ở một mức độ trừu tượng nhất định. Trong định nghĩa của chúng tôi, chúng tôi cũng phân biệt với môi trường các tác nhân khác có thể tương tác với hệ thống (ví dụ: người dùng là con người hoặc các hệ thống nhân tạo khác) để xem xét yếu tố xã hội của hệ thống intelligent.

Bài báo này cũng sử dụng một tập hợp các khả năng thường được liệt kê bởi các định nghĩa dựa trên tác nhân (ví dụ: nhận thức, lý luận, học tập, v.v.). Tuy nhiên, chúng tôi tách các khả năng này thành hai phần. Một mặt, có bốn phẩm chất abi nhận thức chính: nhận thức, kiểm soát hành động, lý luận cân nhắc và sử dụng ngôn ngữ (để tương tác với người khác). Mặt khác, chúng tôi phân biệt khả năng thích ứng với những thay đổi trên thế giới thông qua học tập. Năng lực này được áp dụng cho mỗi khả năng nhận thức chính .

Cuối cùng, đặc tính của chúng tôi về một hệ thống thông minh cũng tính đến việc hệ thống hoạt động theo hai nguyên tắc hành vi: hành động hợp lý và tuân theo các chuẩn mực xã hội. Nguyên tắc đầu tiên cho rằng hành vi của hệ thống phù hợp với các mục tiêu cá nhân cần đạt được (xem các phần sau). Nguyên tắc thứ hai cho rằng hệ thống có hành vi đạo đức nhất định để tuân theo các chuẩn mực được thiết lập bởi nhóm xã hội mà nó hợp tác.

Theo đó, định nghĩa mà chúng tôi tuân theo trong bài báo này một sơ đồ cơ bản để cấu trúc đặc tính hơn nữa của một hệ thống thông minh là như sau:

Định nghĩa. Một hệ thống thông minh (1) hoạt động trong môi trường với các tác nhân khác, (2) possesses các khả năng nhận thức như nhận thức, kiểm soát hành động, lý luận cân nhắc hoặc sử dụng ngôn ngữ, (3) tuân theo các nguyên tắc hành vi dựa trên tính hợp lý và các chuẩn mực xã hội, và (4) có khả năng thích ứng thông qua học tập.

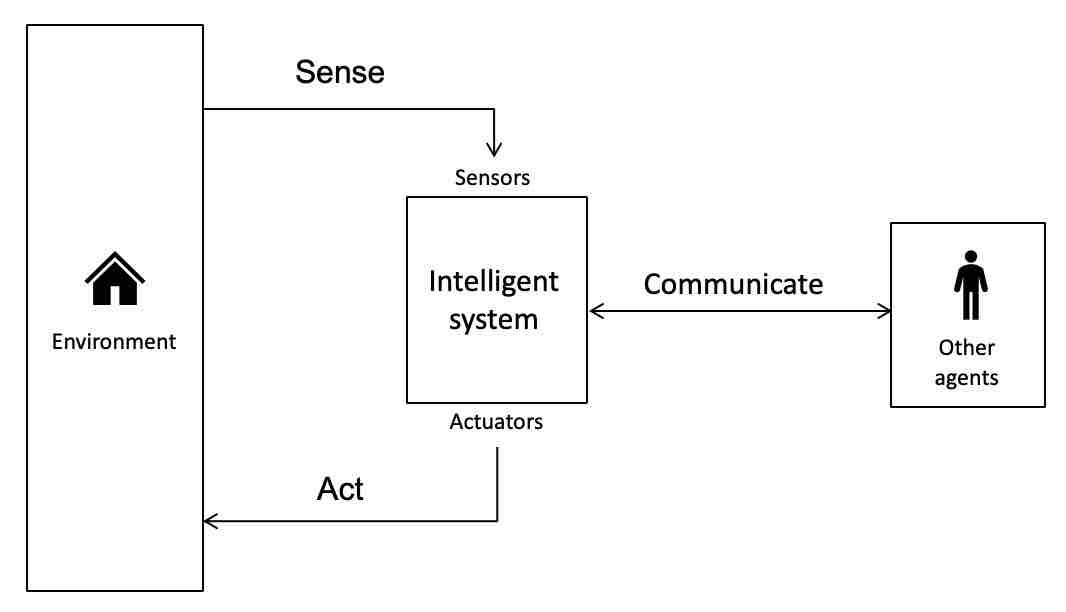
Điều quan trọng cần lưu ýlà t định nghĩa của anh ta không nhằm mục đích là một đặc tính cứng nhắc để xác định xem một hệ thống có thông minh hay không. Thay vào đó, định nghĩa bao gồm các đặc điểm thông thường liên quan đến trí thông minh có thể có hoặc không có trong phân tích hoặc phát triển các hệ thống thông minh cụ thể.

Ví dụ, một trường hợp đơn giản của hệ thống thông minh có thể là một tác nhân phản xạ đơn giản, vì nó được gọi bởi Russel và Norvig (Russell và Norvig 2014). Loại hệ thống này sở hữu khả năng nhận thức và kiểm soát hành động và nókhông có các đặc điểm khác như học tập hoặc lý luận cân nhắc. Một robot tự động cộng tác là một ví dụ phức tạp hơn về hệ thống thông minh. Trong trường hợp này, robot tương tác với người dùng và với môi trường mà nó hoạt động. Robot có khả năng nhận thức về nhận thức, kiểm soát hành động, lý luận cân nhắc (ví dụ: lập kế hoạch chuyển động) và sử dụng ngôn ngữ (để giao tiếp với người dùng). Cũng có thể giả định rằng robot có khả năng thích ứng với sự trợ giúp của cáckỹ thuật học tập machin e.

Các phần sau đây mô tả chi tiết hơn bốn thuộc tính trong định nghĩa của chúng tôi về hệ thống thông minh, trình bày các khái niệm và danh mục bổ sung trong đó các yếu tố của một hệ thống có thể được phân loại.

## Tài sản 1: Làm việc trong môi trường với các đại lý khác

Như đã đề cập trong phần trước, một trong những thuộc tính của hệ thống thông minh là nó hoạt động trong môi trường cùng với các tác nhân khác mà hệ thống có thể giao tiếp (ví dụ: người dùng là con người, tác nhân dựa trên máy tính nhân tạo, v.v.). System quan sát các đặc điểm từ môi trường thông qua các cảm biến và thực hiện các hành động sử dụng bộ truyền động (xem Hình 1). Việc sử dụng các cảm biến và bộ truyền động (thực hoặc ảo) ngăn cách phần thân của hệ thống thông minh với phần còn lại của môi trường. Nhân vật nàyesttic được gọi là *hiện thân*.



Hình 1: Một hệ thống thông minh tương tác với môi trường và với các tác nhân khác.

Ví dụ. Bộ điều nhiệt là một ví dụ đơn giản minh họa khái niệm về loại tương tác này. Mục tiêu của bộ điều nhiệt là điều chỉnh nhiệt độ trong môi trường (ví dụ: phòng trong nhà). Bộ điều chỉnh nhiệt sử dụng nhiệt kế để cảm nhận nhiệt độ của căn phòng và kích hoạt bằng cách khởi động hoặc dừng lò sưởi. Bộ điều chỉnh nhiệt cũng giao tiếp với người dùng là người bắt đầu và dừng chỉsố nhiệt điện trở, hoặc thiết lập nhiệt độ mong muốn.

### Môi trường

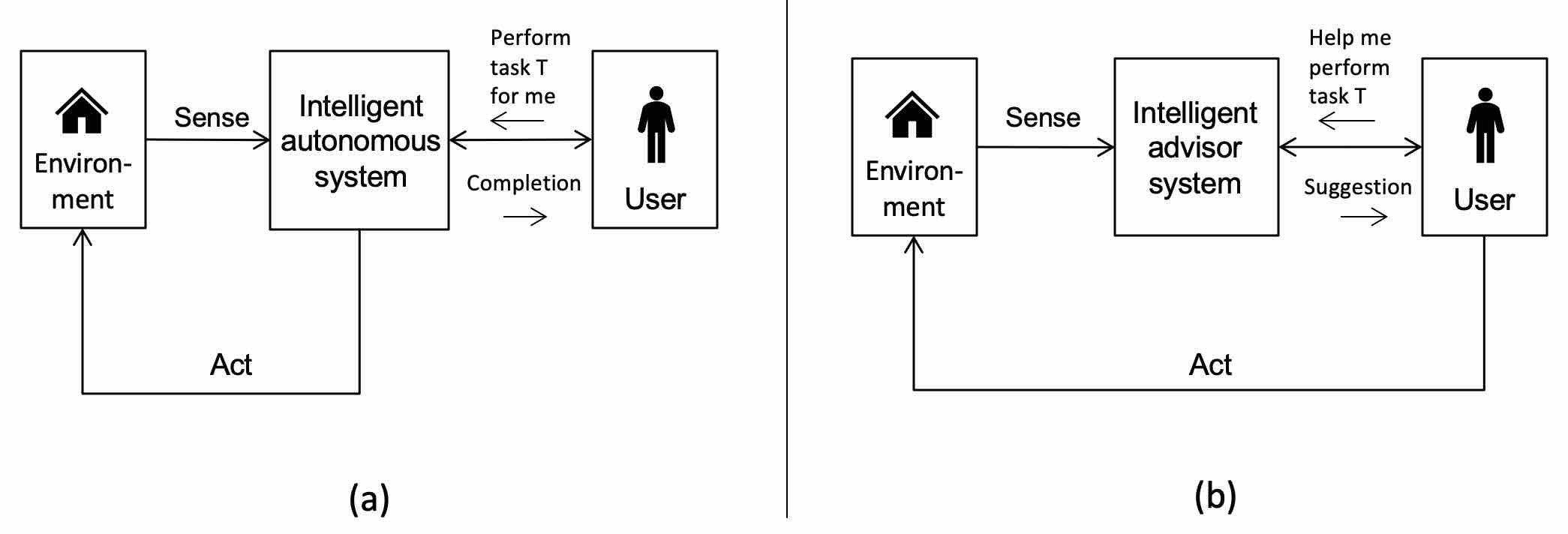
Thông thường, độ phức tạp của môi trường cao hơn đáng kể so với khả năng xử lý thông tin của hệ thống. Số lượng các thành phần của môi trường, tương tác của chúng hoặchiệu ứng giải phóng kiến của các hành động tiềm năng trong môi trường thường không thể được thể hiện hoàn toàn trong bộ nhớ của hệ thống thông minh. Ví dụ, trong trò chơi cờ vua, không thể dự đoán chính xác hiệu quả của tất cả các chuyển động tiềm năng, bởi vì số lượng kết hợp quá cao. Mức độ phức tạp của môi trường đối với tác nhân có thể được mô tả với các tính năng sau (Wooldridge 2009; Russell và Norvig 2014):

