

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC MỎ - ĐỊA CHẤT

BÁO CÁO SEMINAR

TÊN ĐỀ TÀI NỀN TẢNG CỦA TRÍ TUỆ NHÂN TẠO

Người thực hiện: PGS. TS. Lê Văn Hưng (Mã cán bộ: 0801-01)

Đơn vị: Bộ môn Công nghệ phần mềm
Khoa Công nghệ Thông tin

Hà Nội - 2023

MỤC LỤC

DANH MỤC CÁC HÌNH VẼ	2
1. Đặt vấn đề	3
2. Triết học	3
3. Toán học.....	7
4. Kinh tế.....	10
5. Thần kinh học.....	12
6. Tâm lý học.....	15
7. TÀI LIỆU THAM KHẢO.....	16

DANH MỤC CÁC HÌNH VẼ

Hình 5-1. So sánh giữa máy tính và bộ não con người.....	15
--	----

1. Đặt vấn đề

Mục đích của chuyên đề này là trình bày một bức tranh tổng quan về những lĩnh vực nghiên cứu đã góp phần với những ý tưởng, quan điểm và kỹ thuật trong việc phát triển Trí tuệ Nhân tạo (AI). Nội dung của chuyên đề được trình bày thông qua các câu hỏi chính và phân tích câu trả lời từ các ngành khoa học có liên quan.

2. Triết học

Triết học giúp trả lời các câu hỏi sau:

Có thể sử dụng các luật suy diễn hình thức để rút ra các kết luận đúng đắn không?

Trí tuệ phát sinh từ bộ não vật lý như thế nào?

Tri thức đến từ đâu?

Tri thức dẫn đến hành động như thế nào?

Aristotle (384–322 TCN), một nhà triết học cổ đại Hy Lạp, đã có những đóng góp quan trọng trong nhiều lĩnh vực như triết học, logic, khoa học tự nhiên, chính trị và đạo đức. Là học trò của Plato và sau đó là giáo viên của Alexander Đại Đế, Aristotle nổi bật với việc phát triển hệ thống logic của mình. Ông là người đầu tiên đề xuất một hệ luật suy diễn để mô tả cách thức suy luận của con người, cho phép suy ra kết luận một cách tự động từ các tiền đề ban đầu.

Ramon Llull (khoảng 1232–1315), một nhà triết học, logic học, toán học và thần học từ Catalonia, nổi tiếng với những đóng góp của mình trong lĩnh vực logic và là người tiên phong trong việc phát triển phương pháp tính toán và suy luận logic. Ông đã tạo ra "Ars Magna" (Nghệ Thuật Vĩ Đại) vào năm 1305, một hệ thống suy luận đặc biệt, và cố gắng chế tạo một thiết bị cơ học với bánh xe giấy có thể xoay để tự động hóa quá trình suy luận, tạo ra các hoán vị khác nhau.

Vào khoảng năm 1500, Leonardo da Vinci (1452–1519), một trong những thiên tài vĩ đại nhất mọi thời đại, nổi tiếng với các tác phẩm nghệ thuật như "Mona Lisa" và "Bữa tiệc cuối cùng", đã sáng tạo ra thiết kế cho một

máy tính cơ học. Mặc dù ông không thực hiện việc chế tạo nó, nhưng những mô phỏng hiện đại đã chứng minh rằng thiết kế của ông hoàn toàn khả thi và có thể hoạt động.

Tiếp theo, vào khoảng năm 1623, Wilhelm Schickard (1592–1635), một nhà khoa học Đức đa năng, đã chế tạo thành công máy tính toán đầu tiên. Sự thành công này mở đường cho một kỷ nguyên mới trong lịch sử của máy tính, đánh dấu một bước tiến đáng kể trong công nghệ tính toán.

Blaise Pascal (1623–1662), một nhà toán học và nhà khoa học người Pháp, cũng góp phần vào lĩnh vực này với việc phát triển máy Pascaline vào năm 1642. Ông, được biết đến với những đóng góp quan trọng trong xác suất học, đã tạo ra một thiết bị cơ học có khả năng xử lý thông tin một cách tự động, mô tả rằng nó tạo ra các hiệu ứng tương tự như suy nghĩ của con người.

Gottfried Wilhelm Leibniz (1646–1716), một nhà toán học và triết học nổi tiếng, đã tiếp tục phát triển lĩnh vực này bằng cách chế tạo một thiết bị cơ học nhằm thực hiện các phép toán dựa trên các khái niệm thay vì số. Tuy nhiên, thiết bị của ông bị hạn chế về phạm vi ứng dụng.

