TCVN **TIÊU CHUẨN QUỐC GIA**

TCVN 13902:2023
ISO/IEC 22989:2022

Xuất bản lần 1

CÔNG NGHỆ THÔNG TIN –TRÍ TUỆ NHÂN TẠO – CÁC KHÁI NIỆM VÀ THUẬT NGỮ TRÍ TUỆ NHÂN TẠO

Information technology – Artificial intelligence – Artificial intelligence concepts and terminology

HÀ NỘI – 2023

Mục lục

Lời nói đầu

TCVN 13902:2023:2023 hoàn toàn tương đương với ISO/IEC 22989:2022.

TCVN 13902:2023:2023 do Viện Công nghiệp Phần mềm và Nội dung số Việt Nam biên soạn, Bộ Thông tin và Truyền thông đề nghị, Tổng cục Tiêu chuẩn Đo lường Chất lượng thẩm định, Bộ Khoa học và Công nghệ công bố.

Công nghệ thông tin – Trí tuệ nhân tạo – Các khái niệm và thuật ngữ trí tuệ nhân tạo

Information technology – Artificial intelligence – Artificial Intelligence Concepts and Terminology

1 Phạm vi áp dụng

Tiêu chuẩn này đưa ra các thuật ngữ và mô tả các khái niệm trong lĩnh vực AI.

Tiêu chuẩn này có thể được sử dụng để tham chiếu trong việc xây dựng và công bố các tiêu chuẩn khác và trong việc hỗ trợ trao đổi thông tin giữa các bên hoặc các đối tác liên quan

Tiêu chuẩn này áp dụng cho tất cả các cá nhân, tổ chức (ví dụ cơ quan chính phủ, tổ chức, doanh nghiệp).

2 Tài liệu viện dẫn

Tiêu chuẩn này không có tài liệu viện dẫn.

3 Các thuật ngữ và định nghĩa

Tiêu chuẩn này sử dụng các thuật ngữ và định nghĩa dưới đây.

ISO và IEC duy trì cơ sở dữ liệu thuật ngữ sử dụng trong hoạt động tiêu chuẩn hóa tại các địa chỉ dưới đây:

- Nền tảng trình duyệt trực tuyến của ISO: tại địa chỉ https://www.iso.org/obp
- Từ vựng kỹ thuật điện của IEC: tại địa chỉ https://www.electropedia.org/

3.1 Các thuật ngữ liên quan đến AI

3.1.1

Tác tử AI (AI agent)

Thực thể *tự động* (3.1.7) cảm nhận và phản ứng với môi trường của nó và thực hiện các hành động để đạt được mục tiêu của nó.

3.1.2

Phần tử AI (AI component)

Phần tử chức năng để xây dựng *hệ thống AI* (3.1.4)

3.1.3

Trí tuệ nhân tạo (artificial intelligence)

AI

Nghiên cứu và phát triển các cơ chế và ứng dụng của *hệ thống AI* (3.1.4)

CHÚ THÍCH 1: Nghiên cứu và phát triển có thể diễn ra trên bất kỳ lĩnh vực nào, chẳng hạn như khoa học máy tính, khoa học dữ liệu, khoa học tự nhiên, nhân văn, toán học.

3.1.4

Hệ thống trí tuệ nhân tạo (artificial intelligence system)

Hệ thống AI (AI system)

Hệ thống được thiết kế tạo ra các kết quả đầu ra như nội dung, dự báo, khuyến nghị hoặc quyết định cho một tập hợp các mục tiêu xác định bởi con người

CHÚ THÍCH 1: Hệ thống được thiết kế có thể sử dụng các kỹ thuật và phương pháp tiếp cận khác nhau liên quan đến *trí tuệ nhân tạo* (3.1.3) để phát triển *mô hình* (3.1.23) biểu diễn dữ liệu, *tri thức* (3.1.21), các quá trình v.v. được sử dụng để tiến hành các *tác vụ* (3.1.35).

CHÚ THÍCH 2: Hệ thống AI được thiết kế để hoạt động với các mức độ *tự động hóa* (3.1.7) khác nhau.

3.1.5

Tự trị (autonomy)

Tính tự trị (autonomous)

Đặc tính của một hệ thống có khả năng sửa đổi phạm vi sử dụng hoặc mục tiêu dự kiến của nó mà không cần sự can thiệp, kiểm soát hoặc giám sát từ bên ngoài.

3.1.6

Mạch tích hợp được thiết kế cho các ứng dụng cụ thể (application specific integrated circuit)

ASIC

Mạch tích hợp được thiết kế cho các ứng dụng cụ thể.

[nguồn: ISO/IEC/IEEE 24765:2017, 3.193 được sửa đổi – Từ viết tắt đã được chuyển sang dòng riêng].

3.1.7

Tự động (automatic)

Tự động hóa (automation)

Được tự động hóa (automated)

Liên quan đến một quá trình hoặc hệ thống hoạt động trong các điều kiện cụ thể mà không cần sự can thiệp của con người.

[nguồn: ISO/IEC 2382:2015, 2121282 được sửa đổi - Định nghĩa *"một quá trình hoặc thiết bị"* được thay thế bằng "một quy trình hoặc hệ thống" và bổ sung các thuật ngữ "tự động hóa", "được tự động hóa" vào tiêu chuẩn này].

3.1.8

Điện toán nhận thức (cognitive computing)

Loại hình *hệ thống AI* (3.1.4) cho phép người và máy móc tương tác tự nhiên hơn.

CHÚ THÍCH 1: Các tác vụ điện toán nhận thức liên quan tới *học máy* (3.3.5), xử lý tiếng nói, *xử lý ngôn ngữ tự nhiên* (3.6.9), *thị giác máy tính* (3.7.1) và giao diện người-máy.

3.1.9

Học liên tục (continuous learning)

Học không ngừng (continual learning)

Học suốt đời (lifelong learning)

Học tăng cường trên một *hệ thống AI* (3.1.4) diễn ra mang tính liên tục trong giai đoạn vận hành của vòng đời hệ thống AI đó.

3.1.10

Thuyết liên kết (Connectionism)

Hình mẫu theo thuyết liên kết (connectionist paradigm)

Mô hình theo thuyết liên kết (connectionist model)

Tiếp cận theo thuyết liên kết (connectionist approach)

Dạng của mô hình nhận thức sử dụng một mạng liên kết các đơn vị với nhau, thường là các đơn vị tính toán đơn giản.

3.1.11

Khai phá dữ liệu (data mining)

Quá trình điện toán trích xuất các kiểu mẫu bằng cách phân tích dữ liệu định lượng theo bối cảnh hoặc khía cạnh khác nhau để từ đó phân loại, tóm lược các mối quan hệ và tác động tiềm tàng có thể.

[nguồn: ISO 16439: 2014, 3.13 được sửa đổi - thay thế "phân loại nó" bằng "phân loại chúng" vì dữ liệu là số nhiều].

3.1.12

Tri thức khai báo (declarative knowledge)

Tri thức được biểu thị bằng các dữ kiện, quy tắc và định lý.

CHÚ THÍCH 1: Thông thường, tri thức khai báo không thể được xử lý nếu như đầu tiên nó không được chuyển đổi thành *tri thức theo trình tự* (3.1.28).

[nguồn: ISO/IEC 2382-28: 1995, 28.02.22 được sửa đổi - Bỏ dấu phẩy sau "quy tắc" trong định nghĩa].

3.1.13

Hệ chuyên gia (expert system)

Hệ thống AI (3.1.4) tích lũy, kết hợp và đóng gói tri thức (3.1.21) được cung cấp bởi chuyên gia là con người hoặc các chuyên gia trong lĩnh vực cụ thể để suy luận giải pháp cho các vấn đề đặt ra.

3.1.14

AI tổng quát (general AI)

AGI

Trí tuệ tổng quát nhân tạo (artificial general intelligence)

Loại hình *hệ thống AI* (3.1.4) giải quyết một phạm vi rộng các *tác vụ* (3.1.35) với mức độ thỏa đáng về hiệu năng thực thi.

CHÚ THÍCH 1: So sánh với *AI hẹp* (3.1.24).

CHÚ THÍCH 2: AGI thường được sử dụng theo nghĩa mạnh hơn, nghĩa là các hệ thống không chỉ có thể thực hiện phạm vi đa dạng các tác vụ mà còn thực hiện mọi tác vụ do con người thực hiện.

3.1.15

Thuật toán di truyền (genetic algorithm)

GA

Thuật toán mô phỏng chọn lọc tự nhiên bằng cách tạo và làm tiến hóa một quần thể các cá thể (giải pháp) cho các bài toán tối ưu.

3.1.16

Dị thường (heteronomy)

Tính dị thường (heteronomous)

Đặc tính của một hệ thống hoạt động chịu sự ràng buộc bởi các can thiệp, điều khiển, giám sát từ bên ngoài.

3.1.17

Suy luận (inferene)

Lập luận để rút ra các kết luận từ các tiên đề đã biết.

CHÚ THÍCH 1: Trong AI, một tiên đề có thể là một dữ kiện, một quy tắc, một mô hình, một tính năng hoặc một dữ liệu thô.

CHÚ THÍCH 2: Thuật ngữ "suy luận" đề cập đến cả quá trình và kết quả của nó.

[nguồn: ISO/IEC 2382:2015, 2123830 được sửa đổi – Bổ sung mô hình, tính năng và dữ liệu thô. Xóa "Chú thích 4 cho mục 28.03.01 (2382)". Bỏ "Chú thích 3 cho mục suy luận: thuật ngữ và định nghĩa được tiêu chuẩn hóa bởi ISO/IEC 2382-28:1995"].

3.1.18

Internet vạn vật (Internet of things)

IoT

Hạ tầng các thực thể, con người, hệ thống và tài nguyên thông tin được kết nối với nhau cùng với các dịch vụ xử lý và phản hồi thông tin trong thế giới thực và thế giới ảo.

[nguồn: ISO/IEC 20924:2021, 3.2.4 được sửa đổi - "... dịch vụ xử lý và phản ứng với..." được thay thế bằng "... dịch vụ xử lý và phản ứng với..." và từ viết tắt đã được chuyển sang dòng riêng].

3.1.19

Thiết bị IoT (IoT device)

Thực thể của một *hệ thống IoT* (3.1.20) tương tác và giao tiếp với thế giới vật chất thông qua cảm nhận hoặc dẫn động

CHÚ THÍCH 1: Thiết bị IoT có thể là thiết bị cảm biến hoặc thiết bị dẫn động

[nguồn: ISO/IEC 20924:2021, 3.2.6].

3.1.20

Hệ thống IoT (IoT system)

Hệ thống cung cấp các chức năng về *IoT* (3.1.18)

CHÚ THÍCH 1: Hệ thống IoT có thể bao gồm nhưng không giới hạn các thiết bị IoT, thiết bị cổng IoT, thiết bị cảm biến và thiết bị dẫn động.

[nguồn: ISO/IEC 20924:2021, 3.2.9].

3.1.21

Tri thức (knowledge)

<Trí tuệ nhân tạo> thông tin tóm lược về các đối tượng, sự kiện, khái niệm, quy tắc, các mối quan hệ và đặc tính của chúng, được tổ chức để sử dụng một cách có hệ thống theo mục tiêu có định hướng.

CHÚ THÍCH 1: Tri thức trong lĩnh vực AI không bao hàm khả năng nhận thức, khác với cách sử dụng thuật ngữ này trong một số lĩnh vực khác. Cụ thể tri thức không bao hàm hoạt động nhận thức của quá trình hiểu biết.

CHÚ THÍCH 2: Thông tin có thể tồn tại ở dạng số hoặc ký hiệu.

CHÚ THÍCH 3: Thông tin là dữ liệu đã được ngữ cảnh hóa để có thể diễn giải được. Dữ liệu được tạo ra thông qua sự trừu tượng hóa hoặc đo lường từ các đối tượng.

3.1.22

Vòng đời (life cycle)

Sự phát triển của một hệ thống, sản phẩm, dịch vụ, dự án hoặc các thực thể khác do con người tạo ra,

từ lúc hình thành cho đến khi kết thúc hoạt động.

[nguồn: ISO/IEC/IEEE 15288: 2015, 4.1.23]

3.1.23

Mô hình (model)

Dạng biểu diễn vật lý, toán học hoặc logic khác của một hệ thống, thực thể, hiện tượng, quá trình, dữ liệu.

[nguồn: ISO/IEC 18023-1: 2006, 3.1.11, được sửa đổi - Bỏ dấu phẩy sau "toán học" trong định nghĩa, "hoặc dữ liệu" được thêm vào cuối].

3.1.24

AI hẹp (narrow AI)

Loại hình *hệ thống AI* (3.1.4) tập trung vào các *tác vụ* (3.1.35) xác định để giải quyết một vấn đề cụ thể.

CHÚ THÍCH 1: So sánh với *AI tổng quát* (3.1.14).

3.1.25

Hiệu năng (Performance)

Kết quả có thể đo lường được

CHÚ THÍCH 1: Hiệu năng có thể liên quan đến các phát hiện định lượng hoặc định tính.

CHÚ THÍCH 2: Hiệu năng có thể liên quan đến việc quản lý các hoạt động, quá trình, sản phẩm (bao gồm cả dịch vụ), hệ thống hoặc tổ chức.

3.1.26

Lập kế hoạch (planning)

Trong trí tuệ nhân tạo, các quá trình tính toán tạo ra một luồng công việc từ một tập hợp các hành động nhằm đạt được một mục tiêu cụ thể

CHÚ THÍCH 1: Ý nghĩa của "lập kế hoạch" sử dụng trong vòng đời hoặc theo các tiêu chuẩn quản lý AI cũng có thể là các hoạt động được thực hiện bởi con người.

3.1.27

Dự đoán (prediction)

Đầu ra chính của *hệ thống AI* (3.1.4) khi được cung cấp thông tin hoặc *dữ liệu đầu vào* (3.2.9).

CHÚ THÍCH 1: Các dự đoán có thể sau đó là đầu ra bổ sung chẳng hạn như các khuyến nghị, quyết định và hành động.

CHÚ THÍCH 2: Dự đoán không nhất thiết đề cập đến việc dự đoán điều gì đó trong tương lai.

CHÚ THÍCH 3: Các dự đoán có thể đề cập đến các loại hình phân tích hoặc tạo lập dữ liệu khác nhau, áp dụng cho dữ liệu mới hoặc dữ liệu lịch sử (bao gồm dịch văn bản, tạo hình ảnh tổng hợp hoặc chẩn đoán sự cố mất điện trước đó).

3.1.28

Tri thức theo trình tự (procedural knowledge)

Tri thức thức chỉ thị rõ ràng các bước cần thực hiện để giải quyết một vấn đề hoặc để đạt được một mục tiêu.

[nguồn: ISO/IEC 2382-28: 1995, 28.02.23]

3.1.29

Người máy (robot)

Hệ thống tự động hóa với cơ cấu chấp hành để thực hiện các *tác vụ* (3.1.35) dự kiến trong thế giới vật lý bằng cảm nhận môi trường của nó và hệ thống phần mềm điều khiển.

CHÚ THÍCH 1: người máy bao gồm hệ thống điều khiển và giao diện của hệ thống điều khiển.

CHÚ THÍCH 2: Việc phân loại người máy là người máy công nghiệp hoặc người máy dịch vụ được thực hiện tùy thuộc mục đích ứng dụng của nó.