* *Tĩnh* (hoặc *động*), môi trường không thay đổi (hoặc thay đổi) trong khi một tác nhân đang cân nhắc,
* *Rời rạc* (hoặc *liên tục*), trạng thái của môi trường, thời gian, nhận thức hoặc hành động là rời rạc (hoặc liên tục),
* *Hoàn toàn có thể quan sát được* (hoặc *có thể quan sát một phần*), các cảm biến phát hiện (hoặc không phát hiện) tất cả các khía cạnh được relevant để lựa chọn hành động,
* *Xác định* (hoặc *ngẫu nhiên*), trạng thái tiếp theo của môi trường là (hoặc không) hoàn toàn được xác định bởi trạng thái hiện tại và hành động,
* *Tình tiết* (hoặc *tuần tự*), các hành động không ảnh hưởng (hoặc chúng ảnh hưởng đến) các hành động trong tương lai,
* *Đã biết* (hoặc *chưa biết*), kết quả cho tất cả các hành động được đại lý biết trước (hoặc chúng không được biết trước).

Ví dụ. Môi trường của một người chơi cờ là tĩnh, rời rạc, hoàn toàn có thể quan sát được, xác định, tuần tự và được biết đến. Trong trường hợp của một chiếc xe tự lái, môi trường là liên tục, quan sát một phần, ngẫu nhiên, tuần tự và được biết đến.

### Các đại lý khác

Hệ thống intelligent có thể tương tác với các tác nhân khác như một phần của tổ chức. Một tổ chức như vậy có thể được xem như một hệ thống đa tác nhân, trong đó các tác nhân riêng lẻ tương tác bằng cách sử dụng các cơ chế phối hợp xã hội để hợp tác nhằm đạt được các mục tiêu chung hoặc để cạnh tranh cho cácnguồn lực bắt chước.

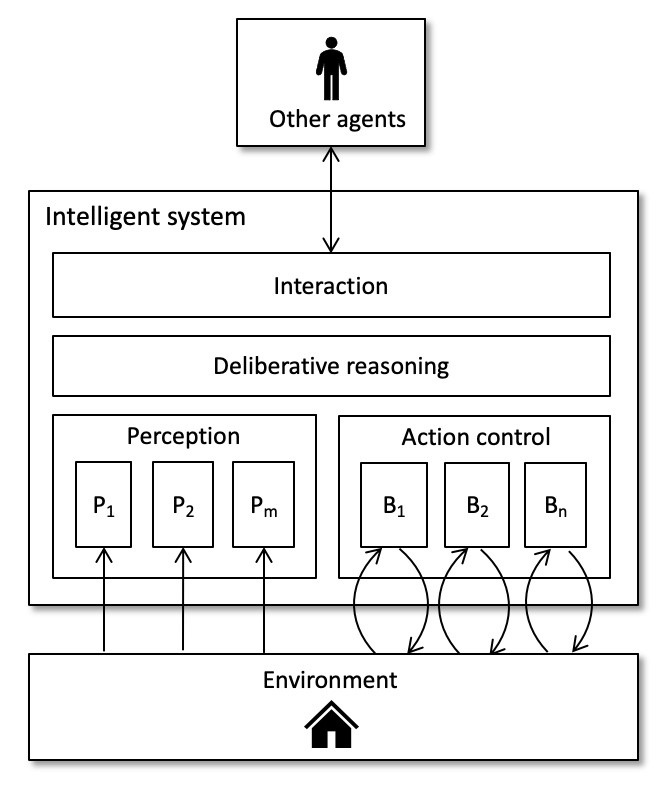


Hình 2: So sánh giữa (a) hệ thống tự trị và (b) hệ thống cố vấn.

Một hệ thống thông minh nhân tạo tương tác ít nhất với người dùng là con người đóng vai trò là chủ sở hữu của hệ thống. Người dùng bắt đầu và dừng thực thi hệ thống thông minh và có thể điều chỉnh các thông số cụ thể để sửa đổi hành vi của hệ thống (ví dụ: chủ sở hữu của người chơi cờ vua có thể điều chỉnh cấp độ của trò chơi). Tùy thuộc vào người đang tác động lên môi trường (hệ thống hoặc người dùng), một hệ thống nhân tạo có thể đóng hai vai trò khác nhau trong quá trình tương tác với người dùng (Hình 2):

* *Tự trị*. Một hệ thống tự trị *hoạt động trong môi trường để giúp người dùng*. Điều này có nghĩa là người dùng ủy quyền cho hệ thống một nhiệm vụ sẽ được thực hiện và hệ thống đưa ra quyết định của riêng mình on làm thế nào để hành động trong môi trường để tự động thực hiện nhiệm vụ đó. Một hệ thống tự trị cũng có thể từ chối các tác vụ được yêu cầu từ người dùng theo một số lý do nhất định, dựa trên tình hình hiện tại của môi trường hoặc các mục tiêu riêng của nó (ví dụ: mục tiêu an toàn hoặccác tiêu chuẩn xã hội). Vì mục đích này, hệ thống có thể xác minh tính đúng đắn và khả thi của một tác vụ được yêu cầu trước khi nó được thực hiện và giải thích cho người dùng những lý do biện minh cho lý do tại sao một tác vụ được yêu cầu bị từ chối.
* *Cố vấn*. Một hệ thống cố vấn *giúp người dùng hành động trong môi trường*. Trong trường hợp này, người dùng là người đưa ra quyết định về những hành động cần làm trong môi trường. Vai trò của hệ thống là đưa ra lời khuyên bằng cách cung cấp thông tin hữu ích để tạo điều kiện thuận lợi cho các quyết định đó cho người dùng. Hệ thống có thể anh talp người dùng trong các nhiệm vụ quy định như lập kế hoạch, lập lịch trình hoặc phân công tài nguyên, nhưng cũng trong các nhiệm vụ phân tích như phát hiện vấn đề, chẩn đoán vấn đề và chiếu thời gian.

Trong quá trình tương tác, một hệ thống thông minh có thể *chủ động* thay vì thụ động[[2]](#footnote-2). Điều này có nghĩa là hệ thống không phải đợi cho đến khi người dùng yêu cầu một nhiệm vụ, nhưng nó chủ động thực hiện một nhiệm vụ dựa trên các mục tiêu của chính nó và những gì nó cảm nhận được từ môi trường.



Hình 3: Khả năng nhận thức chính của một hệ thống thông minh.

## Tài sản 2: Có khả năng nhận thức chính

Trong bài báo này, chúng tôi phân biệt bốn khả năng nhận thức chính thường có trong các hệ thống thông minh:

* *Nhận thức*. Nhận thức có thể không đượccoi là sự giải thích thông tin cảm giác để hình thành một đại diện tinh thần của thế giới (Schacter et al. 2020). Ví dụ, một chiếc xe tự lái cảm nhận thế giới với sự trợ giúp của nhiều cảm biến (lidar, máy ảnh, gia tốc kế, cảm biến âm thanh, etc.). Nhận thức được phân phối trong nhiều thành phần tri giác theo cách mà mỗi thành phần nhận ra một loại đối tượng (ví dụ: người đi bộ, tín hiệu giao thông, các phương tiện khác, v.v.) và cập nhật một đại diện động của môi trường. Nói chung, thế giới nhận thức có thể bao gồm các bộ phận của cơ thể cũng có thể được hệ thống cảm nhận[[3]](#footnote-3).
* *Kiểm soát hành động*. Kiểm soát hành động đề cập đến khả năng của một tác nhân kiểm soát việc thực hiện các hành động của chính nó. Kiểm soát hành động gửi đơn đặt hàng đến bộ truyền động và sử dụng data từ môi trường được đo bằng các cảm biến cụ thể hoạt động trong một vòng kín liên tục. Ví dụ, việc kiểm soát tốc độ của một chiếc xe tự hành tác động lên chân ga bằng cách liên tục quan sát dữ liệu từ một cảm biến đo tốc độ. Cont rol toàn cầucó thể được chia thành một tập hợp các *hành vi*

*bộ điều khiển* và mỗi bộ điều khiển chịu trách nhiệm về một khía cạnh điều khiển cụ thể (ví dụ: phanh, đánh lái, tăng tốc, v.v.) [[4]](#footnote-4).

* *Tương tác*. Khả năng tương tác với các tác nhân khác dựa trên việc sử dụng một ngôn ngữ chung. Cáchệ thống artificial có thể giao tiếp với người dùng bằng các cơ chế đơn giản hóa (ví dụ: giao diện người dùng đồ họa, lệnh có tiền tố, v.v.). Ví dụ, trong một chiếc xe tự lái, hành khách có thể chỉ ra điểm đến để đi và chiếc xe trình bày con đường cần đi cùng với thông tin về môi trường giúp hành khách giám sát việc thực hiện chính xác chuyến đi. Bên cạnh các cơ chế này, tương tác với người dùng cũng có thể dựa trên ngôn ngữ của con người bằng cách sử dụng các thuật toán về xử lý ngôn ngữ tự nhiên.
* *Deliblý luận sai lầm*. Một tác nhân thông minh có thể suy luận để, ví dụ: lập kế hoạch hành động (ví dụ: tìm một chuỗi các bước để di chuyển từ nơi này đến nơi khác) hoặc chẩn đoán các vấn đề (ví dụ: xác định nguyên nhân thất bại). Trong bài báo này, chúng tôi sử dụng thuậtngữ cân nhắc thuật ngữ của anh ấy để chỉ loại khả năng này, vì nó được sử dụng bởi các tác giả khác nhau (Wooldridge 2009) (Ingrand và Ghallab 2017) (Murphy 2019).