Cuối cùng, Thomas Hobbes (1588–1679), một nhà triết học chính trị với tác phẩm nổi tiếng "Leviathan", đã đề xuất ý tưởng về một máy tính có khả năng suy nghĩ, mô tả nó như một "động vật nhân tạo". Ông so sánh các bộ phận của cơ thể với các bộ phận cơ học và khẳng định rằng quá trình suy luận tương tự như việc thực hiện các phép toán cộng và trừ trong toán học.

Tuy nhiên, việc cho rằng trí tuệ hoạt động một phần theo các quy tắc logic và số học, cùng với việc tạo ra các hệ thống vật lý để mô phỏng những quy tắc này, chỉ là một khía cạnh của vấn đề. René Descartes (1596–1650), một nhà triết học nổi tiếng, là người đầu tiên phân tích sự khác nhau giữa trí tuệ và vật chất. Ông chỉ ra rằng, quan niệm trí tuệ hoàn toàn là vật lý là không có chỗ cho "tự do ý chí". Nếu trí tuệ chỉ tuân theo các quy luật vật lý, thì nó không tự do hơn so với một tảng đá đang "quyết định" rơi xuống. Descartes ủng hộ quan điểm song hành, theo đó có một phần của trí tuệ con người – có thể hiểu là linh hồn hoặc tinh thần – tồn tại ngoài yếu tố vật chất và không

chịu sự ràng buộc của các quy luật vật lý. Trái lại, ông coi động vật không sở hữu đặc tính này và có thể được xem như những cỗ máy.

Một trường phái khác với chủ nghĩa song hành là chủ nghĩa duy vật, cho rằng trí tuệ không có gì khác ngoài hoạt động của não bộ tuân theo các quy luật vật lý. Theo đó, “tự do ý chí” chỉ đơn giản là cách mà thực thể đang phải lựa chọn nhận thức về sự xuất hiện của các lựa chọn có sẵn.

Khi nghiên cứu về khả năng xử lý tri thức của trí tuệ, vấn đề quan trọng tiếp theo là xác định nguồn gốc của tri thức. Chủ nghĩa kinh nghiệm, khởi nguồn từ Francis Bacon (1561–1626), được đặc trưng bởi quan điểm của John Locke (1632–1704) “không sự hiểu biết nào mà không xuất phát từ các giác quan”. Quan điểm này nhấn mạnh vai trò quan trọng của sự trải nghiệm và quan sát thông qua các giác quan như là nguồn gốc cơ bản của mọi tri thức.

David Hume (1711–1776), một trong những nhà triết học nổi tiếng của thế kỷ 18, trong tác phẩm "A Treatise of Human Nature" (1739), đã đưa ra một đóng góp quan trọng trong lĩnh vực triết học với nguyên lý quy nạp. Nguyên lý này cho rằng, các quy luật chung được hình thành thông qua việc quan sát và nhận diện sự xuất hiện lặp đi lặp lại mối quan hệ giữa các yếu tố.

Xây dựng trên nền tảng những đóng góp từ Ludwig Wittgenstein (1889–1951) và Bertrand Russell (1872–1970), Vienna Circle, một nhóm các nhà triết học và nhà toán học, đã phát triển chủ nghĩa Logic Tích Cực. Chủ nghĩa Logic Tích Cực là một trường phái triết học cho rằng tri thức được xây dựng trên cơ sở của những quan sát và suy luận logic từ dữ liệu quan sát.

Lý thuyết xác nhận của Rudolf Carnap (1891–1970) và Carl Hempel (1905–1997) phân tích việc tiếp nhận tri thức từ trải nghiệm bằng cách định lượng độ tin cậy của các phát biểu logic thông qua mối quan hệ của chúng với các quan sát xác nhận hoặc phủ định chúng. Cuốn sách "The Logical Structure of the World" (1928) của Carnap có thể coi là một trong những lý thuyết đầu tiên coi trí tuệ như một quá trình tính toán.

Yếu tố cuối cùng trong bức tranh triết học về trí tuệ là mối liên hệ giữa tri thức và hành động. Câu hỏi này cực kỳ quan trọng đối với Trí tuệ Nhân

tạo vì trí tuệ đòi hỏi cả suy luận và hành động. Hơn nữa, chỉ thông qua việc hiểu hành động được đánh giá như thế nào ta mới có thể hiểu được cách xây dựng các tác tử có hành động hợp lý.