CHÚ THÍCH 3: Để thực hiện đúng các *tác vụ* (3.1.35), người máy sử dụng các loại cảm biến khác nhau để xác nhận trạng thái hiện tại của nó và nhận biết các yếu tố tạo nên môi trường mà nó hoạt động.

3.1.30

Khoa học người máy (robotics)

Khoa học và thực tiễn thiết kế, sản xuất và ứng dụng người máy.

[nguồn: ISO 8373: 2012, 2.16].

3.1.31

Điện toán ngữ nghĩa (semantic computing)

Lĩnh vực điện toán hướng tới xác định ý nghĩa của nội dung tính toán và ý định của người dùng và thể hiện chúng ở dạng máy có thể xử lý được.

3.1.32

Điện toán mềm (soft computing)

Lĩnh vực điện toán có khả năng khai thác sự không chính xác, không chắc chắn và đúng từng phần để đưa ra lời giải có tính thuyết phục và dễ kiểm soát hơn.

CHÚ THÍCH 1: Điện toán mềm bao gồm các kỹ thuật khác nhau như logic mờ, học máy và suy diễn thống kê.

3.1.33

AI biểu trưng (symbolic AI)

AI (3.1.3) dựa trên các kỹ thuật và *mô hình* (3.1.23) thao tác trên các biểu tượng và cấu trúc theo các quy tắc được xác định rõ ràng để có được các suy luận.

CHÚ THÍCH 1: So sánh với AI biểu trưng phụ (3.1.34), AI biểu trưng tạo ra các đầu ra kết quả khai báo, ngược lại AI biểu trưng phụ dựa trên các phương pháp thống kê và tạo ra kết quả đầu ra với một xác suất sai nhất định.

3.1.34

AI biểu trưng phụ (subsymbolic AI)

AI (3.1.3) dựa trên các kỹ thuật và *mô hình* (3.1.23) sử dụng mã hóa ẩn thông tin thu được từ trải nghiệm hoặc dữ liệu thô.

CHÚ THÍCH 1: *So với AI biểu trưng* (3.1.33). Trong khi AI biểu trưng tạo ra các đầu ra kết quả khai báo, AI biểu trưng phụ dựa trên các phương pháp thống kê và tạo ra các kết quả đầu ra với xác suất sai nhất định.

3.1.35

Tác vụ (task)

<Trí tuệ nhân tạo> tác vụ là hoạt động cần thiết để đạt được một mục tiêu cụ thể.

CHÚ THÍCH 1: hoạt động có thể là vật lý hoặc nhận thức. Ví dụ: tính toán hoặc tạo các *dự đoán* (3.1.27), bản dịch, dữ liệu tổng hợp, các tạo tác hoặc điều hướng trong một không gian vật lý.

CHÚ THÍCH 2: Ví dụ về các tác vụ bao gồm phân loại, hồi quy, xếp hạng, phân cụm và giảm kích thước dữ liệu.

3.2 Các thuật ngữ liên quan đến dữ liệu

3.2.1

Chú giải dữ liệu (data annotation)

Quá trình đính kèm một tập hợp thông tin mô tả vào dữ liệu mà không có bất kỳ thay đổi nào đối với dữ liệu đó.

CHÚ THÍCH 1: Thông tin mô tả có thể ở dạng siêu dữ liệu, các nhãn và neo.

3.2.2

Kiểm tra chất lượng dữ liệu (data quality checking)

Quá trình trong đó dữ liệu được kiểm tra tính đầy đủ, độ sai lệch và các yếu tố khác ảnh hưởng đến tính hữu dụng của dữ liệu cho một *hệ thống AI* (3.1.4).

3.2.3

Tăng cường dữ liệu (data augmentation)

Quá trình trình tạo các mẫu tổng hợp bằng cách sửa đổi hoặc tận dụng dữ liệu hiện có.

3.2.4

Lấy mẫu dữ liệu (data sampling)

Quá trình chọn một tập hợp con các kiểu mẫu dữ liệu nhằm thể hiện hình thái và xu hướng tương tự tương tự như *tập dữ liệu* (3.2.5) lớn hơn đang được phân tích

CHÚ THÍCH 1: Lý tưởng nhất là tập hợp con các mẫu dữ liệu sẽ đại diện cho *tập dữ liệu* (3.2.5) lớn hơn.

3.2.5

Tập dữ liệu (dataset)

Tập hợp dữ liệu với một định dạng được chia sẻ

VÍ DỤ 1: Các bài đăng trên blog siêu nhỏ từ tháng 6 năm 2020 được liên kết với thẻ bắt đầu bằng # #rugby và #football.

VÍ DỤ 2: Ảnh chụp cận cảnh (macro photography) của bông hoa có kích thước 256x256 pixel.

CHÚ THÍCH 1: Tập dữ liệu có thể được sử dụng để xác thực hoặc kiểm tra một *mô hình AI* (3.1.23). Trong ngữ cảnh *học máy* (3.3.5), tập dữ liệu có thể được sử dụng để đào tạo một *thuật toán học máy* (3.3.6).

3.2.6

Phân tích dữ liệu thăm dò (exploratory data analysis)

EDA

Kiểm tra ban đầu dữ liệu để phát hiện các đặc điểm nổi bật và đánh giá chất lượng của nó.

CHÚ THÍCH 1: EDA có thể bao gồm việc xác định các giá trị còn thiếu, giá trị ngoại lệ, tính đại diện cho tác vụ cần thực hiện - xem *kiểm tra chất lượng dữ liệu* (3.2.2).

3.2.7

Sự thật nền (ground truth)

Giá trị của biến mục tiêu cho một khoản mục cụ thể của dữ liệu đầu vào được gắn nhãn.

CHÚ THÍCH 1: Thuật ngữ sự thật nền không ngụ ý rằng dữ liệu đầu vào được gắn nhãn phù hợp một cách nhất quán với giá trị trong thế giới thực của các biến mục tiêu.

3.2.8

Áp đặt (imputation)

Quá trình trong đó dữ liệu bị thiếu được thay thế bằng dữ liệu ước tính hoặc dữ liệu được mô hình hóa.

[nguồn: ISO 20252: 2019, 3.45]

3.2.9

Dữ liệu đầu vào (input data)

Dữ liệu để một *hệ thống AI* (3.1.4) tính toán đưa ra một dự báo đầu ra hoặc một suy luận

3.2.10

Nhãn (label)

Biến mục tiêu gán cho một mẫu

3.2.11

Thông tin nhận dạng cá nhân (personally identifiable information)

PII

Dữ liệu cá nhân (personal data)

Bất kỳ thông tin nào: (a) có thể được sử dụng để thiết lập mối liên hệ giữa thông tin và thể nhân mà thông tin đó liên quan đến, hoặc (b): có thể liên kết trực tiếp hoặc gián tiếp đến một thể nhân.

CHÚ THÍCH 1: "thể nhân" trong định nghĩa là người sở hữu PII. Để xác định liệu người sở hữu PII có thể nhận dạng được hay không, cần tính đến tất cả các phương tiện có thể được sử dụng với lý do chính đáng bởi bên liên quan đến quyền riêng tư nắm giữ dữ liệu hoặc bởi bất kỳ bên nào khác để thiết lập mối liên kết giữa tập *PII* và thể nhân.

CHÚ THÍCH 2: Định nghĩa này được đưa vào để định nghĩa khái niệm PII được sử dụng trong phạm vi tiêu chuẩn. Bộ xử lý PII trên đám mây công khai thường không ở vị trí để nhận biết rằng liệu thông tin mà nó xử lý có thuộc bất kỳ danh mục cụ thể nào hay không, trừ phi thông tin đó được minh bạch bởi khách hàng của dịch vụ đám mây.

[nguồn: ISO/IEC 29100: 2011 / Amd1: 2018, 2.9]

3.2.12

Dữ liệu sản xuất (production data)

Dữ liệu thu được trong giai đoạn vận hành của *hệ thống AI* (3.1.4), trong đó *hệ thống AI* (3.1.4) được triển khai (3.1.4) tính toán đưa ra kết quả dự đoán đầu ra hoặc *suy luận* (3.1.17).

3.2.13

Mẫu (sample)

Phần tử dữ liệu không thể phân chia được xử lý theo số lượng bằng *thuật toán học máy* (3.3.6) hoặc *hệ thống AI* (3.1.4)

3.2.14

Dữ liệu kiểm tra (test data)

Dữ liệu đánh giá (evaluation data)

Dữ liệu được sử dụng để đánh giá hiệu năng của một *mô hình* (3.1.23) cuối cùng.

CHÚ THÍCH 1: Dữ liệu kiểm tra tách biệt với *dữ liệu huấn luyện* (3.3.16) và *dữ liệu thẩm định* (3.2.15).

3.2.15

Dữ liệu thẩm định (validation data)

Dữ liệu phát triển (development data)

Dữ liệu được sử dụng để so sánh hiệu năng của các *mô hình* (3.1.23) ứng tuyển khác nhau.

CHÚ THÍCH 1: Dữ liệu thẩm định tách biệt với *dữ liệu kiểm tra* (3.2.14) và cả *dữ liệu huấn luyện* (3.3.16) nói chung. Tuy nhiên, trong trường hợp không có đủ dữ liệu để phân tách thành các tập dữ liệu huấn luyện, xác nhận và kiểm tra cho ba cách thức thực hiện tương ứng thì dữ liệu chỉ được chia thành hai tập: tập dữ liệu kiểm tra và tập dữ liệu huấn luyện hoặc xác nhận. Xác

thực chéo hoặc khởi động mồi là các phương pháp phổ biến để sau đó tạo các tập dữ liệu huấn luyện và xác nhận riêng biệt từ tập huấn luyện hoặc tập dữ liệu thẩm định.

CHÚ THÍCH 2: Dữ liệu thẩm định có thể được sử dụng để điều chỉnh các siêu tham số hoặc xác nhận một số lựa chọn thuật toán, tùy thuộc vào hiệu quả của việc đưa một quy tắc đã cho vào trong một hệ thống chuyên gia.

3.3 Các thuật ngữ liên quan đến học máy

3.3.1

Mạng Bayes (Bayesian network)

Mô hình (3.1.23) xác suất sử dụng *suy luận* (3.1.17) Bayes để tính toán xác suất bằng đồ thị không lặp vòng có hướng.

3.3.2

Cây quyết định (decision tree)

Mô hình (3.1.23) mà *suy luận* (3.1.17) được mã hóa dưới dạng các đường dẫn từ gốc đến nút lá trong cấu trúc của cây.

3.3.3

Hợp tác người – máy (human-machine teaming)

Tích hợp tương tác của con người với khả năng thông minh của máy móc

3.3.4

Siêu tham số (hyperparameter)

Đặc tính của các *thuật toán học máy* (3.3.6) ảnh hưởng đến quá trình học tập của nó.

CHÚ THÍCH 1: Các siêu tham số được chọn trước khi huấn luyện và có thể được sử dụng trong quá trình trợ giúp ước lượng các tham số mô hình.

CHÚ THÍCH 2: Các ví dụ về siêu tham số bao gồm số lượng lớp mạng, độ rộng của mỗi lớp, loại hình chức năng kích hoạt, phương pháp tối ưu hóa, tốc độ học đối với mạng nơ-ron; lựa chọn chức năng lõi trong một máy véc-tơ hỗ trợ; số lượng lá hoặc độ sâu của cây; K cho K- giá trị trung bình phân cụm; số lượng tối đa lần lặp lại của thuật toán tối đa hóa kỳ vọng; số lượng phân bố Gaussian trong một hỗn hợp Gaussian.

3.3.5

Học máy (machine learning)

ML

Quá trình tối ưu hóa các *tham số mô hình* (3.3.8) thông qua các kỹ thuật điện toán, sao cho hành vi của *mô hình* (3.1.23) phản ánh dữ liệu hoặc trải nghiệm.

3.3.6

Thuật toán học máy (machine learning algorithm)

Thuật toán xác định các *tham số* (3.3.8) của một *mô hình học máy* (3.3.7) từ dữ liệu theo tiêu chí đã cho.

VÍ DỤ: Xem xét việc giải một hàm tuyến tính đơn biến $y = \theta_0 + \theta_1 x$ trong đó y là đầu ra hoặc kết quả, x là đầu vào, θ_0 là giá trị chặn (giá trị của y trong đó x = 0) và θ_1 là trọng số. Trong *học máy* (3.3.5), quá trình xác định giá trị chặn và trọng số cho một hàm tuyến tính được gọi là hồi quy tuyến tính.

3.3.7

Mô hình học máy (machine learning model)

Cấu trúc toán học tạo ra một *suy luận* (3.1.17) hoặc *dự đoán* (3.1.27) dựa trên dữ liệu hoặc thông tin đầu vào.

VÍ DỤ: Nếu một hàm tuyến tính đơn biến ($y = \theta_0 + \theta_1 x$) đã được huấn luyện bằng cách sử dụng hồi quy tuyến tính, mô hình thu được có thể là $y = 3 + 7x$.

CHÚ THÍCH 1: Mô hình học máy là kết quả của quá trình huấn luyện dựa trên *thuật toán học máy* (3.3.6).

3.3.8

Tham số (parameter)

Tham số mô hình (model parameter)

Biến nội bộ của một *mô hình* (3.1.23) ảnh hưởng đến cách tính toán các kết quả đầu ra của nó.

CHÚ THÍCH 1: Ví dụ về các tham số bao gồm các trọng số trong mạng nơ-ron và xác suất chuyển đổi trong mô hình Markov.

3.3.9

Học tăng cường (reinforcement learning)

RL

Nhận biết về chuỗi hành động tối ưu để tối đa hóa phần thưởng thông qua tương tác với môi trường.

3.3.10

Tái huấn luyện (retraining)

Cập nhật một *mô hình được huấn luyện* (3.3.14) bằng *huấn luyện* (3.3.15) với các *dữ liệu huấn luyện* (3.3.16) khác nhau.

3.3.11

Học máy bán giám sát (semi-supervised machine learning)

Học máy (3.3.5) sử dụng cả dữ liệu được gắn nhãn và không được gắn nhãn khi *huấn luyện* (3.3.15).

3.3.12

Học máy có giám sát (supervised machine learning)

Học máy (3.3.5) chỉ sử dụng dữ liệu được gắn nhãn khi *huấn luyện* (3.3.15).

3.3.13

Máy vec-tơ hỗ trợ (support vector machine)

SVM

Thuật toán học máy (3.3.6) tìm ra các quyết định về biên với lề cực đại.

CHÚ THÍCH 1: Các véc-tơ hỗ trợ là tập các điểm dữ liệu xác định vị trí của các quyết định về biên (siêu mặt phẳng).

3.3.14

Mô hình được huấn luyện (trained model)

Kết quả của *huấn luyện mô hình* (3.3.15)

3.3.15

Huấn luyện (training)

Huấn luyện mô hình (model training)

Quá trình thiết lập hoặc cải thiện các tham số của *mô hình học máy* (3.3.7) dựa trên *thuật toán học máy* (3.2.10) bằng cách sử dụng *dữ liệu huấn luyện* (3.3.16).

3.3.16

Dữ liệu huấn luyện (training data)

Dữ liệu được sử dụng để huấn luyện *mô hình học máy* (3.3.7).