Như Hình 3 cho thấy, kiểm soát nhận thức và hành động có thể được chia thành nhiều thành phần riêng biệt thường hoạt động song song với tần suất cao. Tuy nhiên, cân nhắc và giao tiếp thường hoạt động tuần tự cần thiết cho sự trấn an mạch lạc(Laird, Lebiere và Rosenbloom 2017) và giao tiếp nhất quán.

### Hành vi phản ứng so với lý luận cân nhắc

Hành vi toàn cầu xuất hiện từ các bộ điều khiển hành vi cá nhân là phản ứng theo nghĩa là nó có thể phản ứng nhanh chóng vớits của môi trường. Phản hồi ngay lập tức của tình trạng thế giới được thực hiện thông qua các cảm biến[[5]](#footnote-5). Hành vi phản ứng này được quan sát, ví dụ, trong phản xạ của động vật với các dạng não nguyên thủy. Nó cũng có thể được quan sát trong suy nghĩ trực giác của con người, khi các quyết định dựa trên trực giác, tức là dựa trên cảm xúc hơn là dựa trên sự thật hoặc bằng chứng.

Để thay thế cho hành vi như vậy, lý luận cân nhắc sử dụng thông tin tưởng tượng về thế giới để đánh giá các giả thuyết thay thế. Ví dụ,người chơi c hess tưởng tượng hậu quả của các bước di chuyển trước khi quyết định một trong những thích hợp nhất. Lý luận cân nhắc sử dụng niềm tin, tức là, các đề xuất mà tác nhân cho là đúng (vĩnh viễn hoặc tạm thời). Niềm tin mới có thể được suy ra bằng suy luận logic hoặc bằng các hình thức lý luận khác (quy nạp, lý luận tương tự, v.v.).

Các kết luận đạt được bằng lý luận cân nhắc có thể được sử dụng để kích hoạt và ức chế các bộ điều khiển hành vi nhất định. Cân nhắc có thể quản lý các mục tiêu mâu thuẫn và có thểinhibi t hành vi phản ứng khi cần phải hoãn các câu trả lời ngay lập tức để đạt được các mục tiêu dài hạn với mức độ ưu tiên cao hơn. Ví dụ, một người có thể chấp nhận cố tình chịu đựng một số cơn đau nhất định trong quá trình điều trị y tế với kỳ vọng cải thiện sức khỏe trong tương lai.

Sự khác biệt giữa lý luận cân nhắc và hành vi phản ứng đã được Daniel Kahneman (Kahneman 2011) xác định trong suy nghĩ của con người với tên Hệ thống 1 và Hệ thống

2. Hệ thống 1 (hoặc tư duy trực quan) nhanh, trực quan, autheo chủ đề, vô thức và phi ngôn ngữ (khó diễn đạt bằng lời nói). Ngược lại, Hệ thống 2 (hoặc lý luận cân nhắc) chậm, logic, nỗ lực, có ý thức, ngôn ngữ (có thể được mô tả bằng lời nói) và đưa ra các quyết định phức tạp. Trong quá trình phát triển cáchệ thống thông minh dựa trên tính toán, sự kết hợp giữa lý luận cân nhắc và hành vi phản ứng là một vấn đề không tầm thường đòi hỏi các quyết định thiết kế cụ thể (ví dụ: khi nào nên cân nhắc hơn phản ứng).

### Sử dụng ký hiệu

Sử dụng các biểu tượng đã được pointed ra như một tính năng liên quan của trí thông minh. Ví dụ, nhà triết học Ernst Cassirer tuyên bố rằng con người nên được định nghĩa là động vật tạo ra biểu tượng. Allen Newell và Herbert Simon đã đề xuất giả thuyết biểu tượng vật lý, tức là, một hệ thống biểu tượng vật lý có phương tiện cần và đủ của hành động thông minh nói chung (Newell và Simon 1976), là một mô hình thống trị trong trí tuệ nhân tạo cho đến những năm 1990. Gary Marcus lập luận rằng thao tác tượng trưng dường như rất cần thiết cho nhận thức của con người (Marcus 2020).

Các biểu tượng có mặt trong ngôn ngữ được sử dụng để tương tác với tác nhân. Ví dụ, ngôn ngữ của con người được đặc trưng bởi việc sử dụng các ký hiệu dưới dạng từ hoặc các biểu diễn khác (ví dụ: ký hiệu toán học). Các biểu tượng đại diện cho các lớp củacác đối tượng ph ysical hoặc các khái niệm trừu tượng. Các đại lý đồng ý trước một ý nghĩa chung của các ký hiệu để hiểu ngôn ngữ được sử dụng trong quá trình giao tiếp.

Một tác nhân có thể thiết lập ý nghĩa của các biểu tượng thông qua nhận thức (để liên kết mation đục lỗ cảm giácvới các biểu tượng) và hành động điều khiển (để dịch các lệnh tượng trưng thành hành động). Điều này liên quan đến vấn đề nối đất biểu tượng (Harnad 1990), tức là, làm thế nào để căn cứ ý nghĩa của một biểu tượng trong bất kỳ thứ gì khác với các biểu tượng khác. Vấn đề này đã được giải quyết trong robot, ví dụ, với ý tưởng neo tri giác (Coradeschi và Saffiotti 2002) theo đó một biểu tượng (ví dụ: cây khái niệm) và dữ liệu cảm biến (ví dụ: hình ảnh của một cây cụ thể được quan sát qua camera) có liên quan đến nhau bằng cách sử dụng cấu trúc dữ liệu (được gọi là neo) cùng với các quy trình xử lý chuyên biệt.

Một tác nhân có thể đại diện cho niềm tin được sử dụng trong quá trình suy luận với sự trợ giúp của các biểu tượng dễ hiểu đối với các tác nhân khác. Điều này rất hữu ích để giải thích cho người khác lý luận của chính mình và knowledge được sử dụng để đưa ra kết luận, cho thấy nhận thức về trạng thái tinh thần của chính mình[[6]](#footnote-6) (như con người thực hiện thông qua nội tâm). Thuộc tính này (đôi khi được gọi là *khả năng giải thích*) là một đặc điểm mong muốn để một hệ thống thông minh được tin cậy.

## Tài sản 3: Tuân theo các nguyên tắc về tính hợp lý và chuẩn mực xã hội

Như đã trình bày ở trên, một hệ thống thông minh dự kiến sẽ tuân theo hai nguyên tắc cơ bản của hành vi. Princip le đầu tiênđang hành động hợp lý và ảnh hưởng đến hành vi cá nhân của hệ thống đối với các mục tiêu và môi trường của nó. Nguyên tắc thứ hai là về việc tuân theo các chuẩn mực xã hội và ảnh hưởng đến hành vi của hệ thống đối với các tác nhân khác. Các phần sau đây mô tả các nguyên tắc này chi tiết hơn.

### Hành động hợp lý

Một đại lý hành động hợp lý nếu các quyết định mà họ đưa ra về hành động của mình tìm cách tối đa hóa thước đo hiệu suất. Biện pháp này định lượng mức độ đạt được của các mục tiêu mà tác nhân theo đuổi. Ví dụ, trong trò chơi cờ vua, người chơi cố gắng giành chiến thắng đối thủ của mình. Một nhà phân tích tài chính đưa ra quyết định về các khoản đầu tư để tối đa hóa lợi nhuận kinh tế thu được. Các quyết định được đưa ra bởi các đại lý này được định hướng để tối đa hóa hiệu suất measure (số lượng đối thủ bị đánh bại hoặc lợi ích kinh tế thu được). Loại hành vi này đã được mô tả trong trí tuệ nhân tạo là nguyên tắc hợp lý (Newell 1982) và nó cũng đã được phân tích trong các ngành khác như triết học hoặc kinh tế họcy (Edwards 1954).

Hành động hợp lý có liên quan chặt chẽ đến khái niệm trí thông minh. So sánh các biện pháp hiệu suất trong việc đạt được mục tiêu là một cách để so sánh mức độ thông minh. Chúng tôi nói rằng một người chơi cờ vua A (hoặc cố vấn đầu tư A) thông minh hơn một người chơi khác B (hoặc một cố vấn đầu tư khác B) nếu A thắng nhiều trò chơi hơn (hoặc kiếm được nhiều lợi nhuận hơn). Quan điểm dựa trên mục tiêu này là một trong những cách tiếp cận được theo sau bởi các tác giả khác nhau để xác định trí thông minh (Kurzweil 2000) (Goertzel 2006) (Legg và Hutter 2006). Ví dụ Kurzweil định nghĩa trí thông minh là "khả năng sử dụng các nguồn lực hạn chế tối ưu để đạt được mục tiêu".

Lưu ý rằng loại định nghĩa này cung cấp một phương pháp để kiểm tra trí thông minh dựa trên đánh giá của một hệ thống trong việc tối đa hóa một func tion khách quan. Cách tiếp cận này rất linh hoạt vì nó không tạo ra sự tách biệt cứng nhắc nhân tạo giữa các hệ thống thông minh và không thông minh. Ngoài ra, nó có một cách tiếp cận thực tế để xây dựng các máy móc thông minh, tránh các câu hỏi khoa học mở related đến trí thông minh (ví dụ, ý thức phi thường).

### Chuẩn mực xã hội

Hành vi của một hệ thống trí tuệ nhân tạo với sự tôn trọng các tác nhân khác thường bị hạn chế bởi các chuẩn mực xã hội, đặc biệt là khi nó hợp tác với người dùng. Điều này rất quan trọng để đảm bảo rằng hệ thống hoạt động phù hợp với các giá trị đạo đức nhưđảm bảo sự công bằng (tránh thành kiến nhóm) hoặc hạn chế sử dụng có hại.

Ví dụ: một hệ thống thông minh được đào tạo bằng cách sử dụng dữ liệu thiên vị có thể không thể đảm bảo hoạt động công bằng. Tình huống này đã xảy ra với Tay, một chatbot hội thoại thử nghiệmdo Microsoft phát triển vào năm 2016. Chatbot này được đào tạo với dữ liệu công khai không kiểm soát được dựa trên sự tương tác với mọi người, điều này đã tạo ra sự thiên vị phân biệt chủng tộc trong hệ thống. Hành vi đạo đức của xe tự hành đã được phân tích bởi Bonnefon et al. (Bonnefon, Shariff và Rahwan 2016) xem xét các tình huống cực đoan trong đó các phương tiện phải lựa chọn giữa việc chạy qua người đi bộ hoặc hy sinh hành khách của họ để cứu người đi bộ.