Aristotle đã lập luận (trong tác phẩm "De Motu Animalium") rằng các hành động được đánh giá thông qua mối quan hệ logic giữa mục tiêu và kết quả của hành động. Ông tìm hiểu lý do tại sao đôi khi suy nghĩ dẫn tới hành động, nhưng đôi khi lại không. Theo Aristotle, trong trường hợp hành động, kết luận không phải là một đề xuất lý thuyết mà là một hành động thực tế. Ông minh họa điều này qua ví dụ: Một nhu cầu cụ thể ('Tôi cần che chở') kết hợp với một giải pháp cụ thể ('Một chiếc áo choàng là vật che chở') dẫn đến quyết định hành động ('Tôi cần và phải làm một chiếc áo choàng'). Quyết định này, 'Tôi phải làm một chiếc áo choàng', không chỉ là suy nghĩ mà thực sự là một hành động.

Trong nhiều trường hợp, việc tập trung vào các hành động để đạt được mục tiêu là phương pháp hữu ích. Tuy nhiên, không phải lúc nào cách tiếp cận này cũng hiệu quả. Đặc biệt là trong tình huống có nhiều phương án khác nhau để đạt một mục tiêu, chúng ta cần một phương pháp để lựa chọn lựa chọn tốt nhất. Hơn nữa, đôi khi không thể chắc chắn về việc liệu mục tiêu có thể đạt được hay không, nhưng vẫn cần phải đưa ra quyết định. Vậy, cơ sở để quyết định là gì? Antoine Arnauld, vào năm 1662, khi nghiên cứu về quyết định hợp lý trong cờ bạc, đã đề xuất một công thức toán học để tối đa hóa giá trị tiền kỳ vọng từ kết quả. Tiếp theo, vào năm 1738, Daniel Bernoulli đã giới thiệu khái niệm "lợi ích", một khái niệm rộng hơn để đánh giá giá trị nội tại, chủ quan của một kết quả. Trong thời đại hiện đại, quyết định hợp lý trong bối cảnh không chắc chắn dựa vào việc tối đa hóa lợi ích kỳ vọng.

Trong vấn đề đạo đức và chính sách công, người ra quyết định cần phải xem xét lợi ích của nhiều người. Jeremy Bentham vào năm 1823 và John Stuart Mill vào năm 1863 đã phát triển khái niệm chủ nghĩa công ích. Họ cho rằng, phương pháp quyết định hợp lý dựa trên nguyên tắc tối đa hóa lợi ích nên được áp dụng không chỉ trong đời sống cá nhân mà còn trong quá trình đưa ra các quyết định chính sách công, ảnh hưởng đến đông đảo mọi người. Chủ nghĩa công ích là một hình thức của chủ nghĩa hậu quả, tức là

quan điểm cho rằng một hành động đúng hay sai được đánh giá dựa trên kết quả mà nó dự kiến sẽ tạo ra.

Trái với quan điểm của chủ nghĩa công ích, Immanuel Kant vào năm 1875 đã đề xuất một lý thuyết đạo đức dựa trên các quy tắc cố định, hay còn gọi là chủ nghĩa bổn phận. Theo ông, việc 'làm điều đúng đắn' không phụ thuộc vào kết quả của hành động mà phụ thuộc vào việc tuân thủ các quy tắc xã hội chung, chẳng hạn như 'không nói dối' hoặc 'không giết người.' Vì vậy, trong khi một người theo chủ nghĩa công ích có thể chấp nhận nói dối nhẹ nếu việc đó mang lại nhiều lợi ích hơn hại, một người theo triết lý của Kant sẽ kiên quyết không làm như vậy, vì nói dối theo quan điểm của họ là sai trái không thể biện minh. Dù Mill thừa nhận tầm quan trọng của các quy tắc, ông xem chúng như là các phương pháp quyết định hiệu quả, được hình thành từ việc suy luận logic dựa trên nguyên lý và hậu quả. Phương pháp này hiện được nhiều hệ thống Trí tuệ Nhân tạo hiện đại áp dụng.

3. Toán học

Toán học giúp trả lời các câu hỏi sau:

Những luật suy diễn nào giúp rút ra các kết luận đúng đắn?

Những gì có thể tính toán được?

Làm thế nào để suy luận được với thông tin không chắc chắn?

Những tư tưởng ban đầu về Trí tuệ Nhân tạo đã được các nhà triết học xác lập, nhưng để biến nó thành một ngành khoa học chính thức, điều cần thiết là việc chuyển đổi logic và xác suất thành ngôn ngữ toán học và tạo ra một lĩnh vực toán học mới: lĩnh vực tính toán. Quá trình này đòi hỏi việc áp dụng các phương pháp toán học để cấu trúc hóa và hệ thống hóa logic và lý thuyết xác suất, từ đó tạo nền móng cho việc phát triển các thuật toán và mô hình máy tính có khả năng thực hiện các tác vụ thông minh, điều này chính là cốt lõi của Trí tuệ Nhân tạo.