3.3.17

Học máy không giám sát (unsupervised machine learning)

Học máy (3.3.5) chỉ sử dụng dữ liệu không được gắn nhãn trong khi *huấn luyện* (3.3.15).

3.4 Các thuật ngữ liên quan đến mạng nơ-ron

3.4.1

Hàm kích hoạt (activation function)

Hàm được áp dụng tổ hợp có trọng số đối với tất cả các đầu vào cho một *nơ-ron* (3.4.9)

CHÚ THÍCH 1: hàm kích hoạt cho phép mạng nơ-ron tìm hiểu các tính năng phức tạp trong dữ liệu. Chúng thường là phi tuyến tính.

3.4.2

Mạng nơ-ron tích chập (convolutional neural network)

CNN

Mạng nơ-ron tích chập sâu (deep convolutional neural network)

DCNN

Mạng nơ-ron tiến (3.4.6) sử dụng *tích chập* (3.4.3) trong ít nhất một trong các lớp của nó.

3.4.3

Tích chập (convolution)

Phép toán liên quan đến tích vô hướng trượt hoặc tương quan chéo của dữ liệu đầu vào

3.4.4

Học sâu (deep learning)

Học mạng nơ-ron sâu (deep neural network learning)

<Trí tuệ nhân tạo> cách tiếp cận để tạo ra các biểu diễn phân cấp phong phú thông qua *huấn luyện* (3.3.15) *mạng nơ-ron* (3.4.8) với nhiều lớp ẩn.

CHÚ THÍCH 1: Học sâu là một tập con của *học máy* (3.3.5).

3.4.5

Bùng nổ gradient (exploding gradient)

Hiện tượng của quá trình *huấn luyện* (3.3.15) lan truyền ngược trong mạng nơ-ron, nơi tích lũy các gradient có sai số lớn, dẫn đến việc cập nhật rất nhiều cho các trọng số và làm cho *mô hình* (3.1.23) không ổn định.

3.4.6

Mạng nơ-ron tiến (feed forward neural network)

FFNN

Mạng nơ-ron (3.4.8) trong đó thông tin được đưa từ lớp đầu vào đến lớp đầu ra chỉ theo một hướng.

3.4.7

Bộ nhớ ngắn – dài hạn (long short-term memory)

Loại *mạng nơ-ron hồi quy* (3.4.10) xử lý dữ liệu tuần tự với hiệu suất thỏa đáng cho cả khoảng phụ thuộc dài và ngắn.

3.4.8

Mạng nơ-ron (neural network)

NN

Mạng nơ-ron (neural net)

Mạng nơ-ron nhân tạo (artifical neural network)

<Trí tuệ nhân tạo> mạng của một hoặc nhiều lớp nơ-ron được kết nối bằng các liên kết có trọng số có thể điều chỉnh để lấy dữ liệu đầu vào và tạo ra đầu ra.

CHÚ THÍCH 1: Mạng nơ-ron là một ví dụ nổi bật của *tiếp cận theo thuyết liên kết* (3.1.10).

CHÚ THÍCH 2: Mặc dù thiết kế ban đầu của mạng nơ-ron bắt nguồn từ nguyên lý hoạt động của các nơ-ron sinh học, nhưng hầu hết các công việc nghiên cứu về mạng nơ-ron hiện nay không tuân theo nguyên lý đó nữa.

3.4.9

Nơ-ron (neuron)

Trong trí tuệ nhân tạo, phần tử xử lý cơ bản nhất nhận một hoặc nhiều giá trị đầu vào và tạo ra giá trị đầu ra bằng cách kết hợp các giá trị đầu vào và áp dụng *hàm kích hoạt* (3.4.1) trên kết quả.

CHÚ THÍCH 1: Ví dụ về các hàm kích hoạt phi tuyến là hàm ngưỡng, hàm sigmoid và hàm đa thức.

3.4.10

Mạng nơ-ron hồi quy (recurrent neural network)

RNN

Mạng nơ-ron (3.4.8) trong đó kết quả đầu ra từ cả lớp trước và bước xử lý trước đó được đưa vào lớp hiện tại.

3.5 Các thuật ngữ liên quan đến tính đáng tin cậy

3.5.1

Chịu trách nhiệm (accountable)

Có thể trả lời cho các hành động, quyết định và hiệu năng.

[nguồn: ISO/IEC 38500:2015, 2.2]

3.5.2

Trách nhiệm giải trình (accountability)

Trạng thái chịu trách nhiệm *giải trình* (3.5.1)

CHÚ THÍCH 1: Trách nhiệm giải trình liên quan đến trách nhiệm được phân công. Trách nhiệm có thể dựa trên quy định hoặc thỏa thuận hoặc thông qua sự phân công như một phần của sự ủy quyền.

CHÚ THÍCH 2: Trách nhiệm giải trình liên quan đến việc một cá nhân hoặc tổ chức chịu trách nhiệm về một điều gì đó trước một cá nhân hoặc tổ chức khác, thông qua các phương tiện cụ thể và theo các tiêu chí cụ thể.

[nguồn: ISO/IEC 38500:2015, 2.3 được sửa đổi – Bổ sung CHÚ THÍCH 2].

3.5.3

Tính khả dụng (availability)

Thuộc tính có thể được truy cập và sử dụng theo yêu cầu bởi một thực thể được ủy quyền

[nguồn: ISO/IEC 27000:2018, 3.7].

3.5.4

Thiên vị (bias)

Sự khác biệt mang tính hệ thống trong cách đối xử với một số đối tượng, người hoặc nhóm nhất định so với các đối tượng, người hoặc nhóm khác.

CHÚ THÍCH 1: Cách đối xử là bất kỳ loại hành động nào, bao gồm nhận thức, quan sát, biểu diễn, *dự đoán* (3.1.27) hoặc quyết định.

[nguồn: ISO/IEC TR 24027: 2021, 3.3.2 được sửa đổi - loại bỏ dấu phẩy oxford trong định nghĩa và chú thích].

3.5.5

Điều khiển (control)

Hành động có mục đích trên hoặc trong một quá trình để đáp ứng các mục tiêu cụ thể.

[nguồn: IEC 61800-7-1: 2015, 3.2.6].

3.5.6

Khả năng điều khiển (controllability)

Điều khiển được (controllable)

Thuộc tính của *hệ thống AI* (3.1.4) cho phép con người hoặc tác nhân bên ngoài khác can thiệp vào chức năng của hệ thống.

3.5.7

Tính diễn giải (explainability)

Thuộc tính của *hệ thống AI* (3.1.4) thể hiện các yếu tố quan trọng tác động đến kết quả của *hệ thống AI* (3.1.4) theo cách mà con người có thể hiểu được

CHÚ THÍCH 1: Nó nhằm trả lời câu hỏi "Tại sao?" mà thực ra không cần phải cố tranh luận rằng tiến trình hoạt động đã được thực hiện nhất thiết phải là tối ưu.

3.5.8

Tính dự đoán (predictability)

Thuộc tính của *hệ thống AI* (3.1.4) cho phép các bên liên quan (3.5.13) đưa ra các giả định đáng tin cậy về đầu ra.

[Nguồn: ISO/IEC TR 27550:2019, 3.12, "của các cá nhân, chủ sở hữu và nhà điều hành về PII và quá trình xử lý của nó bởi một hệ thống" đã được thay thế bằng "bởi các bên liên quan về đầu ra"].

3.5.9

Tính tin cậy (reliability)

Thuộc tính về sự phù hợp của hành vi và kết quả dự kiến

[nguồn: ISO/IEC 27000:2018, 2.55].

3.5.10

Khả năng phục hồi (resilience)

Khả năng của một hệ thống để phục hồi tình trạng hoạt động một cách nhanh chóng sau sự cố.

3.5.11

Rủi ro (risk)

Ảnh hưởng của tính bất định đến các mục tiêu

CHÚ THÍCH 1: Ảnh hưởng là độ lệch so với cái được kỳ vọng. Nó có thể tích cực, tiêu cực hoặc cả hai, và nó tạo ra hoặc dẫn đến các cơ hội và mối đe dọa.

CHÚ THÍCH 2: Mục tiêu có thể có các khía cạnh và phạm trù khác nhau và có thể được áp dụng ở các cấp độ khác nhau.

CHÚ THÍCH 3: Rủi ro thường được thể hiện dưới dạng các nguồn gốc rủi ro, các sự kiện tiềm ẩn, hậu quả và khả năng xảy ra của chúng.

[nguồn: ISO 31000: 2018, 3.1 được sửa đổi - Bỏ dấu phẩy sau "cả hai" trong CHÚ THÍCH 1. Bỏ dấu phẩy sau "danh mục" trong CHÚ THÍCH 2].

3.5.12

Độ bền vững (robustness)

Khả năng của một hệ thống duy trì mức độ thực thi của nó trong bất kỳ hoàn cảnh nào.

3.5.13

Bên liên quan (stakeholder)

Bất kỳ cá nhân, nhóm hoặc tổ chức nào có thể ảnh hưởng, bị ảnh hưởng hoặc tự nhận thức bị ảnh hưởng bởi một quyết định hoặc hành động.

[nguồn: ISO/IEC 38500:2015, 2.24 được sửa đổi - Bỏ dấu phẩy sau "bị ảnh hưởng bởi" trong định nghĩa].

3.5.14

Tính minh bạch (transparency)

<tổ chức> Thuộc tính của một tổ chức mà các hoạt động và quyết định phù hợp được truyền đạt cho các *bên liên quan* (3.5.13) một cách toàn diện, dễ tiếp cận và dễ hiểu.

CHÚ THÍCH 1: Thông tin không phù hợp về các hoạt động và quyết định có thể vi phạm an ninh, quyền riêng tư hoặc các yêu cầu về tính bảo mật.

3.5.15

Tính minh bạch (transparency)

<hệ thống> Thuộc tính của một hệ thống mà thông tin phù hợp về hệ thống tạo ra để khả dụng đối với các *bên liên quan* (3.5.13) có liên đới đến các thông tin đó.

CHÚ THÍCH 1: Thông tin phù hợp cho tính minh bạch của hệ thống có thể bao gồm các khía cạnh như tính năng, hiệu năng, các giới hạn, thành phần, thủ tục, biện pháp, các mục tiêu thiết kế, lựa chọn thiết kế và giả định, các nguồn dữ liệu và giao thức gắn nhãn.

CHÚ THÍCH 2: Tiết lộ không phù hợp về một số khía cạnh của hệ thống có thể vi phạm các yêu cầu về bảo mật, quyền riêng tư hoặc các yêu cầu về tính bảo mật.

3.5.16

Tính đáng tin cậy (trustworthiness)

Khả năng đáp ứng kỳ vọng của các *bên liên quan* (3.5.13) theo cách thức có thể xác minh được.

CHÚ THÍCH 1: Tùy thuộc vào bối cảnh hoặc lĩnh vực, cũng như sản phẩm hoặc dịch vụ cụ thể, dữ liệu và công nghệ được sử dụng, các đặc điểm khác nhau được áp dụng và cần được xác minh để đảm bảo đáp ứng kỳ vọng của các *bên liên quan* (3.5.13).

CHÚ THÍCH 2: Các đặc điểm của tính đáng tin cậy bao gồm: ví dụ như độ tin cậy, tính khả dụng, khả năng phục hồi, bảo mật, quyền riêng tư, an toàn, trách nhiệm giải trình, tính minh bạch, tính toàn vẹn, tính xác thực, chất lượng và khả năng sử dụng.

CHÚ THÍCH 3: Tính đáng tin cậy là một thuộc tính có thể được áp dụng cho các dịch vụ, sản phẩm, công nghệ, dữ liệu và thông tin, cũng như trong bối cảnh quản trị đối với các tổ chức.

[nguồn: ISO/IEC TR 24028: 2020, 3.42 được sửa đổi - Kỳ vọng của các bên liên quan được thay thế bằng kỳ vọng các bên liên quan; dấu phẩy giữa chất lượng và khả năng sử dụng được thay thế bằng "và"].

3.5.17

Xác minh (verification)

Xác nhận, thông qua việc cung cấp bằng chứng khách quan, rằng các yêu cầu cụ thể đã được đáp ứng.

CHÚ THÍCH 1: Việc xác minh chỉ cung cấp sự đảm bảo rằng một sản phẩm phù hợp với đặc điểm kỹ thuật của nó.

[nguồn: TCVN 13266:2021 ISO/IEC 27042:2015, 3.21].

3.5.18

Thẩm định (validation)

Xác nhận, thông qua việc cung cấp bằng chứng khách quan, rằng các yêu cầu cho mục đích sử dụng hoặc ứng dụng cụ thể đã được đáp ứng.

[nguồn: TCVN ISO 27043:2019 ISO/IEC 27043:2015, 3.16]

3.6 Các thuật ngữ liên quan đến xử lý ngôn ngữ tự nhiên

3.6.1

Tóm tắt tự động (automatic summarization)

Tác vụ (3.1.35) rút ngắn một phần nội dung hoặc văn bản của *ngôn ngữ tự nhiên* (3.6.7) nhưng vẫn giữ lại thông tin ngữ nghĩa quan trọng.

3.6.2

Quản lý đối thoại (dialogue management)

Tác vụ (3.1.35) chọn bước tiếp theo thích hợp trong một đối thoại dựa trên thông tin đầu vào của người dùng, lịch sử đối thoại và *tri thức* (3.1.21) ngữ cảnh khác để đạt được mục tiêu mong muốn.

3.6.3

Nhận biết cảm xúc (emotion recognition)

Tác vụ (3.1.35) nhận dạng và phân loại một cách có tính toán các cảm xúc được thể hiện trong một đoạn văn bản, bài phát biểu, video, hình ảnh hoặc tổ hợp của chúng.

CHÚ THÍCH 1: Ví dụ về cảm xúc bao gồm hạnh phúc, buồn bã, tức giận và vui mừng

3.6.4

Truy xuất thông tin (information retrieval)

IR

Tác vụ (3.1.35) truy xuất các tài liệu liên quan hoặc các phần của các tài liệu từ *tập dữ liệu* (3.2.5), thường dựa trên các truy vấn từ khóa hoặc *ngôn ngữ tự nhiên* (3.6.7)

3.6.5

Dịch máy (machine translation)

MT

Tác vụ (3.1.35) dịch tự động văn bản hoặc lời nói từ *ngôn ngữ tự nhiên* (3.6.7) này sang ngôn ngữ tự nhiên khác bằng hệ thống máy tính.

[nguồn: ISO 17100: 2015, 2.2.2]

3.6.6

Nhận biết thực thể được đặt tên (named entity recognition)

NER

Tác vụ (3.1.35) nhận biết và gắn nhãn tên biểu thị các thực thể và loại hình của chúng cho các chuỗi từ trong một luồng văn bản hoặc lời nói.

CHÚ THÍCH 1: Thực thể đề cập đến sự vật cụ thể hoặc trừu tượng cần quan tâm, bao gồm các mối liên hệ giữa các sự vật.

CHÚ THÍCH 2: "Thực thể được đặt tên" đề cập đến một thực thể có tên biểu thị mà tại đó tồn tại một ý nghĩa cụ thể hoặc duy

nhất.

CHÚ THÍCH 3: Tên biểu thị bao gồm tên cụ thể của người, địa điểm, tổ chức và các tên riêng khác dựa trên lĩnh vực hoặc ứng dụng.