Trong bối cảnh này, Russell lập luận rằng các hệ thống trí tuệ nhân tạo nên hoạt động như những cỗ máy có lợi để tối đa hóa việc thực hiện sở thích của con người để đối phó với vấn đề kiểm soát AI[[7]](#footnote-7) (Russell 2019). Theo Russell, hành động của máy móc thông minh dự kiến sẽ đạt được mục tiêu của con người và vìmục tiêu t hese là không chắc chắn, máy móc sẽ trì hoãn việc con người xin phép, chấp nhận sửa chữa và cho phép tắt. Ngoài ra, hành vi của những cỗ máy này nên bị hạn chế bởi các quy tắc và lệnh cấm giống như hành động của con người bị hạn chế bởi luật pháp và các chuẩn mực xã hội.

Loại hành vi này cũng đã được phân tích, ví dụ, sử dụng khái niệm *tác nhân từ* trong các tổ chức xã hội như hệ thống đa tác nhân (Castelfranchi 1998). Lòng nhân từ có thể được hiểu là một loại áp dụng mục tiêu. Một đại lý là nhân từ nếu đại lý thông qua các mục tiêu của người khác và việc áp dụng các mục tiêu đó không giúp đạt được mục tiêu của riêng mình.

## Tài sản 4: Khả năng thích ứng

Khả năng thích ứng cho phép một hệ thống thông minh hoạt động hiệu quả trong các môi trường động phức tạp. Nói chung, một hệ thống được cho là tự thích ứng nếu nó có thể sửa đổi cách nó thực hiện một nhiệm vụ nhất định trong response với những thay đổi trong môi trường để hệ thống có thể thực hiện nhiệm vụ đó hiệu quả hơn.

Cách tiếp cận đầu tiên để đạt được sự tự thích nghi là bằng cách kết hợp lại các hành vi năng động (HayesRoth 1995) (Oreizy et al. 1999) (Molina và Santamaria 2021). Trong trường hợp này, tác nhân sử dụng logic thích ứng với thông tin về hiệu quả của các phương pháp hành vi của chính nó. Trong một tình huống nhất định, tác nhân sử dụng logic thích ứng để quyết định kết hợp phương pháp nào là thích hợp nhất để thực hiện một nhiệm vụ nhất định , theo trạng thái hiện tại của môi trường.

Cách tiếp cận thứ hai để thích ứng là thông qua học tập. Các phần sau đây mô tả chi tiết hơn hai hình thức thích ứng (cá nhân và xã hội) tương ứng dựa trên việc học hỏi từ experience và học hỏi từ những người khác.

### Rút kinh nghiệm

Một hệ thống có năng lực học tập có thể cải thiện hiệu suất của nó trong quá trình tương tác nhiều với môi trường. Ví dụ, một người chơi cờ vua có thể học các chiến lược hiệu quả hơn sau khi chơi nhiều trò chơi. Cách tiếp cận học tập này đã được Tom Mitchell (Mitchell 1997) sử dụng trong bối cảnh học máy, người đã xây dựng định nghĩa sau: "một chương trình máy tính được cho là học hỏi từ kinh nghiệm E đối với một số nhiệmvụ T và đo lường hiệu suất P, nếu hiệu suất của nó tại các nhiệm vụ trong T, như được đo bằng P, cải thiện với kinh nghiệm E ".

Một hệ thống thông minh có khả năng học hỏi cần một phương pháp phán đoán giá trị để đánh giá hành vi của nó. Kết quả của đánh giá như vậy được sử dụng bởi hệ thống thông minh để thay đổi hành vi của nó nhằm cải thiện kết quả đánh giá trong tương lai. Phương pháp đánh giá giá trị có thể là một tác nhân *bên ngoài* làm việc như một người hướng dẫn (hoặc huấn luyện viên hoặc người giám sát) quan sát và đánh giá hành vi của hệ thống thông minh để cung cấp phản hồi (ví dụ: với phần thưởng tích cực hoặc tiêu cực). Hệ thống thông minh cũng có thể sử dụng một phương pháp *nội bộ* đánh giá hành động của chính nó để tạo ra phần thưởng tích cực hoặc tiêu cực bằng cách sử dụng, ví dụ, cảm xúc mô phỏng (như pain hoặc niềm vui) hoặc các cơ chế phức tạp khác dựa trên tự quan sát,

Như lập luận của Silver et al. (Silver et al. 2021), các hệ thống học hỏi thông qua kinh nghiệm thử và sai để tối đa hóa phần thưởng có thể giúp có được nhiều khả năng liên quan đến trí thông minh. Do đó, học tập nên được hiểu là một năng lực phân biệt của các hệ thống thông minh ảnh hưởng đến khả năng nhận thức chính (nhận thức, kiểm soát hành động, lý luận hoặc tương tác). Ví dụ, một robot trên không có thể cải thiện khả năng kiểm soát hành động bằng cách học cách hạ cánh chính xác hơn trên một bề mặt được đánh dấu sau nhiều lần thử nghiệm.

### Học hỏi từ các đại lý khác

Ngoài cách tiếp cận được mô tả ở trên, chúng ta có thể xác định một hình thức học tập khác dựa trên việc chuyển giao kiến thức giữa các cá nhân. Với sơ đồ này, một tác nhân A học được niềm tin mới từ một tác nhân B khác trên cơ sở rằng tác nhân B có thể mô tả kiến thức của chính mình bằng một ngôn ngữ tượng trưng có nghĩa là tác nhân A có thể hiểu được. Hình thức học tập này, cho phép các nhóm tác nhân tiếp thu kiến thức trực tiếp từ các tác nhân khác, thúc đẩy sự thích nghi xã hội với môi trường bởi vì một số tác nhân có thể tiết kiệm nỗ lực phải học hỏi cá nhân từ kinh nghiệm của chính họ.

Ví dụ, hãy xem xét niềm tin được thể hiện là ngựa vằn có hình ngựa và có làn da sọc. Một đặc vụ chưa bao giờ nhìn thấy ngựa vằn có thể nhận ra loài động vật này bằng cách sử dụng mô tả tượng trưng. Hình thức học tập này đã được gọi là *học tập không bắn* (Xian et al. 2018), bao gồm học cách nhận ra các đối tượng mới vớibất kỳ ví dụ nào, chỉ có mô tả tượng trưng của chúng.

# Xây dựng hệ thống thông minh

Phần này tóm tắt hai khía cạnh quan trọng trong việc xây dựng một hệ thống thông minh: (1) lựa chọn các phương pháp trí tuệ nhân tạo và (2) việc tiếp thu kiến thức được sử dụng bởi hệ thống để thực hiện các nhiệm vụ mong muốn (có thể được thực hiệnbằng cách sử dụng con người hoặc với sự trợ giúp của các phương pháp tự động).

## Phương pháp trí tuệ nhân tạo

Sự phát triển của một hệ thống thông minh có thể đòi hỏi phải kết hợp một số phương pháp trí tuệ nhân tạo. Bảng 2 cho thấy các chức năng liên quan đến khả năng nhận thức cùng với một mẫu các phương pháp phổ biến được sử dụng để xây dựng các hệ thống thông minh.

Những phương pháp này có thể được chia theo hai cách tiếp cận phân biệt. Một cách tiếp cận tương ứng với các phương pháp sử dụng các biểu diễn kết nối (ví dụ: mạng nơ-ron phức tạp,mạng nơ-ron nt định kỳ hoặc học tăng cường sâu). Cách tiếp cận khác sử dụng các phương pháp tượng trưng theo ngữ nghĩa chính thức dựa trên logic (ví dụ: suy luận logic tự động, hệ thống dựa trên quy tắc và thuật toán thỏa mãn ràng buộc), dựa trên xác suất (ví dụ: mạng bayes) hoặc phương pháp tiếp cận lai (ví dụ: logic mờ).

Các phương pháp mang tính biểu tượng và kết nối tương ứng với các phương pháp thay thế có điểm mạnh và điểm yếu bổ sung (Franklin 1995). Trên thực tế, có sự thiếu hiểu biết về cáchxử lý thông báo được thực hiện bằng các phương pháp tiếp cận tượng trưng có thể được ánh xạ vào các tính toán được thực hiện bằng các phương pháp kết nối. Thuật ngữ *khoảng cách giải thích tính toán* (Reggia 2014) đã được sử dụng để diễn tả rằng không rõ hai cách tiếp cận có liên quan như thế nào.

Các khả năng cogni tive có thể truy cập có ý thức (chẳng hạn như các quy trình liên quan đến cân nhắc hoặc tương tác) có thể được thực hiện bằng các phương pháp tượng trưng. Những phương pháp này có thể mô tả các ký hiệu đại diện cho kiến thức để biện minh cho cách đạt được một kết luận nhất địnhn.

Mặt khác, kiểm soát nhận thức và hành động thường được thực hiện một cách vô thức và chúng được thực hiện tốt hơn với các cách tiếp cận phi biểu tượng như phương pháp kết nối. Trong thực tế, thật thú vị khi lưu ý rằng việc thực hiện các kỹ năng cảm biến vô thức đòi hỏi tính toán nhiều hơn so với lý luận cân nhắc. Điều này đã được tuyên bố là *nghịch lý của Moravec* (Moravec 1988).

## Tiếp thu kiến thức

Các hệ thống thông minh có thể sử dụng kiến thức khó được chính thức hóa uhát đại diện thủ tục của các ngôn ngữ máy tính như Java, Python hoặc C ++. Ví dụ, hãy xem xét một hệ thống chuyên gia y tế sử dụng kiến thức liên quan đến các triệu chứng và bệnh tật. Kiến thức này có thể được mô hình hóa bằng cách sử dụng biểu diễn khai báo tượng trưngs (ví dụ: quy tắc) được lưu trữ trong một *cơ sở kiến thức*. Một thuật toán suy luận xử lý các quy tắc như vậy để chẩn đoán bệnh của một bệnh nhân nhất định. Thuật toán suy luận là chung chung và do đó, nó có thể được thực hiện bằng các công cụ phần mềm chung. Tuy nhiên, các quy tắc của mô hình kiến thức là cụ thể và phải được viết cho hệ thống chuyên gia cụ thể. Việc xây dựng một cơ sở tri thức của các nhà phát triển được gọi là *tiếp thu kiến thức*. Nhiều cơ sở kiến thức của các hệ thống chuyên gia cổ điển đại diện cho kiến thức ic heuristđược thu thập thủ công từ các chuyên gia miền và được chính thức hóa bằng cách sử dụng các biểu diễn khai báo (ví dụ: quy tắc hoặc logic bậc nhất). Việc tạo ra và duy trì các cơ sở kiến thức như vậy thường đòi hỏi nỗ lực đáng kể. Do đó, uisition acq thủ côngbị hạn chế cho việc phát triển các cơ sở kiến thức không quá lớn. Vấn đề này đã được gọi là *nút thắt cổ chai thu nhận kiến thức*.