Khái niệm về logic hình thức xuất phát từ ý tưởng của các nhà triết học cổ đại tại Hy Lạp, Ấn Độ và Trung Quốc. Tuy nhiên, sự phát triển theo hướng toán học đáng chú ý của logic này bắt đầu với George Boole (1815–1864). Boole đã đưa ra và phát triển logic mệnh đề, còn được gọi là logic Boolean,

vào năm 1847. Sau đó, vào năm 1879, Gottlob Frege (1848–1925) đã mở rộng logic của Boole, bổ sung các khái niệm về đối tượng và quan hệ, để tạo nên logic vị từ cấp 1. Logic này ngày nay vẫn là công cụ không thể thiếu trong lĩnh vực Trí tuệ Nhân tạo. Ngoài ra, logic vị từ cấp 1 còn truyền cảm hứng cho các công trình nghiên cứu của Gödel và Turing, là nền tảng quan trọng cho lý thuyết tính toán.

Lý thuyết xác suất, rất quan trọng trong lĩnh vực Trí tuệ Nhân tạo, có thể được xem như một sự mở rộng của logic để áp dụng trong các tình huống có thông tin không chắc chắn. Khái niệm này được Gerolamo Cardano (1501–1576) đề xuất lần đầu, khi ông phân tích xác suất dựa trên các kết quả có thể xảy ra trong đánh bạc. Blaise Pascal (1623–1662) sau đó, vào năm 1654, trong một bức thư gửi đến Pierre Fermat (1601–1665), đã phát triển ý tưởng này bằng cách dự đoán kết quả của một trò chơi đánh bạc. Xác suất trở thành một công cụ không thể thiếu trong các ngành khoa học định lượng, giúp xử lý các giá trị đo không chắc chắn và thiếu thông tin. Các nhà khoa học như Jacob Bernoulli (1654–1705, chú của Daniel), Pierre Laplace (1749–1827) đã đóng góp vào việc phát triển lý thuyết này và giới thiệu các phương pháp thống kê mới. Thomas Bayes (1702–1761) đưa ra quy tắc Bayes, một quy tắc cập nhật xác suất dựa trên bằng chứng mới, trở thành một công cụ cốt lõi trong các hệ thống Trí tuệ Nhân tạo.

Việc hình thành và phát triển lý thuyết xác suất, cùng với việc dữ liệu trở nên dễ tiếp cận hơn, đã mở đường cho sự ra đời và phát triển của ngành thống kê. Một trong những ví dụ sớm nhất về ứng dụng thống kê là công trình phân tích dữ liệu điều tra dân số của London do John Graunt thực hiện vào năm 1662. Ronald Fisher, người xuất bản công trình vào năm 1922, được xem là người đặt nền móng cho ngành thống kê hiện đại. Ông đã tổng hợp và phát triển các khái niệm từ xác suất, thiết kế thí nghiệm, phân tích dữ liệu, đến việc sử dụng máy tính. Đáng chú ý, vào năm 1919, Fisher nhấn mạnh rằng ông không thể thực hiện công việc của mình mà không có máy tính cơ học MILLIONAIRE - máy tính đầu tiên có thể thực hiện phép nhân - mặc dù giá của máy này cao hơn lương hàng năm của ông.

Lĩnh vực tính toán có bề dày lịch sử ngang bằng với lịch sử của việc sử dụng số lượng. Thuật toán có ý nghĩa quan trọng đầu tiên được ghi nhận là thuật toán của Euclid, được dùng để tính ước số chung lớn nhất. Từ 'algorithm' có nguồn gốc từ tên của Muhammad ibn Musa al-Khwarizmi, một nhà toán học ở thế kỷ thứ 9, người đã giới thiệu hệ thống số Ả Rập và đại số tới châu Âu. George Boole và các nhà toán học khác đã phát triển các thuật toán cho suy luận logic. Vào cuối thế kỷ 19, đã có những nỗ lực để hình thức hóa quá trình suy luận toán học, biến nó thành một phần của suy luận logic.

Kurt Gödel, một nhà toán học nổi tiếng, đã làm sáng tỏ một khía cạnh quan trọng trong logic vị từ cấp 1. Ông chứng minh rằng, mặc dù có phương pháp hiệu quả để chứng minh bất kỳ phát biểu đúng nào trong logic này, nó lại không thể biểu diễn đầy đủ nguyên lý quy nạp toán học, một nguyên lý cốt lõi để mô tả các số tự nhiên. Điều này dẫn đến phát hiện của ông vào năm 1931 về những hạn chế trong suy diễn. Định lý không hoàn chỉnh của Gödel khẳng định rằng trong bất kỳ một hệ hình thức nào có sức mạnh tương đương với số học Peano - hệ lý thuyết cơ bản về số tự nhiên - sẽ luôn có những phát biểu đúng không thể chứng minh được bên trong chính hệ đó.