3.6.7

Ngôn ngữ tự nhiên (natural language)

Ngôn ngữ đã và đang được sử dụng tích cực trong một cộng đồng người và các quy tắc của nó được rút ra từ quá trình sử dụng.

CHÚ THÍCH 1: Ngôn ngữ tự nhiên là ngôn ngữ bất kỳ nào của con người, có thể được diễn đạt bằng văn bản, lời nói, ngôn ngữ ký hiệu v.v..

CHÚ THÍCH 2: Ngôn ngữ tự nhiên là bất kỳ ngôn ngữ nào của con người, chẳng hạn như tiếng Anh, tiếng Tây Ban Nha, tiếng Ả Rập, tiếng Trung hoặc tiếng Nhật, được phân biệt với ngôn ngữ lập trình và ngôn ngữ hình thức, chẳng hạn như Java, Fortran, C ++ hoặc logic bậc nhất.

[nguồn: ISO/IEC 15944-8: 2012, 3.82 được sửa đổi - "và các quy tắc trong đó chủ yếu được rút ra từ việc sử dụng" được thay thế bằng "và các quy tắc của nó được rút ra từ quá trình sử dụng". Xóa dấu phẩy sau "Tiếng Trung" trong CHÚ THÍCH 2].

3.6.8

Tạo ngôn ngữ tự nhiên (natural language generation)

NLG

Tác vụ (3.1.35) chuyển đổi dữ liệu mang ngữ nghĩa thành *ngôn ngữ tự nhiên* (3.6.7).

3.6.9

Xử lý ngôn ngữ tự nhiên (natural language processing)

NLP

<hệ thống> Xử lý thông tin dựa vào hiểu biết ngôn ngữ tự nhiên ((3.6.11) hoặc *tạo ngôn ngữ tự nhiên* (3.6.8).

3.6.10

Xử lý ngôn ngữ tự nhiên (natural language processing)

NLP

<quy tắc> Quy tắc liên quan đến cách thức hệ thống thu nhận, xử lý và diễn giải *ngôn ngữ tự nhiên* (3.6.7).

3.6.11

Hiểu biết ngôn ngữ tự nhiên (natural language understanding)

NLU

Hiểu ngôn ngữ tự nhiên (natural language Comprehension)

Thông tin dạng văn bản hoặc lời nói được trích xuất bởi một khối chức năng và truyền đạt nó thành một *ngôn ngữ tự nhiên* (3.6.7), tạo một mô tả cho cả văn bản hoặc lời nói đã cho và những gì nó thể hiện.

[Nguồn: ISO/IEC 2382:2015, 2123786 đã được sửa đổi – bỏ ghi chú, bỏ gạch nối trong "ngôn ngữ - tự nhiên", bổ sung từ viết tắt NLU].

3.6.12

Nhận dạng ký tự quang học (optical character recognition)

OCR

Chuyển đổi hình ảnh của văn bản được đánh máy, in hoặc viết tay thành văn bản được mã hóa bằng máy.

3.6.13

Gán nhãn từ loại (part-of-speech tagging)

Tác vụ (3.1.35) gán nhãn phân loại (ví dụ: động từ, danh từ, tính từ) cho một từ dựa trên các thuộc tính ngữ pháp của nó.

3.6.14

Trả lời câu hỏi (question answering)

Tác vụ (3.1.35) xác định câu trả lời phù hợp nhất cho một câu hỏi bằng *ngôn ngữ tự nhiên* (3.6.7).

CHÚ THÍCH 1: Một câu hỏi có thể ở dạng bỏ ngỏ hoặc nhắm tới một câu trả lời cụ thể.

3.6.15

Trích xuất mối quan hệ (relationship extraction)

Trích xuất quan hệ (relation extraction)

Tác vụ (3.1.35) nhận dạng các mối quan hệ giữa các thực thể được đề cập trong văn bản.

3.6.16

Phân tích cảm xúc (sentiment analysis)

Tác vụ (3.1.35) tính toán để nhận dạng và phân loại các ý kiến được thể hiện trong một đoạn văn bản, bài phát biểu hoặc hình ảnh, để xác định một dải mức độ của cảm nghĩ, chẳng hạn như từ tích cực đến tiêu cực

CHÚ THÍCH 1: Ví dụ về tình cảm bao gồm tán thành, không tán thành, chiều hướng tích cực, chiều hướng tiêu cực, đồng ý và không đồng ý.

3.6.17

Nhận dạng lời nói (speech recognition)

Chuyển lời nói thành văn bản (speech-to-text)

STT

Chuyển đổi tín hiệu lời nói bởi một khối chức năng thành dạng trình bày nội dung của lời nói.

[NGUỒN: ISO/IEC 2382:2015, 2120735 được sửa đổi – xóa chú thích].

3.6.18

Tổng hợp lời nói

Tổng hợp lời nói (speech synthesis)

Chuyển đổi văn bản thành lời nói (text-to-speech)

TTS

Tạo lời nói nhân tạo

[nguồn: ISO/IEC 2382:2015, 2120745]

3.7 Các thuật ngữ liên quan đến thị giác máy tính

3.7.1

Thị giác máy tính (computer vision)

Khả năng của một khối chức năng thu nhận, xử lý và biên dịch dữ liệu đại diện cho hình ảnh hoặc video.

CHÚ THÍCH 1: Thị giác máy tính liên quan đến việc sử dụng các bộ cảm biến để tạo ra hình ảnh số của cảnh tượng trực quan. Điều này có thể bao gồm các hình ảnh, chẳng hạn như hình ảnh thu được các bước sóng ngoài ánh sáng nhìn thấy, chẳng hạn như ảnh hồng ngoại.

3.7.2

Nhận dạng khuôn mặt (face recognition)

Tự động nhận dạng kiểu mẫu, so sánh các hình ảnh được lưu trữ của khuôn mặt người với hình ảnh của khuôn mặt thực, chỉ thị bất kỳ sự trùng khớp nào về hình ảnh, dữ liệu nếu chúng tồn tại để xác định người có khuôn mặt đó.

[nguồn: ISO 5127: 2017, 3.1.12.09]

3.7.3

Hình ảnh (image)

<kỹ thuật số> Nội dung đồ họa để trình bày trực quan

CHÚ THÍCH 1: Điều này bao gồm đồ họa được mã hóa ở bất kỳ định dạng điện tử nào, bao gồm nhưng không giới hạn ở các

định dạng, chẳng hạn như các ảnh điểm riêng lẻ (ví dụ: hình ảnh được tạo ra bởi các chương trình vẽ hoặc bởi các phương tiện chụp ảnh) và các ảnh chứa các công thức (ví dụ: chúng được tạo ra dưới ảnh véc-tơ có thể phóng to hoặc thu nhỏ).

[nguồn: ISO/IEC 20071-11: 2019, 3.2.1].

3.7.4

Nhận dạng hình ảnh (image recognition)

Quá trình phân loại hình ảnh để phân loại (các) đối tượng, (các) kiểu mẫu hoặc (các) bố cục trong một ảnh.

4 Chữ viết tắt

API	Application programming interface	Giao diện lập trình ứng dụng
CPS	Cyber-physical systems	Hệ thống vật lý – mạng
CPU	Central processing unit	Đơn vị xử lý trung tâm
CRISP-DM	Cross-industry process model for data mining	Mô hình quy trình liên ngành cho khai phá dữ liệu
DNN	Deep neural network	Mạng nơ-ron sâu
DSP	Digital signal processor	Bộ xử lý tín hiệu số
FPGA	Field-programmable gate array	Cổng có thể lập trình ứng dụng cấu trúc mảng
GPU	Graphics processing unit	Khối xử lý đồ họa
HMM	Hidden Markov model	Mô hình Markov ẩn
IT	Information technology	Công nghệ thông tin
KDD	Knowledge discovery in data	Khai phá tri thức trong dữ liệu
NPU	Neural network processing unit	Khối xử lý mạng nơ-ron
OECD	Organization for economic co-operation and development	Tổ chức hợp tác và phát triển kinh tế
POS	Part of speech	Từ loại

5 Các khái niệm AI

5.1 Tổng quan

Việc nghiên cứu và phát triển liên ngành về các hệ thống AI nhằm mục đích xây dựng các hệ thống máy tính có thể thực hiện các tác vụ thường đòi hỏi sự thông minh. Các máy hỗ trợ AI hướng đến nhận thức một số môi trường nhất định và thực hiện các hành động đáp ứng các yêu cầu của chúng.

AI sử dụng các kỹ thuật từ nhiều lĩnh vực, chẳng hạn như khoa học máy tính, toán học, triết học, ngôn ngữ học, kinh tế học, tâm lý học và khoa học nhận thức.

So với hầu hết các hệ thống thông thường không phải AI; một vài hoặc toàn bộ các hệ thống AI có thể cung cấp một số tính năng thú vị:

a) Tương tác - đầu vào của hệ thống AI được cung cấp bởi bộ cảm biến hoặc thông qua tương tác với con người; đầu ra có thể dẫn sinh ra các kích thích thiết bị dẫn động hoặc cung cấp phản ứng cho con người hoặc máy móc. Một ví dụ có thể là nhận dạng đối tượng bởi hệ thống AI được hiển thị bằng hình ảnh của đối tượng đó.

b) Theo ngữ cảnh - một số hệ thống AI có thể làm việc với nhiều nguồn thông tin, bao gồm cả thông tin có cấu trúc và không có cấu trúc, cũng như các đầu vào cảm nhận.

c) Giám sát - Hệ thống AI có thể hoạt động với nhiều mức độ giám sát và kiểm soát của con người tùy thuộc vào ứng dụng. Một ví dụ là phương tiện tự lái với các mức độ tự động hóa khác nhau.

d) Thích ứng - một số hệ thống AI được thiết kế để sử dụng dữ liệu động trong thời gian thực và tái huấn luyện để cập nhật hoạt động của chúng dựa trên dữ liệu mới.

5.2 Từ AI mạnh và yếu đến AI tổng quát và hẹp

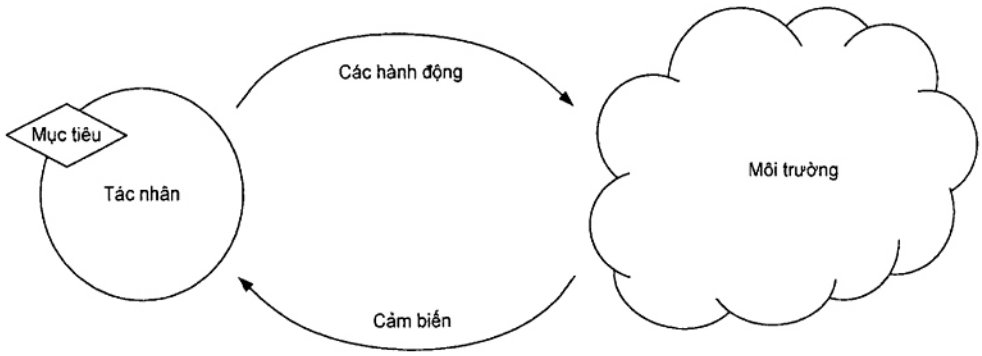
Từ quan điểm triết học, tính khả thi của những cỗ máy sở hữu trí thông minh hiện đang được thảo luận.

Từ những cuộc thảo luận này đã dẫn đến việc giới thiệu hai loại AI khác nhau và được gọi là AI yếu và AI mạnh. Trong AI yếu, hệ thống chỉ có thể xử lý các ký hiệu (chữ cái, số v.v..) mà không bao giờ hiểu những gì nó làm. Trong AI mạnh, hệ thống cũng xử lý các ký hiệu, nhưng thực sự nó hiểu là nó làm gì. Cho nên việc gọi tên "AI yếu" và "AI mạnh" chủ yếu quan trọng đối với các nhà triết học chứ không thích hợp với các nhà nghiên cứu và và triển khai AI.

Theo những thảo luận đó, khả năng của "AI hẹp" so với "AI tổng quát" hiện tại là phù hợp hơn trong lĩnh vực ứng dụng AI. Hệ thống "AI hẹp" có thể giải quyết các tác vụ xác định cho một vấn đề cụ thể (có thể tốt hơn nhiều so với con người làm). Hệ thống "AI tổng quát" giải quyết một loạt các tác vụ để đáp ứng hiệu suất thực thi theo yêu cầu. Các hệ thống AI hiện tại được coi là "hẹp" và vẫn chưa thể biết liệu các hệ thống AI "tổng quát" có khả thi về mặt kỹ thuật trong tương lai hay không.

5.3 Tác nhân

Có thể nhìn nhận các hệ thống AI từ quan điểm mô hình tác nhân vì một số ứng dụng của AI nhằm mục đích mô phỏng trí thông minh và hành vi của con người. Được định nghĩa là một kỹ thuật nghiêm ngặt, AI có thể xem như lĩnh vực đang cố gắng xây dựng các tác nhân nhân tạo thể hiện hành vi có lý trí. Mô hình tác nhân thiết lập một ranh giới rõ ràng ngăn cách tác nhân và môi trường tiến hóa của. Mô hình tác nhân được minh họa trong Hình 1.



Hình 1 – Mô hình tác nhân

Tác nhân AI tương tác với môi trường của nó thông qua các bộ cảm biến và thiết bị dẫn động, thực hiện các hành động nhằm tối đa hóa cơ hội để thành công trong việc đạt được các mục tiêu.

Môi trường có các đặc điểm khác nhau tùy thuộc vào tác vụ được thực hiện và những đặc điểm này ảnh hưởng đến độ phức tạp của việc giải quyết vấn đề.

Trong mô hình này, một vài loại tác nhân AI có thể được định nghĩa, tùy thuộc vào kiến trúc của chúng [29]:

- Tác nhân phản xạ, chỉ dựa vào tình hình hiện tại để chọn một hành động;
- Tác nhân dựa trên mô hình, dựa vào mô hình môi trường cho phép chúng xem xét kết quả các hành động khả dụng;
- Tác nhân dựa trên mục tiêu hoặc tiện ích, dựa vào chức năng tiện ích bên trong cho phép chúng lựa chọn các hành động để đạt được mục tiêu; và trong số các mục tiêu đó, tìm kiếm những hành động mong muốn nhất;
- Tác nhân học tập, có thể thu thập thông tin về môi trường và học hỏi để cải thiện hiệu năng của chúng.

Một số kiến trúc cấp cao và phức tạp dựa trên các lý thuyết khác nhau đang được phát triển để thực hiện các tác nhân.

5.4 Tri thức

Ý nghĩa AI cụ thể của "tri thức" sẽ được thảo luận chi tiết hơn do sự phổ biến của khái niệm này trong tiêu chuẩn và trong thực tiễn.

Trong các lĩnh vực khác, thuật ngữ này có thể liên kết với khả năng nhận thức. Trong bối cảnh AI, nó là một thuật ngữ kỹ thuật thuần túy đề cập đến nội dung và không phải khả năng. Khái niệm tri thức là một phần của cấu trúc phân cấp dữ liệu – thông tin – tri thứ. Theo đó dữ liệu có thể được sử dụng để tạo ra thông tin và thông tin có thể được sử dụng để tạo ra tri thức. Trong ngữ cảnh AI, điều đó thuần túy mang tính kỹ thuật, là các quá trình phi nhận thức.