Để tạo điều kiện thuận lợi cho nhiệm vụ này, lĩnh vực kỹ thuật bản thể học trong trí tuệ nhân tạo đã đề xuấtcác olutions kỹ thuật để hình thành các biểu diễn chung, được hiểu là bản thể học, cho phép chia sẻ và tái sử dụng nội dung của các cơ sở kiến thức trên các hệ thống thông minh khác nhau.

## Phương pháp tự động xây dựng hệ thống thông minh

Trái ngược với kiến thức thủ côngacqu isition, có thể áp dụng các phương pháp tự động để xây dựng các hệ thống thông minh, đã trở nên phổ biến trong những thập kỷ qua do tính khả dụng ngày càng tăng

|  |  |  |  |
| --- | --- | --- | --- |
| Khả năng nhận thức | Hàm ví dụ | Ví dụ về các chức năng chi tiết | Ví dụ về phương pháp tính toán |
| Tri giác | Trích xuất tính năng | Nhận dạng hình ảnh, trích xuất thông tin, xử lý tín hiệu, cơ chế chú ý | Mạng nơ-ron (ví dụ: mạng nơ-ron phức tạp), nhận dạng mẫu methods |
| Dữ liệu  Giải thích | Trừu tượng hóa dữ liệu, hợp nhất dữ liệu, nối đất biểu tượng | Phân tích dữ liệu thống kê, mở rộng  Bộ lọc Kalman, neo tri giác |
| Thảo luận | Lý luận mô tả | Phân loại, chẩn đoán, chiếu thời gian (dự đoán), sửa đổi niềm tin | Suy luận logic tự động, mạng bayes, hệ thống dựa trên quy tắc, mạng ngữ nghĩa, hệ thống duy trì sự thật |
| Lý luận quy định | Lập kế hoạch, lên lịch, cấu hình, phân công, suy luận mục tiêu | Lập kế hoạch tự động, thuật toán thỏa mãn ràng buộc, mô hình BDI |
| Tương tác | Nhận dạng và tạo ngôn ngữ | Hiểu ngôn ngữ tự nhiên, nhận dạng giọng nói, tạoion ngôn ngữ tự nhiên | Phương pháp xử lý ngôn ngữ tự nhiên  (ví dụ: mạng lưới thần kinh tái phát, máy biến áp) |
| Quản lý hộp thoại | Phối hợp đối thoại, lập kế hoạch diễn ngôn | Lý thuyết hành động giao tiếp, giao thức tương tác, lập kế hoạch tự động |
| Kiểm soát hành động | Lựa chọn hành động | Lựa chọn hành động theo mục tiêu hoặc theo sự kiện, quản lý xung đột hành động | Máy trạng thái hữu hạn, cây hành vi,  Lưới Petri |
| Kiểm soát thực thi hành động | Điều khiển chuyển động, điều khiển thao tác | Lý thuyết điều khiển, điều khiển mờ, học tăng cường sâu |

Bảng 2: Các chức năng liên quan đến khả năng nhận thức cùng với các ví dụ về phương pháp tính toán được sử dụng để xây dựng các hệ thống thông minh.

của một lượng lớn dữ liệu và sức mạnh tính toán cao hơn. Các phương pháp này có thể sử dụng các nguồn thông tin với các định dạng khác nhau như hình ảnh (ảnh hoặc video), dữ liệu có cấu trúc (ví dụ: cơ sở dữ liệu có dữ liệu chữ và số hoặc chuỗi thời gian) hoặc văn bản phi cấu trúc được viết bằng ngôn ngữ tự nhiên.

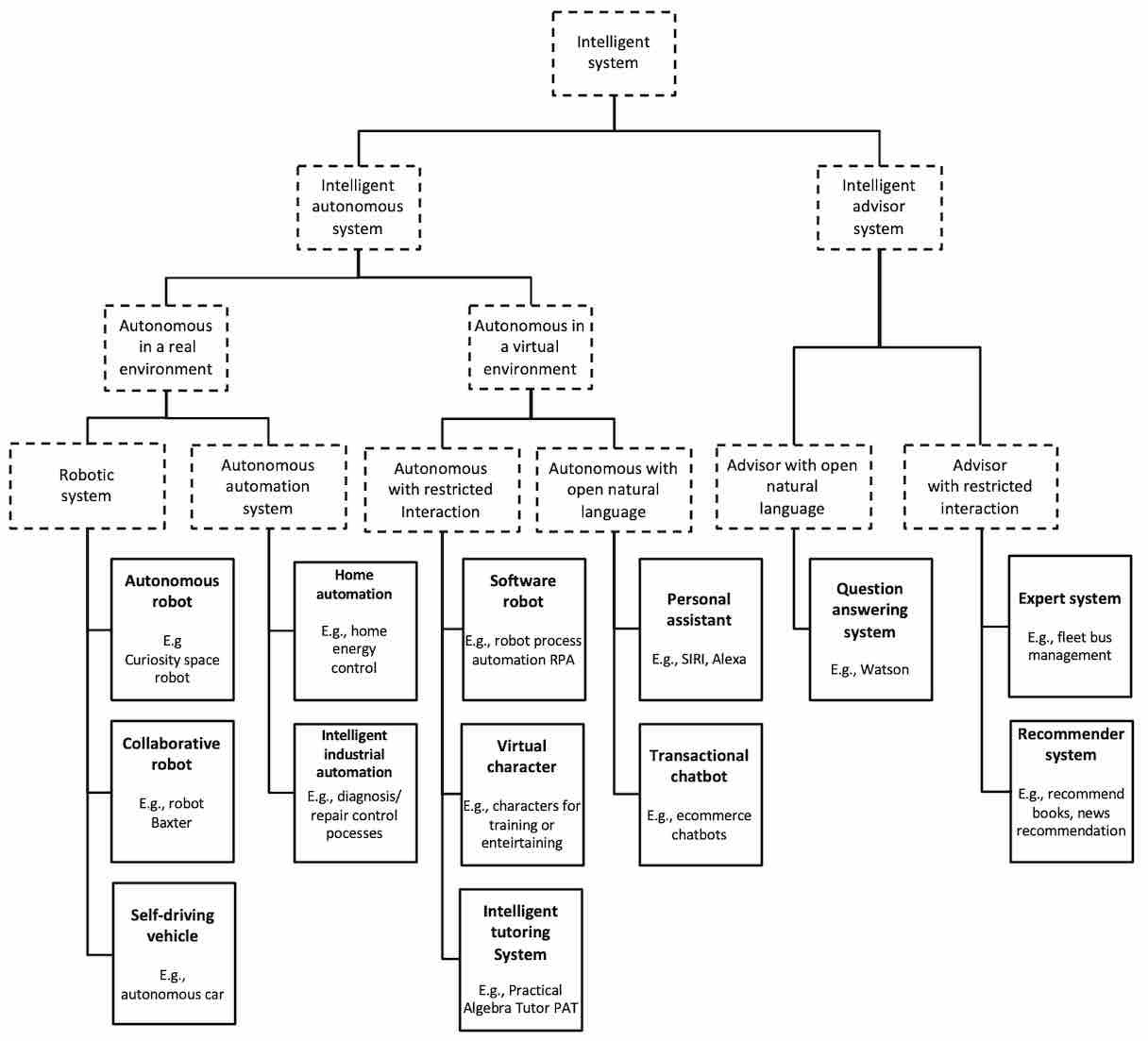
Các phương pháp học máy có thể được sử dụng để phát triển các thành phần thực hiệncác khả năng nhận thức ific thông số kỹ thuật sẽ là một phần của hệ thống thông minh. Ví dụ, một mạng lưới thần kinh có thể được đào tạo bằng cách sử dụng hàng nghìn hình ảnh của các khối u được dán nhãn là ví dụ tích cực và tiêu cực. Mạng lưới thần kinh được đào tạo có thể được sử dụng như một phần của hệ thống intelligent giúp các bác sĩ nhận ra sự hiện diện của các khối u. Để chế tạo một robot trên không có thể hạ cánh trên một bề mặt chuyển động, cũng có thể áp dụng một phương pháp dựa trên học tăng cường sâu và mô phỏng robot hoạt động trong môi trường al điêu luyệnđể huấn luyện trước một mạng lưới thần kinh có thể điều khiển chuyển động của robot trong quá trình hạ cánh trên bề mặt. Để xây dựng một hệ thống chuyên gia dựa trên quy tắc để chẩn đoán y tế, một phương pháp cảm ứng quy tắc có thể được áp dụng cho một cơ sở dữ liệu với các ví dụ về chẩn đoán liên quan đến các bệnh và triệu chứng để tạo ra các quy tắc ứng cử viên cho cơ sở kiến thức.

Phương pháp xử lý ngôn ngữ tự nhiên cũng có thể được sử dụng như một phương pháp tự động để hỗ trợ xây dựng các hệ thống thông minh. Các phương pháp này có thể được sử dụng để trích xuấtcạnh biết từ các tài liệu văn bản được viết bằng ngôn ngữ tự nhiên. Ví dụ: một kho tài liệu văn bản lớn tương ứng với một miền chuyên nghiệp có thể được phân tích bằng cách sử dụng các kỹ thuật xử lý ngôn ngữ tự nhiên để tạo ra các trường hợp của các lớp và và các mối quan hệ được sử dụng bởi hệ thống câu hỏi / câu trả lời.