Phát hiện quan trọng này cũng chỉ ra rằng có những hàm toán học áp dụng trên số nguyên không thể định nghĩa hoặc thực hiện được bằng một thuật toán, tức là chúng không thể tính toán được. Sự nhận thức này đã thúc đẩy Alan Turing (1912–1954) tìm cách xác định rõ ràng các hàm nào là có thể tính toán được, hay có thể thực hiện thông qua một quy trình hiệu quả. Luận đề Church-Turing, đề xuất vào năm 1936, gợi ý rằng khái niệm tính toán chung nên được hiểu là các hàm có thể xử lý được bởi máy Turing. Ngoài ra, Turing cũng chứng minh rằng có những hàm mà không máy Turing nào có thể tính toán được. Một ví dụ điển hình là không có thuật toán nào có thể tổng quát hóa việc xác định liệu một chương trình máy tính nào đó sẽ trả về kết quả cho một đầu vào xác định hay sẽ chạy vô hạn.

Dù khái niệm về tính toán được là quan trọng trong tính toán, nhưng khái niệm tính khả thi lại có ảnh hưởng lớn hơn tới Trí tuệ Nhân tạo. Một vấn đề được xem là khó giải quyết khi thời gian cần thiết để giải quyết nó tăng theo cấp số mũ của kích thước. Sự khác nhau giữa độ phức tạp theo đa

thức và theo cấp số mũ đã được nhấn mạnh lần đầu tiên vào giữa những năm 1960, qua công trình của Cobham và Edmonds. Điều này rất quan trọng bởi vì với độ phức tạp theo cấp số mũ thì ngay cả những vấn đề có kích thước vừa phải cũng không thể xử lý được trong thời gian hợp lý.

Phát triển bởi Cook vào năm 1971 và Karp vào năm 1972, lý thuyết NP-đầy đủ đã thiết lập một nền tảng để đánh giá tính khả thi của các bài toán. Theo lý thuyết này, một bài toán có thể được coi là không khả thi nếu nó có thể chuyển đổi về dạng của một bài toán NP-đầy đủ. Dù chưa có bằng chứng cụ thể cho thấy tất cả các bài toán NP-đầy đủ đều không khả thi, nhưng phần lớn các nhà lý thuyết tin rằng điều này là đúng. Nhận thức này đã tạo nên một sự đối lập mạnh mẽ so với quan điểm lạc quan ban đầu của giới truyền thông đối với máy tính, khi chúng được ví von như là 'Não Điện Tử Siêu Việt' và được ca ngợi là 'Nhanh hơn cả Einstein'. Tuy nhiên, ngay cả khi tốc độ máy tính ngày càng tăng, việc sử dụng nguồn lực một cách thông minh và chấp nhận những hạn chế là điều cần thiết cho các hệ thống thông minh.

4. Kinh tế

Kinh tế học trả lời các câu hỏi sau:

Chúng ta nên đưa ra quyết định như thế nào để phù hợp với sở thích của mình?

Chúng ta nên làm như thế nào khi người khác có thể không đồng tình?

Chúng ta nên làm như thế nào khi lợi ích có thể chỉ đến sau một thời gian dài?

Ngành khoa học kinh tế ra đời vào năm 1776, đánh dấu bởi sự ra đời của tác phẩm 'Nghiên cứu về Bản chất và Nguyên nhân của Sự giàu có của các Quốc gia' của Adam Smith (1723–1790). Trong tác phẩm này, Smith đã phân tích nền kinh tế dưới góc độ các cá nhân, mỗi người tập trung vào lợi ích riêng của mình. Tuy nhiên, Smith không coi lòng tham về mặt tài chính là quan điểm đạo đức đúng đắn. Trong cuốn sách của ông viết trước đó, 'Lý thuyết về Tình cảm Đạo đức' phát hành vào năm 1759, ông bắt đầu nhấn mạnh rằng việc quan tâm đến sự phúc lợi của người khác là một phần không thể thiếu trong lợi ích của mỗi cá nhân.