Tri thức khác với thông tin ở chỗ thông tin được quan sát bởi hệ thống, trong khi tri thức là những gì hệ thống giữ lại từ những quan sát đó. Tri thức được tổ chức trong cấu trúc xác định và có tổ chức; nó tóm lược thông tin từ các đặc điểm cụ thể của các quan sát riêng lẻ. Tùy thuộc vào mục tiêu, cùng một thông tin có thể rút ra các tri thức khác nhau.

Tri thức khác với sự biểu đạt ở chỗ cùng một tri thức có thể có những cách biểu đạt khác nhau: nó có thể xuất hiện dưới những dạng cụ thể khác nhau, mỗi dạng có ưu và nhược điểm riêng, nhưng chúng đều có cùng ý nghĩa.

Những sự khác biệt này có tác động về mặt kỹ thuật, vì một số cách tiếp cận, phương pháp và các chủ đề nghiên cứu khác của AI hoàn toàn dựa vào khả năng tạo ra các kiến thức khác nhau từ cùng một thông tin hoặc các cách biểu đạt khác nhau cho cùng một tri thức.

5.5 Nhận thức và điện toán nhận thức

Nhận thức bao gồm việc thu nhận và xử lý tri thức thông qua lập luận, kinh nghiệm, trải nghiệm độc quyền hoặc chia sẻ, học tập và lĩnh hội. Nó bao gồm các khái niệm như sự chú tâm, hình thành tri thức, trí nhớ, phán đoán, đánh giá, suy diễn và tính toán, giải quyết vấn đề và ra quyết định, hiểu và tạo lập ngôn ngữ.

Điện toán nhận thức là một trong những phân ngành của AI [27]. Nó thực hiện nhận thức bằng cách sử dụng các khả năng như nhận dạng kiểu mẫu từ việc xử lý một lượng lớn thông tin. Nó cho phép mọi người tương tác với máy móc một cách tự nhiên hơn. Các tác vụ điện toán nhận thức được liên kết với học máy, xử lý giọng nói, xử lý ngôn ngữ tự nhiên, thị giác máy tính và giao diện người – máy.

5.6 Điện toán ngữ nghĩa

Điện toán ngữ nghĩa giải quyết sự phù hợp của điện toán về ngữ nghĩa nội dung với ý định của con người. Nó cung cấp các biểu đạt mô tả thông tin và sử dụng các biểu đạt này để truy xuất, quản lý, thao tác và tạo nội dung (chẳng hạn như văn bản, video, âm thanh, quy trình, chức năng, thiết bị và mạng). Mô tả ngữ nghĩa của nội dung cho phép giảm sự không chắc chắn trong các quá trình nhận thức và suy diễn logic về thông tin. Điều này giúp làm giàu thông tin, loại bỏ mâu thuẫn, tóm tắt và so sánh. Do đó, điện toán ngữ nghĩa là một cách tiếp cận để kết hợp giữa quá trình tiền xử lý thông tin và quá trình học tập.

5.7 Điện toán mềm

Điện toán mềm áp dụng phương pháp kết hợp các kỹ thuật khác nhau để có thể dung nạp sự không chính xác, không chắc chắn và biết một phần sự thật để giải quyết các vấn đề phức tạp. Các phương pháp tính toán thông thường được áp dụng để tìm ra các giải pháp chính xác và chặt chẽ cho các vấn đề. Tuy nhiên, các giải pháp như vậy có thể không phù hợp hoặc cực kỳ phức tạp. Điện toán mềm được xây dựng dựa trên nhận thức rằng thông tin về các sự vật, hiện tượng trong thế giới thực thường được phản ánh không đầy đủ, không chính xác và không chắc chắn. Do đó, việc cố gắng tìm các giải pháp

chính xác cho các vấn đề trong thế giới thực sẽ liên quan đến chi phí và độ phức tạp cho việc thực hiện. Do đó, điện toán mềm hướng tới tận dụng dung sai cho phép đối với sự không chính xác, không chắc chắn và một phần sự thật để đạt được các giải pháp có thể kiểm soát, mạnh mẽ và có chi phí thấp [24]. Ví dụ về các kỹ thuật điện toán mềm là các hệ thống mờ, các thuật toán tiến hóa, trí thông minh bầy đàn và hệ thống mạng nơ-ron.

5.8 Thuật toán di truyền

Các thuật toán di truyền mô phỏng chọn lọc tự nhiên bằng cách tạo ra và phát triển một quần thể các cá thể (giải pháp) cho bài toán tối ưu hóa. Việc tạo ra các giải pháp mới dựa trên một quần thể ban đầu được lấy cảm hứng từ đột biến gen. Nhiễm sắc thể (tập hợp "gen") được biểu diễn dưới dạng chuỗi số 0 và 1. Khi một quần thể nhiễm sắc thể ban đầu được tạo ra, bước đầu tiên chỉ là tính toán mức độ phù hợp của từng nhiễm sắc thể. Giá trị hàm phù hợp định lượng mức độ tối ưu của một giải pháp bằng cách xếp hạng nó so với các giải pháp khác. Nếu giải pháp được tạo ra không phải là tối ưu, thì một cặp nhiễm sắc thể được chọn để trao đổi các phần (trao đổi chéo) và tạo ra hai nhiễm sắc thể con. Trong bước tiếp theo, một đột biến mang tính ngẫu nhiên làm thay đổi ít nhất một gen trong các nhiễm sắc thể. Sau đó quần thể ban đầu được thay thế bằng quần thể mới và lại bắt đầu một bước lặp mới. Các lần lặp lại GA kết thúc khi một trong các tiêu chí kết thúc (thường là số lần lặp được xác định trước) được thỏa mãn. Cuối cùng thì các nhiễm sắc thể phù hợp nhất được giữ lại [25].

5.9 Các phương pháp tiếp cận biểu tượng và biểu tượng phụ cho AI

Trong lĩnh vực AI, tồn tại nhiều quan điểm khác nhau được thể hiện bằng các mô hình tương ứng với chúng. Không có cách thức phân loại nào để phân biệt rõ ràng giữa các loại AI hình khác nhau. Tuy nhiên có một vài cách thức theo đó các hệ thống AI có thể được định vị.

Kể từ khi lĩnh vực AI được hình thành, đã có hai mô hình được phát triển mang tính cạnh tranh với nhau: đó là AI biểu trưng và AI biểu trưng phụ.

AI biểu trưng liên quan đến việc mã hóa tri thức bằng các ký hiệu và cấu trúc, nó chủ yếu sử dụng quan hệ logic để mô hình hóa các quá trình suy diễn. Trong mô hình này, thông tin được mã hóa tường minh bằng cách sử dụng hình thức diễn đạt theo quy tắc, các cú pháp có thể xử lý được bằng máy tính và ngữ nghĩa của nó được hiểu bởi con người.

Cách tiếp cận khác là AI biểu trưng phụ sử dụng mô hình theo thuyết liên kết. Mô hình này không dựa trên suy diễn biểu trưng; đúng hơn là nó dựa vào mã hóa tri thức ẩn. Hình thức diễn đạt tri thức ngầm này chủ yếu dựa vao các phương pháp tiếp cận thống kê để xử lý dữ liệu kinh nghiệm hoặc dữ liệu thô. Ví dụ về loại hình hệ thống AI này là các hệ thống học máy, bao gồm các dạng khác nhau của mạng nơ-ron sâu.

Các hệ thống AI hiện đại thường chứa các yếu tố của cả loại hình AI biểu trưng và AI biểu tượng phụ. Các hệ thống như vậy được gọi là AI lai.

5.10 Dữ liệu

Dữ liệu là trung tâm của hệ thống AI. Phần lớn các hệ thông A được thiết kế để xử lý dữ liệu, cần thiết cho việc sử dụng dữ liệu để thử nghiệm. Trong các hệ thống học máy, toàn bộ vòng đời của chúng phụ thuộc vào tính khả dụng của dữ liệu.

Dữ liệu có thể ở dạng có cấu trúc (ví dụ: cơ sở dữ liệu quan hệ) hoặc dạng không có cấu trúc (ví dụ: email, tài liệu văn bản, hình ảnh, âm thanh và tệp). Dữ liệu là một khía cạnh quan trọng của hệ thống AI và chúng trải qua các quá trình bao gồm:

- Thu thập dữ liệu, trong đó dữ liệu được thu thập từ một hoặc nhiều nguồn. Dữ liệu có thể được lấy từ nguồn trong một tổ chức hoặc đưa từ bên ngoài vào. Sự phù hợp của dữ liệu cần được đánh giá, ví dụ như liệu nó có bị sai lệch theo cách thức nào đó hay nó có đủ lớn để để đại diện cho dữ liệu đầu vào hoạt động theo dự kiến;
- Phân tích dữ liệu thăm dò, trong đó các thuộc tính dữ liệu được kiểm tra để tìm ra kiểu mẫu, mối quan hệ, xu hướng và các ngoại lệ. Phân tích như vậy có thể định hướng thực hiện cho các bước sau này, chẳng hạn như huấn luyện và xác minh;
- Chú giải dữ liệu, trong đó các phần tử quan trọng của dữ liệu được thêm vào dưới dạng siêu dữ liệu (ví dụ: thông tin về nguồn gốc dữ liệu hoặc nhãn để hỗ trợ huấn luyện cho một mô hình);
- Chuẩn bị dữ liệu, trong đó dữ liệu được đưa vào ở dạng có thể sử dụng được bởi hệ thống AI;
- Lọc dữ liệu, là loại bỏ dữ liệu không mong muốn. Các tác động của bộ lọc cần được kiểm tra cẩn thận để tránh đưa ra các sai lệch không mong muốn cũng như các vấn đề khác;
- Chuẩn hóa dữ liệu, là việc điều chỉnh các giá trị dữ liệu theo một thang đo hoặc đánh giá chung để chúng có thể so sánh được với nhau về mặt toán học;
- Khử nhận dạng hoặc các quy trình khác, bước này có thể được thực hiện nếu tập dữ liệu bao gồm thông tin nhận dạng cá nhân (PII) hoặc liên kết với các cá nhân hoặc tổ chức trước khi dữ liệu được sử dụng bởi hệ thống AI (ví dụ: xem ISO/IEC 20889);
- Kiểm tra chất lượng dữ liệu, nội dung của dữ liệu được kiểm tra tính đầy đủ, độ lệch và các yếu tố khác ảnh hưởng đến tính hữu dụng của nó đối với hệ thống AI. Kiểm tra sự nhiễm độc dữ liệu là rất quan trọng để đảm bảo rằng dữ liệu huấn liện không bị nhiễm dữ liệu có thể gây ra các kết quả có hại hoặc không mong muốn;
- Lấy mẫu dữ liệu, tạo tập con đại diện cho dữ liệu được trích xuất;
- Tăng cường dữ liệu, dữ liệu có sẵn nhưng số lượng quá nhỏ cần phải qua một loại hình chuyển đổi đề mở rộng tập dữ liệu;
- Gán nhãn dữ liệu, trong đó các tập dữ liệu được gán nhãn, có nghĩa là các mẫu được liên kết với các biến mục tiêu. Nhãn thường cần cho dữ liệu kiểm tra và dữ liệu thẩm định. Một số phương pháp tiếp cận ML cũng dựa vào khả dụng của các nhãn để huấn luyện mô hình (xem 5.11.1 và 5.11.3).

Tùy thuộc vào trường hợp sử dụng và cách tiếp cận sử dụng, dữ liệu trong hệ thống AI có thể bao hàm ở các dạng:

- Dữ liệu sản xuất là dữ liệu được xử lý bởi hệ thống AI trong giai đoạn vận hành. Không phải tất cả các hệ thống AI đều liên quan đến dữ liệu sản xuất vì còn tùy thuộc vào trường hợp sử dụng, nó độc lập với thiết kế kỹ thuật và cách tiếp cận thực thi một hệ thống AI.
- Dữ liệu kiểm tra là dữ liệu sử dụng để đánh giá hoạt động của hệ thống AI trước khi triển khai. Nó được kỳ vọng là tương tự với dữ liệu sản xuất, và để đánh giá chính xác phải tách biệt dữ liệu kiểm tra khỏi bất kỳ dữ liệu nào được sử dụng trong quá trình phát triển. Tất cả phương pháp tiếp cận thực thi AI đều đảm bảo thực hiện đánh giá nhưng tùy thuộc vào tác vụ cụ thể, vì việc sử dụng dữ liệu kiểm tra không phải lúc nào cũng thích hợp.
- Dữ liệu thẩm định tương ứng với dữ liệu được nhà phát triển sử dụng để thẩm định một số thuật toán lựa chọn (tìm kiếm siêu tham số, thiết kế quy tắc v.v.). Nó có nhiều tên khác nhau tùy thuộc vào lĩnh vực AI, chẳng hạn trong xử lý ngôn ngữ tự nhiên nó thường được gọi là dữ liệu phát triển. Có những trường hợp không cần dữ liệu thẩm định.
- Dữ liệu huấn luyện sử dụng đặc thù trong ngữ cảnh học máy: nó đóng vai trò là nguyên liệu thô mà từ đó thuật toán học máy trích xuất bởi mô hình của nó để giải quyết tác vụ đã cho.

CHÚ THÍCH 1: Trong các khung đánh giá phần mềm, thẩm định là quá trình kiểm tra xem các yêu cầu nhất định đã được đáp ứng hay chưa. Nó là một phần của quá trình đánh giá. Trong ngữ cảnh riêng của AI, thuật ngữ "thẩm định" được sử dụng để chỉ quá trình tận dụng dữ liệu thiết lập các giá trị và thuộc tính nhất định có liên quan đến thiết kế hệ thống. Nó không phải là đánh giá hệ thống theo các yêu cầu đưa ra và thực hiện trước giai đoạn đánh giá.

CHÚ THÍCH 2: Trong các khung đánh giá phần mềm, "kiểm tra" đề cập đến các quá trình đa dạng khác nhau, chẳng hạn như tìm kiếm lỗi, bài kiểm tra các khối thực thi, đo thời gian xử lý. Ý nghĩa của nó trong AI đề cập cụ thể đến việc đánh giá mang tính thống kê về hiệu năng hệ thống dựa trên một tập dữ liệu chuyên dụng.

5.11 Các khái niệm về học máy

5.11.1 Học máy có giám sát

Học máy có giám sát được định nghĩa là "học máy chỉ sử dụng dữ liệu được gắn nhãn khi huấn luyện" (3.3.12). Trong trường hợp này, các mô hình ML được huấn luyện với dữ liệu huấn luyện bao gồm đầu ra hoặc biến mục tiêu (nhãn) đã biết hoặc đã xác định. Giá trị của biến mục tiêu đối với một mẫu nhất định còn được gọi là sự thật nền. Nhãn có thể thuộc bất kỳ loại nào bao gồm các giá trị phân loại, nhị phân, số hoặc các đối tượng có cấu trúc (ví dụ: chuỗi, hình ảnh, cây hoặc đồ thị) tùy thuộc vào tác vụ. Các nhãn có thể là một phần của tập dữ liệu gốc nhưng trong nhiều trường hợp chúng được xác định bằng thủ công hoặc thông qua các quy trình khác.

Học có giám sát có thể sử dụng cho các tác vụ phân loại và hồi quy, cũng như cho các tác vụ phức tạp hơn liên quan đến dự đoán có cấu trúc.

Để biết thông tin về học máy có giám sát, xem thêm trong ISO/IEC 23053.