Lưu ý rằng hai phương pháp này, học máy và xử lý ngôn ngữ tự nhiên, có thể đóng các vai trò khác nhau trong sự phát triển của các hệ thống thông minh. Một mặt, chúng có thể được sử dụng ngoại tuyến như được mô tả trong giáo phái này, đóng vai trò là công cụ được nhà phát triển sử dụng để tạo các phần tự động của hệ thống. Mặt khác, chúng có thể được sử dụng trực tuyến, tức là, hệ thống thông minh cuối cùng bao gồm các thành phần sử dụng các phương pháp đó để thực hiện khả năng thích ứng (sử dụngphương pháp học mac hine) hoặc khả năng tương tác bằng ngôn ngữ tự nhiên (sử dụng phương pháp xử lý ngôn ngữ tự nhiên).

# Ví dụ về các hệ thống thông minh

Phần này trình bày các ví dụ minh họa về các hệ thống thông minh trong thế giới thực bằng cách sử dụng đặc tính được trình bày trong bài báo này. Để cung cấp một cái nhìn toàn cầu, Hình 4 cho thấy các hệ thống thông minh đa dạng được tổ chức trong một hệ thống phân cấp các danh mục. Tìnhtrạng hỗn loạn hie này không nhằm mục đích chính xác và đầy đủ mà là minh họa cho loại hệ thống hiện có. Các phần sau đây mô tả chi tiết hơn một số danh mục này.



Robot tự hành

Một vụ cướp tự trịot có thể được coi là một trường hợp đại diện của hệ thống thông minh. Một ví dụ về hệ thống này là một robot trên không để kiểm tra máy bay được phát triển như một nguyên mẫu trình diễn vào năm 2019 cho công ty Airbus (Bavle 2019). Hệ thống này thực hiện các chức năng sau:

* *Nhận thức*. Robot trên không được trang bị nhiều cảm biến như (1) đơn vị đo lường quán tính (IMU) để cung cấp thông tin về hướng, vận tốc góc và gia tốc tuyến tính, (2) camera để tính toán vận tốc bằng cách sử dụng phép đo đường thị giác, (3) lidar sensor tạo ra đám mây điểm 3D và (4) camera phía trước có độ phân giải cao. Hệ thống nhận thức thực hiện hợp nhất dữ liệu từ nhiều cảm biến để bản địa hóa chính xác. Camera độ phân giải cao được sử dụng để phát hiện sự bất thường trên bề mặt của airplane bằng thị giác máy tính.
* *Cân nhắc*. Trong hệ thống này, chức năng cân nhắc chính là lập kế hoạch đường dẫn. Hệ thống có thể tạo ra một kế hoạch kiểm tra để che phủ bề mặt của máy bay. Ngoài ra, hệ thống lưu trữ hình ảnh để phân tích sau.
* *Hành động control*. Điều khiển hành động tương ứng với việc thực hiện kế hoạch và điều khiển chuyển động để hoàn thành kế hoạch điều hướng. Điều khiển hành động cũng bao gồm một cơ chế phản ứng để tránh các chướng ngại vật bất ngờ được phát hiện với sự trợ giúp của cảm biến lidar.
* *Tương tác*. Người vận hành chuyên nghiệpvide hình dạng của máy bay và bắt đầu nhiệm vụ. Người vận hành có thể dừng thực hiện nhiệm vụ nếu trường hợp khẩn cấp xảy ra.

Nói chung, một hệ thống robot thông minh là một trong những loại hệ thống thông minh phức tạp nhất vì nó thường bao gồm phần lớn các đặc điểm được xem xét trong bài báo này và ngoài ra, nó hoạt động trong môi trường thế giới thực vớicác ràng buộc vật lý deman ding. Trong danh mục này, có nhiều loại hệ thống với các loại phức tạp khác nhau như robot cộng tác công nghiệp (ví dụ: Baxter), xe tự hành (ví dụ: xe tự lái, thám hiểm hành tinh Curiosity), rô bốt hỗ trợ , robot nội địa (ví dụ: robot dọn dẹp nhà cửa Roomba), rô bốt giải trí (ví dụ: Sony Aibo, Furby), v.v.

## Hệ thống chuyên gia

Sự phát triển của các hệ thống chuyên gia đã góp phần vào sự thành công thực tế của AI trong những năm 1980 với nhiều hệ thống thương mại trongcác lĩnh vực multipl e. Chúng được áp dụng để thực hiện các tác vụ tự động như sau:

* *Phân loại*: Tìm danh mục của một đối tượng bằng cách sử dụng các quan sát (ví dụ: xác định sự hiện diện của khoáng sản, đề xuất một loại hình đầu tư, phát hiện các hành vi bất thường).
* *Tiên lượng hạt dẻ:* Tìm nguyên nhân gây ra các triệu chứng (ví dụ: chẩn đoán y tế, chẩn đoán cơ học).
* *Dự báo thời gian*: Ước tính hành vi trong tương lai của một hệ thống động (ví dụ: dự đoán thị trường kinh tế).
* *Phân công/Lên lịch*: Chỉ định một tập hợp các tài nguyên cho một nhóm các nhu cầu(ví dụ: chỉ định máy bay cho các cổng trong sân bay).
* *Cấu hình*: Tạo một thiết kế đáp ứng một số yêu cầu (ví dụ: thiết kế máy móc của thang máy).
* *Lập kế hoạch*: Tìm một tập hợp các hành động để đạt đến trạng thái mong muốn (ví dụ: lập kế hoạch sơ tán khẩn cấp)

### Hệ thống chuyên gia thế hệ đầu tiên

Thế hệ đầu tiên của các hệ thống chuyên gia được sử dụng chủ yếu là lý luận heuristic với các đại diện kiến thức tượng trưng như các quy tắc. Ví dụ, Mycin là một trong những hệ thống chuyên gia đầu tiên được sử dụng thành công như một mô hình trong các loại vấn đề khác. Mycin được phát triển tại đại học Stanford vào những năm 1970 (Buchanan và Shortliffe 1984) và nó đã có thể chẩn đoán các bệnh truyền nhiễm và đề xuất các liệu pháp. Hệ thống này có hai thành phần chức năng chính sau:

* *Tương tác.* Thông tin về bệnh nhâns có được thông qua giao tiếp với người dùng. Đối thoại với người dùng được thực hiện trong quá trình trả lời câu hỏi với các câu hỏi và câu trả lời có tiền tố. Mycin đặt câu hỏi về bệnh nhân và nuôi cấy. Các câu hỏi được điều chỉnh cho phù hợp với câu trả lời trước đós (câu hỏi có ngữ cảnh). Người dùng có thể trả lời không chắc chắn và có thể yêu cầu giải thích. Mycin đưa ra khuyến nghị chẩn đoán và điều trị
* *Thảo luận.* Mycin chẩn đoán nguyên nhân của các triệu chứng tìm thấy các sinh vật truyền nhiễm có thể. Mycin sử dụngphương pháp phân loại heuristic với nền tảng kiến thức với 450 quy tắc và chuỗi ngược. Mycin bao gồm một phương pháp ban đầu cho lý luận gần đúng (các yếu tố chắc chắn dưới dạng các giá trị trong [-1, +1]). Ngoài ra, Mycin khuyến nghị các liệu pháp sử dụng phương pháp based trên một quy trình tạo và thử nghiệm.

Các phương pháp được sử dụng để xây dựng hệ thống chuyên gia đã chứng minh rằng chúng hợp lệ trong nhiều vấn đề, nhưng chúng đưa ra những hạn chế được sử dụng trong một số trường hợp nhất định. Một trong những hạn chế quan trọng nhất là vấn đề được đề cập above gọi là *nút thắt cổ chai thu nhận kiến thức* hạn chế cách tiếp cận này để phát triển các cơ sở tri thức không quá lớn.

Hệ thống chuyên gia có thể giải quyết các vấn đề khó khăn, nhưng chúng hoạt động trong các lĩnh vực chuyên ngành hẹp, không có kiến thức chung về phần còn lại của thế giới . Do đó, so với các chuyên gia con người, những người cũng có kiến thức chung, các hệ thống chuyên gia rất giòn và không thể phản ứng chính xác trong các tình huống bất ngờ.

Một vấn đề khác liên quan đến cách các hệ thống này obtain thông tin về môi trường. Ví dụ, Mycin nhận thông tin về bệnh nhân bằng cách sử dụng một bộ câu hỏi và câu trả lời có tiền tố. Cơ chế giao tiếp này có thể quá hẹp và cứng nhắc để được sử dụng trong các tình huống khi định dạng của dữ liệu có sẵn đa dạng hơn (ví dụ: văn bản phi cấu trúc). Hạn chế này gây khó khăn cho việc tích hợp hệ thống chuyên gia vào các hoạt động hàng ngày.

### Hệ thống chuyên gia quản lý môi trường cảm biến

Hành vi của môi trường động có thể được theo dõi với sự trợ giúp của mạng cảm biến và hệ thống thông tin. Đây là trường hợp, ví dụ, cơ sở hạ tầng thông tin hiện đại đã được phát triển trong những thập kỷ qua, ví dụ như áp dụng cho các lĩnh vực chiến lược như thành phố thông minh (giao thông, khí hậu, ô nhiễmkhi giám sát) hoặc công nghiệp 4.0 (với các cài đặt công nghiệp cảm biến).

Trong bối cảnh này, nó xuất hiện nhu cầu sử dụng các hệ thống thông minh tư vấn cho các nhà khai thác đưa ra quyết định cho việc quản lý các môi trường như vậy. Có nhiều ví dụ vềcác hệ thống e xpert thuộc thể loại này. Ví dụ, có các hệ thống quản lý giao thông công cộng (Molina 2005), lên lịch và điều phối tại sân bay (Jo et al. 2000), hỗ trợ quyết định khẩn cấp trong lũ lụt (Molina và Blasco 2003) hoặcvận hành điện ele ctric (Filho et al. 2012).