Phần lớn mọi người thường liên kết kinh tế học với tiền. Thực vậy, phân tích toán học đầu tiên về việc đưa ra quyết định trong tình huống không chắc chắn, đó là công thức giá trị kỳ vọng tối đa của Arnauld vào năm 1662, chủ yếu tập trung vào giá trị tiền trong các cuộc đánh cược. Tuy nhiên, Daniel Bernoulli vào năm 1738 đã phát hiện ra rằng công thức này không phù hợp với các khoản tiền lớn, chẳng hạn như trong việc đầu tư vào các chuyến thương mại hàng hải. Thay vì theo công thức này, ông đã đề xuất một nguyên tắc dựa vào việc tối đa hóa lợi ích kỳ vọng. Ông giải thích rằng sự lựa chọn đầu tư của con người có thể được hiểu thông qua việc nhận định rằng giá trị lợi ích cận biên của mỗi đồng tiền bổ sung sẽ giảm dần khi một người có được nhiều tiền hơn.

Léon Walras (1834 - 1910) đã mở rộng lý thuyết lợi ích, không chỉ dựa trên giá trị tiền tệ, mà còn dựa trên sở thích giữa các lựa chọn đánh cược với mọi loại kết quả. Sự phát triển này sau đó được Ramsey nâng cấp vào năm 1931, và được phát triển thêm bởi John von Neumann và Oskar Morgenstern trong tác phẩm 'Lý thuyết Trò chơi và Hành vi Kinh tế' xuất bản vào năm 1944. Nhờ những đóng góp này, kinh tế học không còn giới hạn ở việc nghiên cứu về tiền bạc, mà đã trở thành ngành học nghiên cứu về những mong muốn và sở thích của con người.

Lý thuyết quyết định, một sự kết hợp của lý thuyết xác suất và lý thuyết lợi ích, tạo ra một khung cảnh toàn diện và hệ thống cho việc ra quyết định của cá nhân, dù là trong lĩnh vực kinh tế hay không, khi đối mặt với tình huống không chắc chắn. Trong những tình huống này, việc sử dụng mô tả xác suất phản ánh đúng môi trường mà người ra quyết định phải đối mặt. Đối với các nền kinh tế lớn, mỗi người tham gia không cần phải quan tâm tới hành động của người khác. Tuy nhiên, trong các nền kinh tế nhỏ, tình hình lại giống như một trò chơi, nơi mà hành động của một người chơi có thể ảnh hưởng lớn đến lợi ích của người chơi khác, cả tích cực và tiêu cực. Sự phát triển của lý thuyết trò chơi bởi Von Neumann và Morgenstern, cũng như công trình của Luce và Raiffa vào năm 1957, đã phát hiện ra rằng, trong một số trường hợp, người chơi hợp lý nên theo đuổi các chiến lược dường như ngẫu nhiên. Trái với lý thuyết quyết định, lý thuyết trò chơi không chỉ ra

cách chọn lựa hành động một cách rõ ràng. Trong Trí tuệ Nhân tạo, quyết định liên quan đến nhiều người tham gia được nghiên cứu dưới khía cạnh của hệ thống đa tác tử.

Các nhà kinh tế học, trừ một số trường hợp đặc biệt, thường không tập trung vào vấn đề làm thế nào để ra quyết định hợp lý khi lợi ích từ các hành động không đến ngay lập tức mà phát sinh từ chuỗi các hành động liên tiếp. Vấn đề này được khám phá sâu hơn trong lĩnh vực nghiên cứu vận hành, ngành học phát triển trong Thế chiến II từ nhu cầu tối ưu hóa hệ thống radar ở Anh và sau đó được ứng dụng rộng rãi trong đời sống dân sự. Công trình của Richard Bellman vào năm 1957 đã hình thức hóa một loạt các bài toán ra quyết định tuần tự, được biết đến với tên gọi là quá trình ra quyết định Markov.

Công trình trong lĩnh vực kinh tế và nghiên cứu vận hành đã có những đóng góp đáng kể trong việc hình thành khái niệm về tác tử hợp lý. Tuy nhiên, nghiên cứu về Trí tuệ Nhân tạo đã phát triển theo một hướng hoàn toàn khác biệt trong nhiều năm qua. Một trong những nguyên nhân là sự phức tạp trong việc đưa ra quyết định hợp lý. Herbert Simon, một nhà nghiên cứu tiên phong về Trí tuệ Nhân tạo, đã nhận giải Nobel Kinh tế vào năm 1978 nhờ công trình nghiên cứu của ông. Ông đã chỉ ra rằng việc áp dụng mô hình 'satisficing', tức là đưa ra những quyết định 'đủ tốt' thay vì cố gắng tìm ra quyết định tối ưu, mô tả chính xác hơn hành vi thực của con người (Simon, 1947). Từ những năm 1990 trở lại đây, đã có sự gia tăng mức độ quan tâm đến việc áp dụng các phương pháp dựa trên lý thuyết ra quyết định trong ngành Trí tuệ Nhân tạo.

5. Thần kinh học

Thần kinh học giúp trả lời câu hỏi sau:

Não bộ xử lý thông tin như thế nào?