5.11.2 Học máy không giám sát

Học máy không giám sát được định nghĩa là "học máy chỉ sử dụng dữ liệu không được gắn nhãn trong khi huấn luyện" (3.3.17).

Học máy không giám sát có thể hữu ích trong các trường hợp như phân cụm, trong đó mục tiêu của tác vụ là xác định điểm giống nhau giữa các mẫu trong dữ liệu đầu vào. Giảm kích thước của tập dữ liệu huấn luyện là một ứng dụng khác của học máy không có giám sát, trong đó các tính năng liên quan nhất mang tính thống kê được xác định với bất kể nhãn nào.

Để biết thông tin về học máy không giám sát, xem thêm trong ISO/IEC 23053.

5.11.3 Học máy bán giám sát

Học máy bán giám sát được định nghĩa là "Học máy sử dụng cả dữ liệu được gắn nhãn và không được gắn nhãn khi huấn luyện" (3.3.11). Học máy bán giám sát là sự kết hợp của học máy được giám sát và học máy không có giám sát.

Học máy bán giám sát rất hữu ích trong trường hợp việc gán nhãn tất cả các mẫu trong một tập dữ liệu huấn luyện lớn bị cấm do tốn kém về thời gian hoặc chi phí. Tham khảo ISO/IEC 23053 để biết thêm chi tiết về học máy bán giám sát.

5.11.4 Học tăng cường

Học tăng cường là quá trình huấn luyện (các) tác nhân tương tác với môi trường của nó để đạt được mục tiêu xác định trước. Trong học tăng cường, (các) tác nhân học máy học thông qua một quá trình thử lặp. Mục tiêu của (các) tác nhân là tìm ra chiến lược (tức là xây dựng một mô hình) để đạt được phần thưởng tốt nhất từ môi trường. Đối với mỗi phép thử (thành công hay không), một phản hồi gián tiếp được cung cấp bởi môi trường. Sau đó (các) tác nhân sẽ điều chỉnh hành vi của mình (tức là mô hình của nó) dựa trên phản hồi này. Tham khảo ISO/IEC 23053 để biết thêm thông tin về học tăng cường.

5.11.5 Học chuyển giao

Chuyển giao học đề cập đến một loạt các phương pháp trong đó dữ liệu để giải quyết một vấn đề nào đó được tận dụng, tri thức thu được từ nó được áp dụng để giải quyết một vấn đề khác. Ví dụ: thông tin thu được từ nhận dạng số nhà ở chế độ xem phố có thể được sử dụng để nhận dạng số viết tay. Tham khảo ISO/IEC 23053 để biết thêm chi tiết về học chuyển giao.

5.11.6 Dữ liệu huấn luyện

Dữ liệu huấn luyện bao gồm các mẫu dữ liệu được sử dụng để huấn luyện thuật toán học máy. Thông thường, dữ liệu mẫu liên quan đến một số chủ đề quan tâm cụ thể, chúng có thể bao gồm dữ cấu có trúc và không có cấu trúc. Các mẫu dữ liệu có thể được gắn nhãn hoặc không gắn nhãn.

Trong trường hợp dữ liệu được gắn nhãn thì nhãn được sử dụng để hướng dẫn quá trình đào tạo mô hình học máy. Ví dụ dữ liệu đầu vào là hình ảnh và mục đích là để quyết định xem hình ảnh có hiển thị

liệu có phải là một con mèo hay không, nhãn có thể là "đúng" đối với hình ảnh có mèo và "sai" đối với hình ảnh không có mèo. Điều này cho phép mô hình được huấn luyện thể hiện mối quan hệ thống kê giữa các thuộc tính của mẫu dữ liệu huấn luyện và biến mục tiêu.

Số lượng mẫu của dữ liệu huấn luyện và các tính năng thích hợp được chọn đóng phần vào việc mô hình ML tạo ra kết quả tốt như thế nào để phù hợp với hàm phân bố của dữ liệu hoặc biến mục tiêu. Tuy nhiên, ở đây có sự trả giá về thời gian và tài nguyên yêu cầu cho việc tính toán nếu tập dữ liệu là cực kỳ lớn.

5.11.7 Mô hình được huấn luyện

Tiêu chuẩn này định nghĩa một mô hình được huấn luyện là kết quả của hoạt động huấn luyện, mà mô hình huấn luyện lại được định nghĩa là quy trình thiết lập hoặc cải thiện các thông số của mô hình học máy dựa trên các thuật toán học máy sử dụng dữ liệu huấn luyện. Mô hình học máy là một cấu trúc toán học tạo ra các suy luận hoặc dự đoán dựa trên dữ liệu hoặc thông tin đầu vào. Mô hình được huấn luyện phải được sử dụng bởi hệ thống AI để đưa ra dự đoán dựa trên dữ liệu sản xuất từ lĩnh vực đang quan tâm. Tồn tại một vài định dạng được chuẩn hóa để lưu trữ và truyền tải mô hình đã được huấn luyện dưới dạng tập hợp các số.

5.11.8 Dữ liệu kiểm tra và thẩm định

Để đánh giá mô hình được huấn luyện, người ta thường chia dữ liệu thu được từ quá trình phát triển mô hình thành các tập dữ liệu: huấn luyện, thẩm định và kiểm tra.

Dữ liệu thẩm định được sử dụng trong và sau khi huấn luyện để điều chỉnh các siêu tham số. Dữ liệu kiểm tra được sử dụng để xác minh rằng mô hình đã học được những gì nó cần học. Cả hai đều chứa các dữ liệu không bao giờ được đưa vào mô hình trong quá trình huấn luyện. Nếu sử dụng dữ liệu huấn luyện cho mục đích kiểm tra và thẩm định thì mô hình có khả năng đã "ghi nhớ" dự đoán chính xác mà không thực sự xử lý mẫu dữ liệu. Để tránh hiện tượng đánh giá sai hiệu năng của mô hình, dữ liệu kiểm tra cũng không được đưa ra trong quá trình điều chỉnh.

Khi sử dụng thẩm định chéo, dữ liệu được phân tách sao cho mỗi một mẫu dữ liệu được sử dụng cho cả huấn luyện và thẩm định. Cách tiếp cận này phỏng tạo sử dụng một tập dữ liệu lớn hơn để cải thiện hiệu năng mô hình. Đôi khi dữ liệu không đủ để phân tách riêng rẽ thành các tập huấn luyện, thẩm định và kiểm tra. Trong những trường hợp như vậy dữ liệu chỉ cần chia thành hai tập được đặt tên là: 1. dữ liệu huấn luyện hoặc thẩm định và: 2. dữ liệu kiểm tra. Các tập dữ liệu huấn luyện và thẩm định riêng biệt sau đó lại được tạo từ dữ liệu huấn luyện hoặc thẩm định, ví dụ như thông qua khởi động mồi hoặc thẩm định chéo.

5.11.9 Tái huấn luyện

5.11.9.1 Tổng quan

Tái huấn luyện thực hiện cập nhật một mô hình được huấn luyện bằng cách huấn luyện mô hình đó với

các dữ liệu huấn luyện khác. Nó có thể cần thiết do nhiều yếu tố, bao gồm việc thiếu bộ dữ liệu huấn luyện lớn, xuất hiện sự thay đổi về dữ liệu và khái niệm.

Thay đổi về dữ liệu là độ chính xác các dự đoán của mô hình giảm dần theo thời gian do những thay đổi các đặc điểm thống kê của dữ liệu sản xuất (ví dụ: độ phân giải hình ảnh đã thay đổi hoặc một lớp dữ liệu nào đó xuất hiện thường xuyên hơn so với các lớp dữ liệu khác). Trong trường hợp này, mô hình cần được tái huấn luyện lại với dữ liệu huấn luyện mới đại diện tốt hơn cho dữ liệu sản xuất hiện tại.

Thay đổi về khái niệm là các biên quyết định bị di chuyển (ví dụ: những cái hợp pháp và những cái không có xu hướng thay đổi khi luật mới được công bố), điều này cũng làm giảm độ chính xác của các dự đoán mặc dù dữ liệu không thay đổi. Trong trường hợp khái niệm dịch chuyển, các biến mục tiêu trong dữ liệu đào tạo cần được gán lại nhãn và mô hình cần được tái huấn luyện.

Khi tái huấn luyện một mô hình đã có, một quan tâm cụ thể là khắc phục hoặc giảm thiểu những thách thức của cái gọi là "sự lãng quên thảm khốc". Nhiều thuật toán học máy chỉ thực hiện tốt tác vụ học tập nếu dữ liệu được đưa vào cùng một lúc. Nếu một mô hình được huấn luyện cho một tác vụ cụ thể thì các tham số của nó được điều chỉnh để giải quyết tác vụ đó. Khi dữ liệu huấn luyện mới được đưa vào, các điều chỉnh dựa trên những quan sát mới đó ghi đè tri thức mà mô hình đã có trước đó. Đối với mạng nơ-ron, hiện tượng này là được gọi là "sự lãng quên thảm khốc" và được coi là một trong những hạn chế cơ bản của chúng.

5.11.9.2 Học liên tục

Học liên tục, còn được gọi là học không ngừng, học suốt đời, là gia tăng huấn luyện cho một mô hình diễn ra một cách liên tục trong khi hệ thống đang chạy trong quá trình sản xuất. Đây là một trường hợp đặc biệt của tái huấn luyện, trong đó việc cập nhật cho mô hình được lặp lại và xảy ra với tần suất cao và không gây ra bất kỳ sự gián đoạn hoạt động nào.

Trong nhiều hệ thống AI, hệ thống được đào tạo trong quá trình phát triển trước khi hệ thống được đưa vào sản xuất. Quá trình này về bản chất tương tự như quá trình phát triển phần mềm tiêu chuẩn, trong đó hệ thống được xây dựng và được kiểm tra đầy đủ trước khi đưa vào sản xuất. Hoạt động của các hệ thống như vậy được đánh giá trong quá trình xác minh và dự kiến sẽ không thay đổi trong giai đoạn vận hành.

Các hệ thống AI thể hiện sự học hỏi liên tục liên quan đến việc cập nhật từng bước cho mô hình trong hệ thống đang trong quá trình sản xuất. Dữ liệu đầu vào cho hệ thống trong quá trình vận hành không chỉ được phân tích để tạo ra một đầu ra của hệ thống, mà còn đồng thời được sử dụng để điều chỉnh mô hình trong hệ thống với mục đích cải thiện mô hình trên cơ sở dữ liệu sản xuất. Tùy thuộc vào thiết kế của hệ thống AI học liên tục, có thể có các hành động của con người được yêu cầu trong quá trình này, chẳng hạn như gán nhãn dữ liệu, thẩm định việc áp dụng một bản cập nhật gia tăng cụ thể hoặc theo dõi hiệu năng của hệ thống AI.

Học liên tục có thể giúp giải quyết các hạn chế của dữ liệu huấn luyện ban đầu và nó cũng có thể giúp đối phó với sự thay đổi về dữ liệu và sai lệch khái niệm. Tuy nhiên, học liên tục mang lại những thách thức đáng kể trong việc đảm bảo rằng hệ thống AI vẫn hoạt động chính xác khi nó học. Việc xác minh hệ thống trong quá trình sản xuất là cần thiết, cũng như cần phải trích xuất dữ liệu sản xuất dùng làm một phần của tập dữ liệu huấn luyện để cập nhật hệ thống AI trong tương lai.

Do nguy cơ xuất hiện "sự lãng quên thảm khốc", học liên tục có ngụ ý đề cập đến khả năng học theo thời gian bằng cách cung cấp các quan sát mới được thực hiện trên dữ liệu hiện tại trong khi vẫn giữ lại kiến thức trước đó.

Các đặc điểm của học liên tục bao gồm:

- Học theo thời gian trong môi trường năng động (lý tưởng là trong thế giới mở);
- Bổ sung tri thức đã học trước đó bằng cách học tri thức mới để cải thiện hiệu năng (thông qua dữ liệu mới hoặc suy diễn từ tri thức hiện có);
- Khám phá các khía cạnh mới của tác vụ cần học và gia tăng việc học;
- Học trong công việc hoặc học trong khi hệ thống đang chạy trong quá trình sản xuất.

5.12 Ví dụ về thuật toán học máy

5.12.1 Mạng nơ-ron

5.12.1.1 Yêu cầu chung

Mạng nơ-ron cố gắng mô phỏng khả năng thông minh trong việc quan sát, học hỏi, phân tích và ra quyết định cho các vấn đề phức tạp. Do đó việc thiết kế mạng nơ-ron dựa trên ý tưởng về phương thức kết nối các tế bào thần kinh trong não của con người và động vật. Cấu trúc của mạng nơ-ron bao gồm các phần tử xử lý liên kết với nhau được gọi là nơ-ron. Mỗi nơ-ron nhận một số đầu vào và chỉ tạo ra một đầu ra. Chúng được tổ chức thành các lớp, nơi đầu ra của một lớp trở thành đầu vào cho lớp tiếp theo. Mỗi kết nối có một trọng số được ấn định liên quan đến tầm quan trọng của đầu vào. Mạng nơ-ron "học" bằng cách huấn luyện với các đầu vào đã biết, so sánh đầu ra thực tế với đầu ra mong đợi và sử dụng sai số để điều chỉnh trọng số. Do đó, các liên kết tạo ra câu trả lời đúng được củng cố và những liên kết tạo ra câu trả lời sai sẽ bị yếu đi.

Tiêu chuẩn này định nghĩa học sâu là một cách tiếp cận để tạo ra các biểu diễn phân cấp phong phú thông qua việc đào tạo mạng nơ-ron với nhiều lớp ẩn. Quá trình này cho phép mạng nơ-ron tinh chỉnh dần kết quả đầu ra cuối cùng. Học sâu có thể giảm hoặc loại bỏ yêu cầu về thiết kế tính năng vì các tính năng phù hợp nhất được xác định tự động. Học sâu có thể đòi hỏi đáng kể về thời gian và tài nguyên máy tính

Có rất nhiều "kiến trúc" mạng nơ-ron (về cơ bản là sự sắp xếp của các nơ-ron) và đây là một lĩnh vực nghiên cứu sôi động với nhiều kiến trúc mạng nơ-ron mới đang tiếp tục được giới thiệu. Ví dụ kiến trúc NN có thể bao gồm:

- Mạng nơ-ron tiến;
- Mạng nơ-ron tái phát;
- Mạng nơ-ron tích chập.

Các kiến trúc NN này được mô tả trong các mục từ 5.12.1.2 đến 5.12.1.4

CHÚ THÍCH: Tham khảo ISO/IEC 23053 để biết thêm thông tin về NN.

5.12.1.2 Mạng nơ-ron tiến

FFNN là kiến trúc mạng nơ-ron đơn giản nhất. Nó cung cấp thông tin từ đầu vào đến đầu ra chỉ theo một hướng. Không có kết nối nào giữa các nơ-ron trong cùng một lớp. Hai lớp liền kề thường được "kết nối hoàn toàn", trong đó mỗi rơ-ron trong một lớp có một kết nối với một nơ-ron trong lớp tiếp theo.