Ví dụ đầu tiên tương ứng với một hệ thống chuyên gia phát hiện các sự cố và đề xuất hành động trong mạng lưới giao thông xe buýt đô thị. Hệ thống này được đưa vào năm 2002 như một phần của hệ thống quản lý độixe của trung tâm điều khiển xe buýt t he của thành phố Vitoria ở Tây Ban Nha. Hệ thống này có các khả năng chính sau:

* *Nhận thức*. Hệ thống thu thập dữ liệu từ xe buýt sử dụng thông tin GPS và bản địa hóa vị trí của từng xe buýt trong bản đồ các tuyến giao thông.
* *Cân nhắc*. Hệ thống phát hiện sự hiện diện của các sự cố (ví dụ: phát hiện sự chậm trễ dựa trên vị trí hiện tại và vị trí được lên kế hoạch) và dự đoán hành vi trong tương lai (ví dụ: một mô hình được sử dụng về nhu cầu vận chuyển lịch sử). Hệ thống xác định actions được thực hiện để quản lý sự cố (một kế hoạch tự động dựa trên kế hoạch mạng nhiệm vụ phân cấp (HTN) được sử dụng để tạo kế hoạch hành động).
* *Tương tác*. Hệ thống trình bày cho người vận hành các sự cố được phát hiện (ví dụ: sự chậm trễ của xe buýt). Nhà điều hành có thể thông báocác sự kiện oth er cho hệ thống (ví dụ: xe buýt bị hỏng, đường phố bị chặn, v.v.). Hệ thống đề xuất các hành động (ví dụ: sử dụng thêm xe buýt, gửi xe tải sửa chữa, v.v.). Hệ thống biện minh cho các hành động (ví dụ: tại sao một số dòng được ưu tiên dựa trên nhu cầu dự kiến).

Nói chung, loại hệ thống này bao gồm một thành phần *nhận thức* để quan sát và giải thích các đặc điểm của hành vi môi trường bằng cách sử dụng mạng cảm biến hoặc hệ thống thông tin. Ngoài ra, hệ thống bao gồm một thành phần *cân nhắc* để lấy racác lusions conc từ dữ liệu, chẳng hạn như:

* những gì xảy ra trong môi trường (phát hiện),
* tại sao nó xảy ra (chẩn đoán),
* điều gì có thể xảy ra trong tương lai (dự đoán thời gian, phân tích điều gì sẽ xảy ra nếu), và
* những gì nên được thực hiện (lập kế hoạch).

Hệ thống cũng bao gồm một thành phần *tương tác*  để trình bày thông tin này cho người vận hành. Loại hệ thống thông minh này không có thành phần *truyền động* để sửa đổi môi trường. Người vận hành chịu trách nhiệm đưa ra quyết định quản lý và thực hiện các hành động trong môi trường. Do đó, điều quan trọng là thành phần tương tác trình bày các đề xuất hệ thống với sự biện minh đầy đủ theo cách mà người vận hành có thể chịu trách nhiệm về các quyết định.

## Hệ thống trả lời câu hỏi dựa trên tính toán nhận thức

IBM đãphát triển một hệ thống có tên Watson (Ferrucci et al. 2010) đánh dấu một cột mốc quan trọng trong lịch sử trí tuệ nhân tạo khi giành giải nhất của chương trình truyền hình đố vui "Jeopardy" vào năm 2011 cạnh tranh với các đối thủ của con người. System đã có thể trả lời tốt hơn con người cho các câu hỏi miền mở được viết bằng ngôn ngữ tự nhiên về văn hóa nói chung.

Trong trò chơi đố vui, Watson phải đoán câu trả lời dựa trên manh mối nhất định do người dùng cung cấp. Ví dụ, Watson nhận được một manh mối như "loại thuốc này đã được chứng minh là làm giảm các triệu chứng của ADD với tương đối ít tác dụng phụ". Trong vòng chưa đầy 5 giây, Watson đã phải đưa ra câu trả lời được thể hiện là: "Ritalin là gì?" (theo quy tắc của trò chơi, câu trả lời "Ritalin" được trình bày như một câu hỏi).

Watson đã sử dụng một kho tài liệu văn bản lớn được chọn lọc gồm khoảng 2 triệu trang làm nguồn thông tin (ví dụ: tin tức, bách khoa toàn thư và tác phẩm văn học). Watson tìm kiếm câu trả lời của ứng cử viên trong một kho tư liệu như vậy có thể là câu trả lời có thể có cho các câu hỏi được xây dựng bởi người dùng cuối bằng ngôn ngữ tự nhiên. Watson sử dụng cấu trúc phức tạp của các thuật toán để (1) chọn các từ khóa tìm kiếm có liên quan có thể được sử dụng để tìm ra (một phần) câu trả lời một phần của ứng cử viên và (2) xếp hạng và tích hợp các câu trả lời một phần để tạo ra câu trả lời cuối cùng bằng ngôn ngữ tự nhiên.

Các đặc điểm chung của Watson có thể được mô tả bằng cách sử dụng các thành phần được trình bày trong bài báo này. Trong đặc tính của chúng tôi, chúng tôi có thể coi kho tài liệu là môi trường mà hệ thống hoạt động để thực hiện nhiệm vụ tìm kiếm của nó. According cho điều này, các thành phần của Watson như sau:

* *Tương tác*. Thành phần tương tác của Watson sử dụng xử lý ngôn ngữ tự nhiên để trích xuất cấu trúc ngôn ngữ của manh mối đã cho. Thành phần tương tác cũng tạo ra câu trả lời cuối cùng trongngôn ngữ nat ural.
* *Cân nhắc*. Cân nhắc trong Watson bao gồm hai phần. Phần đầu tiên tạo ra một số khóa tìm kiếm từ cấu trúc ngôn ngữ của manh mối đã cho. Mỗi khóa tìm kiếm được tạo ra để sử dụng sau này (bởi thành phần nhận thức) để truy xuất thông tin từ các tài liệu văn bản. Phần thứ hai của xếp hạng cân nhắc và kết hợp các câu trả lời ứng cử viên được truy xuất. Với mục đích này, Watson đã có thể hợp nhất các câu trả lời (Kalyanpur et al. 2012) thực hiện lý luận thời gian, lý luận không gian địa lý và phân loại reasoning (subsumption, districintness, v.v.) bằng cách sử dụng bản thể và cơ sở dữ liệu (ví dụ: Yago, DBPedia, Freebase). Để ghi điểm câu trả lời, Watson sử dụng các mô hình được tạo ra bởi học máy.
* *Nhận thức*. Để có được câu trả lời của ứng viên từ kho tài liệu, Watson đã sử dụng hàng chục thuật toán chuyên biệt để trích xuất thông tin bằng cách sử dụng xử lý ngôn ngữ tự nhiên.

Watson là một trường hợp đại diện thành công của một hệ thống trả lời câu hỏi sử dụng phương pháp tính toán được gọi là *điện toán nhận thức*. Ý tưởng của approach này là hệ thống sử dụng như một nguồn thông tin một lượng lớn dữ liệu có cấu trúc và phi cấu trúc (văn bản, hình ảnh, v.v.). Hệ thống thực hiện các nhiệm vụ nhận thức tự động bao gồm trích xuất tính năng (ví dụ: sử dụng xử lý ngôn ngữ tự nhiên hoặc thị giác máy tính), tổng hợp câu trả lời (lý luận với bản thể hoặc cơ sở kiến thức), giao tiếp ngôn ngữ tự nhiên (sử dụng sự hiểu biết và tạo ngôn ngữ tự nhiên) và gán bằng chứng cho các giả thuyết bằng cách sử dụng các mô hình được học bằng máy học.

# Conclusions

Bài báo này đã trình bày một đặc điểm của một hệ thống thông minh đã được xây dựng với mục đích hữu ích cho các kỹ sư hệ thống để phân tích và thiết kế loại hệ thống này. Bài báo định nghĩa một hệ thống thông minh bằng cách sử dụng một số terminology nhất định với ý nghĩa thường được chấp nhận trong trí tuệ nhân tạo. Mô tả tập trung vào các đặc điểm chức năng liên quan đến trí thông minh mà máy móc hiện có khả năng thể hiện với tình trạng công nghệ hiện tại.

Trái ngược với các công việc khác, mô tả được trình bày trong bài báo này được định hướng để cung cấp một cái nhìn toàn cầu nói chung, với các tính năng thông thường có thể có hoặc không có trong một hệ thống cụ thể. Toàn bộ quan điểm này tạo ra một bối cảnh thuận tiện để mô tả chính xác hơn các khái niệm và xác địnhmối quan hệ thừa kế t. Ví dụ, đặc tính của chúng tôi mở rộng các định nghĩa khác để xem xét rõ ràng yếu tố xã hội của một hệ thống thông minh. Điều này bao gồm sự tương tác với các tác nhân khác bằng cách sử dụng ngôn ngữ và tuân theo các chuẩn mực xã hội. Ngôn ngữ rấtphấn khích khi sử dụng các biểu tượng, đây là một đặc điểm quan trọng của trí thông minh con người và các chuẩn mực xã hội rất quan trọng, đặc biệt là khi hệ thống thông minh nên hợp tác với người dùng theo các giá trị đạo đức. Ngoài ra, khả năng thích ứng được trình bày theo nghĩa rộng, liên quan đến các hình thức học tập khác nhau (từ kinh nghiệm và từ những người khác).

Đặc tính cung cấp một cái nhìn thống nhất về ý nghĩa của một hệ thống thông minh tích hợp các quan niệm khác nhau về các hệ thống đó với cáchệ thống cổ điển đồng ver (ví dụ: hệ thống dựa trên quy tắc), hệ thống dựa trên hành vi hoặc các quan niệm gần đây hơn cho các hệ thống dựa trên AI tự trị. Các thành phần chức năng được sử dụng trong định nghĩa của một hệ thống thông minh có liên quan đến các phương pháp AI phổ biến (dựa trên cả phương pháp kết nối và biểu tượng). Bài báo cũng chứng minh cách đặc tính được trình bày có thể được sử dụng để mô tả các chức năng của các hệ thống thực khác nhau (ví dụ: hệ thống chuyên gia và robot tự động).

Là công việc trong tương lai, chúng tôi hy vọng sẽ thích nghi và mở rộngđặc điểm chung của anh ấy để tính đến kết quả thu được từ nghiên cứu mới về các mô hình tính toán mới của trí thông minh.

# Tham khảo

Arkin, Ronald C. (1998). *Robot dựa trên hành vi*. Báo chí MIT.

Bavle, Hriday (2019). "Định vị và lập bản đồ cho robot trên không bằng cách sử dụng nhận thức trên tàu cho các nhiệm vụ tự trị". Luận án tiến sĩ. ETSI Industriales, Đại học Politecnica de Madrid.