Neuroscience, ngành khoa học nghiên cứu về hệ thần kinh và đặc biệt là não bộ, chủ yếu tập trung vào cách não hoạt động. Dù cách thức cụ thể mà não tạo ra suy nghĩ vẫn là một bí ẩn lớn, người ta đã nhận thức từ lâu rằng não là nơi tạo ra suy nghĩ. Điều này được chứng minh qua hậu quả của những

chấn thương nặng ở đầu dẫn đến suy giảm trí tuệ. Ngoài ra, từ thời cổ đại, người ta đã biết rằng não người có những đặc điểm riêng biệt; theo Aristotle (335 TCN), con người có bộ não lớn nhất so với kích thước cơ thể. Tuy nhiên, phải đến giữa thế kỷ 18, não mới được công nhận rộng rãi là trung tâm của ý thức, thay vì các bộ phận khác như tim hay lá lách.

Công trình nghiên cứu của Paul Broca vào năm 1861 về chứng rối loạn ngôn ngữ (aphasia) ở những bệnh nhân bị tổn thương não đã mở đường cho việc khám phá tổ chức chức năng của não. Ông đã xác định một khu vực cụ thể ở bán cầu trái của não, sau này được gọi là vùng Broca, chịu trách nhiệm cho việc sản sinh ngôn ngữ. Đến thời điểm đó, người ta đã biết rằng não bộ chủ yếu được cấu thành từ các tế bào thần kinh, hay nơ-ron. Tuy nhiên, phải đến năm 1873, Camillo Golgi mới phát triển một phương pháp nhuộm đặc biệt giúp quan sát từng nơ-ron riêng biệt. Phương pháp này sau đó được Santiago Ramon y Cajal sử dụng trong các nghiên cứu tiên phong của ông về cấu trúc nơ-ron. Ngày nay, người ta hiểu rằng các chức năng nhận thức xuất phát từ hoạt động điện hóa của những cấu trúc này và hệ thống các tế bào đơn giản có thể tạo nên suy nghĩ, hành động và trí tuệ. Như John Searle đã nói vào năm 1992, 'não bộ tạo ra trí tuệ'.

Ngày nay, người ta đã thu thập được dữ liệu về mối liên hệ giữa các vùng não với các bộ phận cơ thể mà chúng điều khiển hoặc nhận thông tin cảm giác. Điều đáng chú ý là những mối liên hệ này có thể thay đổi đáng kể chỉ trong vài tuần. Thêm vào đó, một số loài động vật có vẻ như sở hữu nhiều hệ thống bản đồ não khác nhau. Hiện vẫn còn nhiều hạn chế trong việc hiểu rõ cách thức mà các khu vực khác của não có thể tiếp quản các chức năng từ những phần não bị tổn thương. Hơn nữa, hiện vẫn chưa có lý thuyết rõ ràng nào về cách một ký ức cá nhân được lưu trữ trong não, hay về cách thức hoạt động của các chức năng nhận thức ở cấp độ cao hơn.

Kể từ năm 1929, với phát minh của Hans Berger về máy điện não đồ (EEG), người ta đã bắt đầu có thể đo hoạt động của bộ não còn nguyên vẹn. Tiếp theo đó, với sự phát triển của kỹ thuật chụp cộng hưởng từ chức năng (fMRI), bắt đầu từ năm 1990 bởi Ogawa và các cộng sự, cũng như Cabeza và Nyberg vào năm 2001, đã mở ra cơ hội cho các nhà nghiên cứu thần kinh

quan sát não bộ với độ chi tiết chưa từng có, qua đó cho phép đo lường một cách chính xác các quá trình nhận thức đang diễn ra. Ngoài ra, những bước tiến trong việc ghi chép xung điện từng nơ-ron đơn lẻ và sự phát triển của kỹ thuật optogenetics – bởi Crick và các đồng nghiệp từ năm 1999 đến năm 2007 - đã mở ra khả năng không chỉ đo lường mà còn kiểm soát các nơ-ron đơn lẻ được thay đổi để nhạy sáng.

Sự phát triển của các giao diện não-máy (Lebedev và Nicolelis, 2006) không chỉ cho việc cảm nhận và kiểm soát vận động hứa hẹn khôi phục chức năng cho những cá nhân khuyết tật, mà còn mang lại cái nhìn sâu sắc về nhiều khía cạnh của hệ thống thần kinh. Một phát hiện đáng chú ý từ công trình này là não bộ có khả năng tự điều chỉnh để kết nối thành công với thiết bị bên ngoài, coi nó như một cơ quan cảm giác hoặc chi khác.