5.12.1.3 Mạng nơ-ron tái phát

5.12.1.3.1 Yêu cầu chung

RNN [21] giải quyết các đầu vào xuất hiện theo chuỗi thứ tự, tức là vấn đề về trình tự của các đầu vào. Ví dụ về các đầu vào bao gồm chuỗi động như luồng âm thanh và video, nhưng cũng có thể là chuỗi tĩnh như văn bản, thậm chí là hình ảnh đơn lẻ. Các RNN có các nút vừa lấy thông tin đầu vào từ lớp trước, vừa lấy thông tin từ chính chúng từ lần chuyển tiếp trước đó. RNN có thuộc tính trạng thái bị ảnh hưởng bởi quá trình học tập trong quá khứ. RNN được sử dụng rộng rãi trong nhận dạng giọng nói, dịch máy, dự báo chuỗi thời gian và nhận dạng hình ảnh. Tham khảo ISO/IEC 23053 để biết thêm thông tin về RNN.

5.12.1.3.2 Mạng bộ nhớ ngắn – dài hạn

Một mạng LSTM là một dạng của RNN được thiết kế để ghi các thông tin có sự chênh lệch thời gian dài và ngắn khác nhau để phù hợp các liên kết học dài hạn. Chúng được giới thiệu để giải quyết vấn đề triệt tiêu gradient trong RNN có liên quan đến lan truyền ngược [22].

Mạng LSTM có thể học các trình tự phức tạp, chẳng hạn như văn học của Shakespeare hoặc soạn nhạc. Tham khảo ISO/IEC 23053 để biết thêm thông tin về LSTM.

5.12.1.3.3 Mạng nơ-ron tích chập

CNN là một mạng nơ-ron bao gồm ít nhất một lớp tích chập để lọc thông tin hữu ích từ các đầu vào. Các ứng dụng phổ biến bao gồm nhận dạng hình ảnh, ghi nhãn video và xử lý ngôn ngữ tự nhiên. Tham khảo ISO/IEC 23053 để biết thêm thông tin trên CNN.

5.12.2 Mạng Bayes

Mạng Bayes là các mô hình đồ họa được sử dụng để tạo ra các dự đoán về sự phụ thuộc giữa các biến. Chúng có thể được sử dụng để tính xác suất cho các nguyên nhân hoặc các biến số góp phần tạo kết quả đầu ra. Mối quan hệ nhân quả này rất hữu ích trong các ứng dụng như chẩn đoán y học. Mạng

Bayes có những ứng dụng hữu ích khác chẳng hạn như phân tích dữ liệu, xử lý dữ liệu không đầy đủ và giảm thiểu việc áp dụng quá nhiều mô hình vào xử lý dữ liệu. Mạng Bayes dựa trên xác suất Bayes: khả năng xảy ra của một sự kiện phụ thuộc vào mức độ tin tưởng vào sự kiện đó. Thông tin thêm về mạng Bayes có thể tìm trong [20] và trong ISO/IEC 23053.

5.12.3 Cây quyết định

Cây quyết định sử dụng cấu trúc quyết định dạng cây để mã hóa các kết quả có thể xảy ra. Các thuật toán cây quyết định được sử dụng rộng rãi để phân loại và phân tích hồi quy. Cây được hình thành từ các nút quyết định và nút lá. Mỗi nút quyết định có ít nhất hai nhánh, trong khi các nút lá đại diện cho quyết định hoặc phân loại cuối cùng. Nói chung, các nút được sắp xếp theo thứ tự theo quyết định đưa ra dự đoán mạnh nhất. Dữ liệu đầu vào cần được chia thành nhiều thành tố khác nhau để xác định kết quả tốt nhất. Cây quyết định tương tự như biểu đồ luồng, trong đó tại mỗi nút quyết định một câu hỏi có thể được đặt ra để xác định quá trình sẽ tiếp tục thực hiện theo nhánh nào.

5.12.4 Máy véc-tơ hỗ trợ

SVM là một phương pháp học máy được sử dụng rộng rãi để phân lớp và phân tích hồi quy. SVM đánh dấu các mẫu dữ liệu thành hai hạng mục khác nhau và sau đó gán các mẫu dữ liệu mới cho hạng mục này hoặc hạng mục kia. SVM là các thuật toán phân lớp khoảng cách tối đa. Chúng định nghĩa một siêu mặt phẳng để tách hai lớp trên và dưới nó, cung cấp khoảng cách tối đa giữa mặt phẳng phân lớp và các điểm dữ liệu gần nhất. Các điểm gần nhất với đường biên được gọi là véc-tơ hỗ trợ. Khoảng cách trực giao giữa các véc-tơ hỗ trợ và siêu mặt phẳng là một nửa lề của SVM. Việc đào tạo một SVM liên quan đến việc tối đa hóa lề tùy thuộc vào dữ liệu từ các hạng mục khác nhau nằm đối diện nhau so với siêu mặt phẳng. SVM cũng sử dụng các hàm lõi để ánh xạ dữ liệu từ không gian đầu vào thành không gian có chiều cao hơn (đôi khi là vô hạn) mà trong đó siêu phẳng phân lớp sẽ được chọn.

Các lề-cứng SVM hiếm khi được sử dụng trong thực tế. Bộ phân lớp lề cứng chỉ hoạt động nếu dữ liệu có thể phân tách tuyến tính. Chỉ với một mẫu dữ liệu ở nằm sai phía của siêu mặt phẳng thì bộ phân lớp không giải quyết được.

Ngược lại, bộ phân lớp lề-mềm cho phép các mẫu dữ liệu vi phạm lề (nghĩa là nằm sai phía của siêu mặt phẳng). Các bộ phân lớp lề-mềm cố gắng đạt được tối đa hóa lề trong giới hạn về vi phạm lề.

Việc lớp hóa dữ liệu không gắn nhãn và sử dụng trong dự đoán và nhận dạng kiểu mẫu là những ví dụ về ứng dụng của SVM. Khi sử dụng SVM để phân tích hồi quy thì mục tiêu là ngược lại so với bộ phân lớp SVM. Mục tiêu của phân tích hồi quy SVM là làm phù hợp càng nhiều thực thể dữ liệu bên trong lề càng tốt, đồng thời với việc hạn chế các vi phạm lề (những mẫu bên ngoài lề).

5.13 Tự chủ, can thiệp và tự động hóa

Các hệ thống AI có thể được so sánh dựa trên mức độ tự động hóa và liệu chúng có chịu sự kiểm soát từ bên ngoài hay không. Nếu xét về thang mức độ thì hệ thống tự chủ nằm ở một đầu cuối thang và đầu

cuối kia là hệ thống hoàn toàn do con người kiểm soát, các hệ thống với mức độ can thiệp khác nhau nằm ở đâu đó giữa hai đầu cuối trên. Bảng 1 cho thấy mối quan hệ giữa tự chủ, can thiệp và tự động hóa, bao gồm cả trường hợp rỗng cho việc hoàn toàn không có tự động hóa.

Bảng 1 – Mối quan hệ giữa tự chủ, can thiệp và tự động hóa

		Mức độ tự động hóa	Nhận xét
Hệ thống được tự động hóa	Tự chủ	6 – Tự chủ	Hệ thống có khả năng sửa đổi lĩnh vực sử dụng dự kiến hoặc các mục tiêu của nó mà không cần sự can thiệp, kiểm soát hoặc giám sát từ bên ngoài.
	Can thiệp	5 – Tự động hóa hoàn toàn toàn	Hệ thống có khả năng thực hiện toàn bộ tác vụ của nó mà không cần sự can thiệp từ bên ngoài.
		4 – Tự động hóa cao	Hệ thống thực hiện các phầntác vụ của nó mà không cần sự can thiệp từ bên ngoài
		3 – Tự động hóa có điều kiện	Hệ thống duy trì hiệu năng ổn định ở mức độ cụ thể với một tác nhân bên ngoài sẵn sàng tiếp quản khi cần thiết
		2 – Tự động hóa từng phần	Một số chức năng phụ của hệ thống hoàn toàn tự động trong khi hệ thống vẫn chịu sự kiểm soát của tác nhân bên ngoài.
		1 – Trợ giúp	Hệ thống hỗ trợ cho một nhà điều hành.
		0 – Không tự động hóa	Nhà điều hành hoàn toàn kiểm soát hệ thống

CHÚ THÍCH: Trong luật học, quyền tự chủ đề cập đến năng lực tự quản trị. Theo nghĩa này, "tự trị" là một từ dùng sai áp dụng cho các hệ thống AI tự động, bởi vì ngay cả những hệ thống AI tiên tiến nhất cũng không tự quản trị được. Đúng hơn là các hệ thống AI hoạt động dựa trên các thuật toán và tuân theo lệnh của người vận hành. Vì những lý do này, tài liệu này không sử dụng thuật ngữ mang tính phổ biến là tự trị để mô tả tự động hóa [30].

Các tiêu chí liên quan để phân loại hệ thống theo dải mức độ này bao gồm:

- Có sự hiện diện hoặc không có sự hiện diện đối với giám sát bên ngoài bởi một người điều hành ("con người – trong – vòng lặp") hoặc bằng một hệ thống tự động khác;
- Mức độ hiểu biết chung của hệ thống, bao gồm tính đầy đủ và khả năng vận hành mô hình của hệ thống trong các trạng thái của môi trường, và độ chắc chắn mà hệ thống có thể suy diễn và hành động trong môi trường của nó;
- Mức độ phản ứng hoặc khả năng đáp ứng, bao gồm cả việc liệu hệ thống có thể nhận thấy những thay đổi trong môi trường của nó, liệu nó có thể phản ứng với những thay đổi đó, và liệu nó có thể đặt ra quy định cho những thay đổi trong tương lai hay không;

- Hoạt động của nó có tồn tại cho đến khi hoặc sau khi hoàn thành một tác vụ hoặc xuất hiện một sự kiện trong môi trường (ví dụ: liên quan đến việc đạt được mục tiêu, hết hạn định hoặc các cơ chế khác);
- Mức độ thích ứng với những thay đổi bên trong hoặc bên ngoài, nhu cầu thiết yếu hoặc xu thế;
- Khả năng tự đánh giá hiệu năng hoặc tình trạng phù hợp của hệ thống, bao gồm cả các đánh giá so với các mục tiêu đã định;
- Khả năng quyết định và lập kế hoạch một cách chủ động đối với các mục tiêu, động lực và xu thế phát triển của hệ thống.

Thay vì thay thế công việc của con người, trong một số trường hợp máy móc sẽ bổ trợ cho công việc của con người và được gọi là hợp tác giữa người và máy. Điều này có thể xảy ra như tác dụng một phía đối với sự phát triển AI hoặc một hệ thống có thể được phát triển riêng cho mục tiêu tạo ra một hệ thống người – máy. Hệ thống hướng tới bổ sung khả năng nhận thức của con người đôi khi được gọi là hệ thống tăng cường trí thông minh.

Nhìn chung, sự hiện diện của giám sát có trách nhiệm trong quá trình hoạt động có thể hỗ trợ việc đảm bảo rằng AI hệ thống hoạt động đúng như dự kiến và tránh các tác động không mong muốn đối với các bên liên quan.

5.14 Internet vạn vật và các hệ thống thực - ảo

5.14.1 Yêu cầu chung

AI ngày càng được sử dụng như một thành phần trong các hệ thống nhúng, chẳng hạn như Internet vạn vật và hệ thống vật lý – mạng để phân tích các luồng thông tin về thế giới thực từ các bộ cảm biến, hoặc đưa ra các dự đoán và quyết định về các quá trình vật lý, sử dụng để gửi lệnh cho các cơ cấu chấp hành điều khiển hoặc tác động đến các quá trình vật lý đó.

5.14.2 Internet vạn vật

IoT là hạ tầng các thực thể, con người, hệ thống và tài nguyên thông tin được kết nối với nhau cùng với các dịch vụ xử lý và phản hồi thông tin trong thế giới thực và thế giới ảo (3.1.8). Về cơ bản, một hệ thống IoT là một mạng lưới các nút có các bộ cảm biến để đo lường các thuộc tính vật lý các thực thể và truyền dữ liệu liên quan đến các phép đo đó, và cả các bộ truyền động để thay đổi các thuộc tính của các thực thể vật lý đáp ứng với đầu vào số.

Theo dõi y tế và giám sát trạng thái khí quyển là các ví dụ về hệ thống IoT, trong đó đầu ra của hệ thống là thông tin hỗ trợ con người trong các hoạt động phù hợp (ví dụ: cảnh báo cho nhân viên y tế, dự báo thời tiết).

Hệ thống IoT liên quan đến các hệ thống CNTT kết nối mạng và tương tác với các thực thể vật lý. Nền tảng của IoT hệ thống là các thiết bị IoT, phổ biến nhất ở dạng thiết bị cảm biến và thiết bị truyền động để tương tác với các thực thể. Một bộ cảm biến có thể đo một hoặc nhiều thuộc tính của một hoặc nhiều

thực thể vật lý và dữ liệu đầu ra được truyền qua mạng. Cơ cấu chấp hành thay đổi một hoặc nhiều thuộc tính của vật lý thực thể tương ứng với dữ liệu đầu vào nhận được qua mạng. Vai trò của các bộ cảm biến và thiết bị truyền động có thể thay thế cho nhiều dạng thiết bị, chẳng hạn như nhiệt kế, gia tốc kế, máy quay video, micrô, rơ-le, lò sưởi, người máy hoặc thiết bị công nghiệp để sản xuất hoặc kiểm soát quá trình. Xem TCVN 13117:2020 ISO/IEC 30141 để biết thêm thông tin.

AI có thể đóng một vai trò quan trọng trong bối cảnh ứng dụng của các hệ thống IoT. Nó bao gồm việc phân tích về dữ liệu và ra quyết định để có thể hỗ trợ đạt được các mục tiêu của hệ thống, chẳng hạn như kiểm soát các thực thể vật lý và các quá trình vật lý, bằng cách cung cấp thông tin theo ngữ cảnh, theo thời gian thực và thông tin dự báo.

5.14.3 Các hệ thống thực - ảo

CPS là các hệ thống tương tự như IoT nhưng trong đó logic điều khiển được áp dụng cho đầu vào được lấy từ các bộ cảm biến để chỉ đạo hoạt động của cơ cấu chấp hành và từ đó ảnh hưởng đến các quá trình diễn ra trong thế giới thực.

Người máy là một ví dụ về hệ thống CPS mà tại đó đầu vào cảm biến được sử dụng trực tiếp để điều khiển các hoạt động của người máy và thực hiện các hành động trong thế giới thực.

Khoa học người máy bao gồm tất cả các hoạt động liên quan đến thiết kế, lắp ráp, sản xuất, điều khiển và sử dụng người máy cho các loại ứng dụng khác nhau. Người máy bao gồm các thành phần điện tử, cơ khí, phần sụn và phần mềm tương tác chặt chẽ với nhau để đạt được các mục tiêu đặt ra cho một ứng dụng cụ thể. Người máy thường có các bộ cảm biến để đánh giá tình trạng hiện tại của chúng, bộ vi xử lý cung cấp khả năng kiểm soát thông qua phân tích và lập kế hoạch cho các hành động, cơ cấu chấp hành để thực hiện các hành động. Người máy công nghiệp đặt trong các dây chuyền sản xuất tự động theo công đoạn được lập trình để lặp lại một cách chính xác các quỹ đạo và hành động giống nhau mà không bị sai lệch. Người máy dịch vụ hoặc người máy công tác cần phải thích ứng với các tình huống thay đổi và môi trường động. Lập trình tính linh hoạt này là một thách thức khó khăn vì liên quan đến nhiều yếu tố biến động.