Bonnefon, Jean-François, Azim Shariff và Iyad Rahwan (2016). "Tìnhthế tiến thoái lưỡng nan của xe tự hành". Trong: *Khoa học* 352.6293.

Brooks, Rodney (1991). "Trí thông minh không có đại diện". Trong: *Trí tuệ nhân tạo* 47.1-3, trang 139–159.

Buchanan, BG và E.H. Shortliffe (1984). *Hệ thống chuyên gia dựa trên quy tắc: MYCIN Experiments của Dự án lập trình heuristic Stanford*. Đọc, MA: Addison-Wesley.

Castelfranchi, Cristiano (1998). "Mô hình hóa hành động xã hội cho các tác nhân AI". Trong: *Trí tuệ nhân tạo* 103.1-2.

Coradeschi, S. và A. Saffiotti (2002). "Neo tri giác: Một khái niệm quan trọng để thực hiện kế hoạch trong các hệ thống nhúng". Trong: *Những tiến bộ trong điều khiển dựa trên kế hoạch của các tác nhân robot*. Berlin, Heidelberg: Springer, các trang 89–105.

Duval, Shelley và Robert A. Wicklund (1972). "Một lý thuyết về nhận thức bản thân khách quan." Trong.

Edwards, Ward (1954). "Lý thuyết về việc ra quyết định." Trong: *Bản tin tâm lý* 51.4, tr. 380.

Ferrucci, David và cộng sự (2010). "Building Watson: Tổng quan về dự án DeepQA". Trong: *AI*

*Tạp chí* 31.3, các trang 59–79.

Filho, Joao Inacio Da Silva và cộng sự (2012). " Hỗ trợ quyết định vận hành hệ thống điện bởi hệ thống chuyên gia được xây dựng với logic chú thích paraconsistent". Trong: *Những tiến bộ trong hệ thống chuyên gia*. IntechOpen (bằng tiếng Anh).

Franklin, S. (1995). *Tâm trí nhân tạo*. Báo chí MIT.

Franklin, S. và A. Graesser (1996). "Nó là một Đại lý, hay chỉ là mộtgram Pro?: Một phân loại cho tự trị

Đại lý". Trong: *Hội thảo quốc tế về lý thuyết đại lý, kiến trúc và ngôn ngữ*.

Goertzel, Ben (2006). *Mô hình ẩn: Một triết lý mô hình của tâm trí*. Nhà xuất bản toàn cầu.

Harnad, Stevan (1990). "Vấn đề nối đất biểu tượng". Trong: *Physica D: Hiện tượng phi tuyến*

42.1-3.

Hayes-Roth, Barbara (1995). "Một kiến trúc cho các hệ thống thông minh thích ứng". Trong: *Nhân tạo*

*Tình báo* 72, các trang 329–365.

Ingrand, Felix và Malik Ghallab (2017). "Cân nhắc cho robot tự động: Một cuộc khảo sát". Trong:

*Trí tuệ nhân tạo* 247, trang 10–44.

Jo, G. S. và cộng sự (2000). "Hệ thống chuyên gia hoạt động dốc để lên lịch và điều phối tại sân bay".

Trong: *AI Magazine* 21.4, pp. 75–82.

Kahneman, Daniel (2011). *Suy nghĩ, nhanh và chậm*. New York: Farrar, Straus và Giroux.

Kalyanpur, A. và cộng sự (2012). "Dữ liệu có cấu trúc và suy luận trong DeepQA". Trong: *Ibm Journal of*

*Nghiên cứu và phát triển* 56.3.4.

Kurzweil, Ray (2000). *Thời đại của những cỗ máy tâm linh: Khi máy tính vượt quá trí thông minh của con người*. Chim cánh cụt.

Laird, J. E., C. Lebiere, và P. S. Rosenbloom (2017). "Một mô hình tiêu chuẩn của tâm trí: Hướng tới một khuôn khổ tính toán chung trên thông minh nhân tạo, khoa học nhận thức, khoa học thần kinh và robot". Trong: *AI Magazine* 38.4, pp. 13–26.

Legg, Shane và Marcus Hutter (2006). "Một thước đo chính thức của trí thông minh máy móc". Trong: *Proc. 15th*

*Hội nghị máy học hàng năm của Bỉ và Hà Lan (Benelearn'06)*. Gent. Maes, Pattie (1995). "Cuộc sống nhân tạo đáp ứng giải trí: Cuộc sống như những đặc vụ tự trị". Trong:

*Truyền thông của ACM* 38.11, các trang 108–114.

Marcus, Gary (2020). *Thập kỷ tiếp theo trong LĨNH VỰC AI: Bốn bước tiến tới trí tuệ nhân tạo mạnh mẽ*. arXiv: [2002.06177.](https://arxiv.org/abs/2002.06177)

Meystel, Alexander M. và James S. Albus (2002). *Hệ thống thông minh: kiến trúc, thiết kế và điều khiển*. New York: Wiley.

Mitchell, Tom (1năm 997). *Học máy*. Đồi McGraw.

Molina, Martin (2005). "Một trợ lý thông minh cho quản lý giao thông công cộng". Trong: *Hội nghị quốc tế về điện toán thông minh*. Berlin, Heidelberg: Springer, các trang 199–208.

Molina, Martin và Gemma Blasco (2003). "Một hệ thống đa tác nhân để hỗ trợ quyết định khẩn cấp".

Trong: *Hội nghị quốc tế về kỹ thuật dữ liệu thông minh và học tập tự động*.

Molina, Martin và Pablo Santamaria (2021). "Phối hợp hành vi cho robot tự thích ứng bằng cách sử dụng cấu hình dựa trên ràng buộc". Trong: *arXiv preprint arXiv:2103.13128*.

Moravec, Hans (1988). *Tâm trí trẻ em: Tương lai của robot và trí thông minh của con người*. Nhà in Đại học Harvard.

Morin, Alain (2011). "Tự nhận thức phần 1: Định nghĩa, biện pháp, tác dụng, chức năng và tiền đề".

Trong: *La bàn tâm lý học xã hội và nhân cách* 5.10, trang 807–823.

Muller, Jorg P. (1996). *Thiết kế của các tác nhân thông minh: Một cách tiếp cận* theo lớp. Springer.

Murphy, Robin R. (2019). *Giới thiệu về robot AI*. Báo chí MIT.

Newell, Allen (1982). "Trình độ kiến thức". Trong: *Trí tuệ nhân tạo* 18.1, trang 87–127.

Newell, Allen và Herbert A. Simon (1976). "Khoa học máy tính như một cuộc điều tra thực nghiệm: Biểu tượng và tìm kiếm". Trong: *ACM Turing giải thưởng bài giảng*.

Oreizy, P. và cộng sự (1999). "Một cách tiếp cận dựa trên kiến trúc cho phần mềm tự thích ứng". Trong: *IEEE*

*Hệ thống thông minh vàứng dụng eir*  thứ 14.3, trang 54–62.

Reggia, James A. (2014). "Máy móc có ý thức: Quan điểm AI". Trong: *Aaai Fall Symposium Series*. Arlington, Virginia, Hoa Kỳ.

Russell, Stuart (2019). *Tương thích với con người: Trí tuệ nhân tạo và vấn đề kiểm soát*. Người Viking.

Russell, Stuart và Peter Norvig (2014). *Trí tuệ nhân tạo: Một cách tiếp cận hiện đại (phiên bản thứ 3)*. Giáo dục Pearson.

Schacter, Daniel L. và cộng sự (2020). *Tâm lý học*. Nhà xuất bản đáng giá.

Bạc, David và cộng sự (2021). "Phần thưởng là đủ". Trong: *Trí tuệ nhân tạo* 299, tr. 103535. Wooldridge, M. J. (2009). *Giới thiệu về hệ thống đa tác vụ (Xuất bản lần thứ hai)*. John Wiley và các con trai.

Wooldridge, MJ và N. R. Jennings (1995). "Agents thông minh: Lý thuyết và thực hành". Trong: *Đánh giá kỹ thuật tri thức* 10.2, trang 115–152.

Xian, Yongqin và cộng sự (2018). "Học tập không bắn — một đánh giá toàn diện về điều tốt, điều xấu và điều xấu". Trong: *Giao dịch IEEE về phân tích mẫu vàigence*  intell máy 41.9, trang 2251–2265.

1. Ý nghĩa thiết yếu của từ *người đại lý* có thể được biểu thị là "một người hành động", dựa trên nguồn gốc của nó từ từ tiếng Latin *agens*. [↑](#footnote-ref-1)
2. Một số tác giả sử dụng thuật ngữ này *Phản ứng* ngược lại với chủ động. Tuy nhiên, trong bài viết này, chúng tôi sẽ sử dụng thuật ngữ phản ứng ngược lại với cân nhắc vì nó là khái quáty được sử dụng trong các hệ thống dựa trên tác nhân. [↑](#footnote-ref-2)
3. Loại nhận thức này được gọi là *Proprioception* (nhận thức về sự tự vận động và vị trí cơ thể). [↑](#footnote-ref-3)
4. Hành vi của một số loại hệ thống thông minh nhất định có thể được hiểu là sự kết hợp của nhiều bộ điều khiển hành vi như nó đã được mô tả bởi mô hình dựa trên hành vi trong robot (Arkin 1998) (Brooks 1991). [↑](#footnote-ref-4)
5. Nó nói rằng các đại lý là *Situated* trong một môi trường khi tác nhân hoạt động trong một tương tác liên tục kết hợp chặt chẽ với môi trường trong một chuỗi hành động cảm giác liên tục. Một hệ thống được hình thành theo cách này có thể sử dụng một bộ nhớ hạn chế hoặc không tồn tại về thế giới. [↑](#footnote-ref-5)
6. Trong tâm lý học, tự awareness được hiểu là khả năng trở thành đối tượng chú ý của một người (Duval và Wicklund 1972). Nhận thức về bản thân cũng có thể được hiểu theo quan điểm xử lý thông tin về bản thân (Morin 2011) có thể thích hợp để mô phỏng abi nàylity với các hệ thống dựa trên máy tính. [↑](#footnote-ref-6)
7. Vấn đề kiểm soát AI là vấn đề làm thế nào để xây dựng các hệ thống AI hỗ trợ thay vì gây hại cho người tạo ra chúng. [↑](#footnote-ref-7)