Sự tiến bộ trong việc phát triển các giao diện kết nối giữa não bộ và máy móc, được nghiên cứu bởi Lebedev và Nicolelis vào năm 2006, không chỉ mang lại cơ hội phục hồi chức năng cho những người khuyết tật thông qua việc cảm nhận và kiểm soát chuyển động, mà còn mở ra hiểu biết mới về hoạt động của hệ thống thần kinh. Phát hiện đặc biệt từ nghiên cứu này là não bộ có khả năng tự điều chỉnh để phối hợp hiệu quả với thiết bị ngoại vi, coi chúng như một giác quan hoặc chi của cơ thể.

Não bộ và máy tính số mang những đặc tính riêng biệt (Hình 5-1. So sánh giữa máy tính và bộ não con người). Máy tính có khả năng xử lý dữ liệu nhanh gấp một triệu lần so với não bộ. Tuy nhiên, não bộ lại vượt trội về dung lượng lưu trữ và mạng lưới kết nối giữa các tế bào thần kinh, hơn hẳn so với máy tính cá nhân hiện đại, mặc dù các siêu máy tính lớn nhất cũng có thể tương đương với não bộ theo một số tiêu chí. Các nhà tương lai học thường chú ý nhiều đến những con số này, dự báo về thời điểm quan trọng khi máy tính đạt tới mức hiệu suất vượt trội so với con người, rồi sau đó liên tục tự nâng cấp bản thân. Tuy nhiên, việc so sánh chỉ dựa trên số liệu không

mang lại cái nhìn đầy đủ. Ngay cả khi có một máy tính với khả năng không giới hạn, chúng ta vẫn cần những bước tiến mới để hiểu rõ về trí tuệ.

6. Tâm lý học

	Supercomputer	Personal Computer	Human Brain
Computational units	10^6 GPUs + CPUs	8 CPU cores	10^6 columns
	10^{15} transistors	10^{10} transistors	10^{11} neurons
Storage units	10^{16} bytes RAM	10^{10} bytes RAM	10^{11} neurons
	10^{17} bytes disk	10^{12} bytes disk	10^{14} synapses
Cycle time	10^{-9} sec	10^{-9} sec	10^{-3} sec
Operations/sec	10^{18}	10^{10}	10^{17}

Hình 5-1. So sánh giữa máy tính và bộ não con người

Tâm lý học giúp trả lời câu hỏi sau:

Con người và động vật suy nghĩ và hành động như thế nào?

Tâm lý học được coi là khởi nguồn từ công trình của Hermann von Helmholtz, một nhà vật lý người Đức và học trò của ông là Wilhelm Wundt. Helmholtz đã tiên phong trong việc áp dụng phương pháp khoa học để nghiên cứu thị giác con người và tác phẩm 'Handbook of Physiological Optics' của ông được coi là một trong những công trình quan trọng nhất về lĩnh vực này. Năm 1879, Wundt đã thành lập phòng thí nghiệm tâm lý học thực nghiệm đầu tiên tại Đại học Leipzig, nơi ông thực hiện các thí nghiệm được kiểm soát chặt chẽ. Trong những thí nghiệm này, các nhà nghiên cứu thực hiện các tác vụ liên quan đến cảm nhận và liên tưởng, đồng thời phân tích quá trình suy nghĩ của họ. Cách làm này đã góp phần định hình tâm lý học như một ngành khoa học, mặc dù yếu tố chủ quan trong dữ liệu nghiên cứu có thể hạn chế khả năng phản bác các giả thuyết của các nhà nghiên cứu.

Các nhà sinh học nghiên cứu hành vi của động vật lại không có dữ liệu về tự nhận thức và do đó đã xây dựng một phương pháp luận khách quan, như H. S. Jennings mô tả trong công trình nổi tiếng của ông vào năm 1906, 'Behavior of the Lower Organisms'. Khi áp dụng quan điểm này cho con người, trào lưu hành vi học do John Watson dẫn đầu bác bỏ bất kỳ lý thuyết nào liên quan đến các quá trình hoạt động nội tại của não bộ, bởi vì họ cho rằng tự nhận thức không thể mang lại bằng chứng đáng tin cậy. Các nhà hành vi học chỉ tập trung vào việc nghiên cứu các chỉ số khách quan của cảm giác

(hoặc kích thích) đối với động vật và hành động (hoặc phản ứng) kết quả. Hành vi học đã mang lại nhiều hiểu biết về chuột và chim bồ câu nhưng lại ít hiệu quả hơn trong nghiên cứu về con người.

7. TÀI LIỆU THAM KHẢO

Stuart Russell and Peter Norvig, Artificial Intelligence: A Modern Approach 4th Edition. Pearson, 2021.