Các thành phần của hệ thống AI có thể đóng vai trò là một phần của phần mềm điều khiển và quy trình lập kế hoạch thông qua mô hình "cảm nhận, lập kế hoạch, hành động". Do đó nó cho phép người máy điều chỉnh khi xuất hiện các trở ngại hoặc đối tượng mục tiêu đã di chuyển. Các thành phần ghép nối giữa người máy và hệ thống AI cho phép tương tác thực một cách tự động với các đối tượng, môi trường và con người.

5.15 Tính đáng tin cậy

5.15.1 Yêu cầu chung

Tính đáng tin cậy của hệ thống AI đề cập đến các đặc điểm giúp các bên liên quan hiểu được liệu hệ thống AI có đáp ứng kỳ vọng của họ hay không. Những đặc điểm này có thể giúp các bên liên quan xác

minh rằng:

- Hệ thống AI đã được thiết kế và thẩm định phù hợp theo các quy tắc và tiêu chuẩn hiện đại. Điều này ngụ ý về đảm bảo chất lượng và độ mạnh mẽ;
- Hệ thống AI được xây dựng vì lợi ích của các bên liên quan có mục tiêu phù hợp nhau. Điều này ngụ ý nhận thức về hoạt động của các thuật toán AI và sự hiểu biết về hoạt động tổng thể của các bên liên quan. Nó cũng bao hàm sự đảm bảo đủ điều kiện hoặc chứng nhận về sự phát triển và hoạt động của AI phù hợp với các yêu cầu pháp lý và tiêu chuẩn chuyên ngành khả dụng;
- Hệ thống AI được cung cấp với việc chỉ định đúng các bên chịu trách nhiệm và chịu trách nhiệm giải trình;
- Các hệ thống AI được phát triển và vận hành có cân nhắc đến các mối quan tâm phù hợp của phạm vi, lĩnh vực sử dụng.

Tham khảo ISO/IEC TR 24028 để biết thêm thông tin.

5.15.2 Độ bền vững của AI

Đối với các hệ thống AI, độ bền vững mô tả khả năng duy trì mức hiệu năng của chúng theo dự kiến của các nhà phát triển trong bất kỳ trường hợp nào. Một ví dụ về độ bền vững là khả năng thực thi của một hệ thống trong các giới hạn có thể chấp nhận được bất chấp các điều kiện bên ngoài hoặc sự khắc nghiệt của môi trường. Độ bền vững có thể bao gồm các thuộc tính khác như khả năng phục hồi và độ tin cậy. Hoạt động chắc chắn của một hệ thống AI liên quan đến hoặc dẫn đến tính chất an toàn đối với các bên liên quan trong một môi trường hoặc bối cảnh nhất định (xem ISO/IEC TR 24028).

Ví dụ: một hệ thống AI dựa trên ML mạnh mẽ có thể có khả năng khái quát hóa các nhiễu đầu vào, chẳng hạn như không có mô hình phù hợp. Để đạt được độ bền vững thì có một lựa chọn là huấn luyện mô hình hoặc các mô hình sử dụng bộ dữ liệu huấn luyện lớn, bao gồm cả dữ liệu huấn luyện nhiễu (xem ISO/IEC TR 24018).

Các thuộc tính của độ bền vững chứng minh khả năng (hoặc không có khả năng) hệ thống có hiệu năng có thể so sánh được trong các trường hợp hệ thống hoạt động với dự liệu đầu vào không điển hình, và trái ngược với nó là dữ liệu được mong đợi trong các hoạt động điển hình, hoặc là dữ liệu khác với dữ liệu mà hệ thống được huấn luyện (xem ISO/IEC TR 24029-1).

Khi xử lý dữ liệu đầu vào, hệ thống AI dự kiến sẽ tạo ra các dự đoán (kết quả đầu ra của nó) có thể chấp nhận được, nhất quán hoặc hiệu quả trong một vài phạm vi nào đó. Ngay cả khi một hệ thống có các kết quả đầu ra được coi là không lý tưởng thì hệ thống đó vẫn có thể được coi là mạnh mẽ. Hệ thống AI có đầu ra không nằm trong phạm vi chấp nhận được, không nhất quán và không hiệu quả khi xử lý dữ liệu đầu vào thì hệ thống đó không thể được coi là mạnh mẽ.

Độ mạnh mẽ có thể được coi là khác nhau đối với các loại hệ thống AI khác nhau, chẳng hạn như:

- Độ mạnh mẽ của mô hình hồi quy là khả năng có những khoảng cách chấp nhận về biên độ các

đáp ứng đối với bất kỳ đầu vào hợp lệ nào.

- Độ bền vững một mô hình phân lớp là khả năng tránh chèn các lỗi phân lớp mới khi chuyển từ các đầu vào điển hình sang các đầu vào có thông số thay đổi trong một phạm vi nhất định của chúng.

5.15.3 Tính tin cậy của AI

Tính tin cậy là khả năng của một hệ thống hoặc một thực thể trong hệ thống đó thực hiện các chức năng cần thiết của nó trong các điều kiện đã nêu trong một khoảng thời gian cụ thể (xem ISO/IEC 27040).

Tính tin cậy của hệ thống AI đề cập đến khả năng cho phép nó đưa ra dự đoán cần thiết (3.1.27), đề xuất và quyết định một cách chính xác, nhất quán trong giai đoạn vận hành (6.2.6).

Tính tin cậy có thể bị ảnh hưởng bởi ít nhất là bởi độ bền vững, tính tổng quát, tính nhất quán và khả năng phục hồi của hệ thống AI. Tất cả các đầu vào và cài đặt môi trường đáp ứng các tiêu chí đã nêu phải được xử lý chính xác trong quá trình hoạt động của nó. Một số đầu vào có thể khác với những đầu vào được sử dụng trong giai đoạn phát triển của nó, nhưng có thể xảy ra khi hệ thống được sử dụng. Việc sao lưu hệ thống AI hoặc một thành phần nào đó cũng cải thiện độ tin cậy, điều này sẽ giúp thực thi các logic nghiệp vụ hoạt động giống với bản gốc. Nó hoạt động khi hệ thống AI bị lỗi.

Tính tin cậy có thể hỗ trợ sự an toàn về chức năng của hệ thống AI, theo nghĩa là các hoạt động và bảo vệ tự động được yêu cầu (bởi các bên liên quan) đối với hệ thống hoặc một phần của hệ thống chống lại sự cố đã được xác định.

5.15.4 Khả năng phục hồi của AI

Khả năng phục hồi là khả năng hệ thống phục hồi tình trạng hoạt động một cách nhanh chóng sau khi xảy ra sự cố. Khả năng chịu lỗi là khả năng hệ thống tiếp tục hoạt động khi xảy ra gián đoạn, lỗi và hỏng hóc làm suy giảm năng lực của hệ thống.

Hệ thống AI cũng như các hệ thống phần mềm khác, lỗi hỏng phần cứng có thể ảnh hưởng đến hoạt động hoàn hảo của các thuật toán.

Tính tin cậy liên quan đến khả năng phục hồi, nhưng các cấp độ dịch vụ dự kiến và theo kỳ vọng là khác nhau. Kỳ vọng về khả năng phục hồi có thể thấp hơn theo định nghĩa của các bên liên quan. Hệ thống có khả năng phục hồi có thể đưa ra cấp độ suy giảm hoạt động khi xuất hiện một số dạng hư hỏng mà các bên liên quan có thể chấp nhận được. Hệ thống phục hồi cần phải có các phương pháp phục hồi phù hợp theo yêu cầu.

5.15.5 Khả năng điều khiển AI

Khả năng điều khiển là thuộc tính của hệ thống AI cho phép tác nhân bên ngoài có thể can thiệp vào hoạt động của nó. Khả năng điều khiển có được bằng việc cung cấp các cơ chế đáng tin cậy qua đó tác nhân có thể tiếp quản quyền điều khiển hệ thống AI.

Khía cạnh quan trọng của khả năng điều khiển là việc xác định (các) tác nhân nào có thể điều khiển các thành phần nào của hệ thống AI (ví dụ: nhà cung cấp dịch vụ hoặc nhà cung cấp sản phẩm, nhà cung cấp các hợp phần của hệ thống AI, người dùng hoặc tổ chức có thẩm quyền quản lý).

Thông tin thêm về khả năng điều khiển có thể xem trong ISO/IEC TR 24028: 2020, 10.4.

5.15.6 Tính diễn giải của AI

Tính diễn giải là thuộc tính của hệ thống AI, có nghĩa là các yếu tố quan trọng tác động đến quyết định được thể hiện theo cách mà con người có thể hiểu được. Tính diễn giải đặc biệt quan trọng khi các quyết định do hệ thống AI đưa ra ảnh hưởng đến một hoặc nhiều người. Người có trách nhiệm thường không tin tưởng vào một quyết định nào đó trừ khi họ có thể hiểu được cách thức đưa ra quyết định, đặc biệt khi quyết định đó theo cách thức gây bất lợi cho họ ở cấp độ cá nhân (ví dụ: từ chối đơn xin vay tại ngân hàng).

Tính diễn giải có thể là phương tiện hữu ích để thẩm định một hệ thống AI, ngay cả khi các quyết định không ảnh hưởng trực tiếp đến con người. Ví dụ: nếu một hệ thống AI đang phân tích hình ảnh của một cảnh quan để xác định các thực thể trong cảnh quan đó; các diễn giải về lý do đưa ra quyết định về các nội dung có trong cảnh quan là hữu hích, và coi đó là cách thức để kiểm tra những gì được nhận dạng và những gì cần phải xem xét thêm. Trong lịch sử phát triển các hệ thống AI từng đã có những ví dụ phản bác các quyết định mà không có những giải thích phù hợp, do vậy người ta cho rằng hệ thống AI đang nhận dạng các thực thể trong cảnh quan dựa vào tương quan giả tồn tại trong dữ liệu huấn luyện.

Thực thi tính diễn giải trong một số loại hình hệ thống AI này có thể dễ dàng hơn với những hệ thống khác. Mạng nơ-ron học sâu có thể phát sinh các vấn đề vì sự phức tạp của hệ thống có thể khiến nó khó đưa ra lời giải thích có ý nghĩa về cách thức hệ thống đưa ra một quyết định.

Các thuật toán dựa trên quy tắc, chẳng hạn như các phương pháp biểu trưng hoặc cây quyết định được coi là có tính diễn giải cao vì những quy tắc đó trực tiếp cung cấp các diễn giải. Tuy nhiên tính diễn giải có thể trở nên khó hiểu hơn nếu các mô hình phát triển hơn nữa về quy mô và độ phức tạp.

5.15.7 Tính dự đoán của AI

Tính dự đoán là thuộc tính của hệ thống AI cho phép các bên liên quan đưa ra giả định đáng tin cậy về kết quả đầu ra. Tính dự đoán đóng một vai trò quan trọng trong khả năng chấp nhận các hệ thống AI và thường được đề cập trong các cuộc tranh luận về đạo đức liên quan đến các hệ thống AI. Sự tin tưởng vào công nghệ thường dựa trên tính dự đoán: một hệ thống được tin cậy nếu người dùng có thể suy luận hệ thống sẽ hoạt động như thế nào trong một tình huống cụ thể, ngay cả khi người dùng không thể giải thích các yếu tố đằng sau hành vi của hệ thống. Ngược lại, người dùng có thể ngừng tin tưởng nếu hệ thống hoạt động bất thường trong các tình huống mà câu trả lời đúng dường như là hiển nhiên.

Tuy nhiên tồn tại một số vấn đề với quan niệm mơ hồ về tính dự đoán, dựa vào ý tưởng cho rằng con người có thể dự đoán hành vi của một hệ thống AI:

- Một định nghĩa trực tiếp dựa trên sự hiểu biết của con người vốn dĩ mang tính chủ quan. Định nghĩa tính dự đoán nên sử dụng các tiêu chí khách quan, có thể định lượng được.
- Có thể thiết lập niềm tin vào một hệ thống AI ngay cả khi một (con) người không thể dự đoán hành vi chính xác của nó trong mọi tình huống. Một đảm bảo mang tính thống kê về sự phù hợp trong hành vi của nó có thể hữu ích hơn. Lý do đằng sau tuyên bố này là nhiều phương pháp học máy tạo ra những kết quả không thể đoán trước được.

Khả năng dự đoán gắn liền với độ chính xác. Các phương pháp cải thiện độ chính xác có thể làm giảm khả năng hệ thống AI tạo ra kết quả không thể đoán trước.

5.15.8 Tính minh bạch của AI

Tính minh bạch của các hệ thống AI hỗ trợ các mục tiêu lấy con người làm trung tâm, là chủ đề của các nghiên cứu và thảo luận đang tiến hành. Cung cấp tính minh bạch về hệ thống AI có thể liên quan đến việc truyền đạt các thông tin phù hợp về hệ thống cho các bên liên quan (ví dụ: mục tiêu, giới hạn đã biết, định nghĩa, lựa chọn thiết kế, giả định, tính năng, mô hình, thuật toán, phương pháp huấn luyện và quy trình đảm bảo chất lượng). Ngoài ra, tính minh bạch của hệ thống AI có thể liên quan đến việc thông báo cho các bên liên quan chi tiết dữ liệu được sử dụng (ví dụ: cái gì, ở đâu, khi nào, tại sao dữ liệu được thu thập và cách nó được sử dụng) để tạo ra hệ thống và bảo vệ dữ liệu cá nhân cùng với mục đích hệ thống và cách nó được xây dựng và triển khai. Tính minh bạch cũng có thể bao gồm việc thông báo cho các bên liên quan về quá trình xử lý và mức độ tự động hóa được sử dụng để đưa ra các quyết định liên quan.

CHÚ THÍCH: Việc tiết lộ một số thông tin nhằm đáp ứng tính minh bạch có thể đi ngược lại các yêu cầu về bảo mật, quyền riêng tư hoặc tính bí mật.

5.15.9 Sự thiên vị và công bằng trong AI

Theo nghĩa tổng quan, ý nghĩa của thuật ngữ thiên vị phụ thuộc vào ngữ cảnh của nó.

Trong AI, thuật ngữ thiên vị đề cập đến ý tưởng rằng các trường hợp khác nhau yêu cầu cách đối xử khác nhau. Theo nghĩa này, sự thiên vị cho phép các hệ thống học máy đánh giá rằng một tình huống này là khác với tình huống khác và có cách hành xử khác nhau cho phù hợp. Do đó, sự thiên vị là cơ sở cho quá trình học máy và để điều chỉnh hành vi cho phù hợp với tình huống cụ thể.

Tuy nhiên trong ngữ cảnh xã hội, thuật ngữ thiên vị thường đề cập đến quan điểm cho rằng những khác biệt nhất định trong đối xử là không công bằng. Để tránh nhầm lẫn, trong ngữ cảnh của AI thay vì thiên vị, thuật ngữ không công bằngđược sử dụng để chỉ sự đối xử khác biệt không hợp lý, có lợi cho một số nhóm nhất định hơn so với những nhóm khác.

Hành vi không lành mạnh của hệ thống AI có thể dẫn đến việc không tôn trọng thực tế, niềm tin và chuẩn mực đã được thiết lập, dẫn đến thiên vị hoặc phân biệt đối xử.

Mặc dù sự thiên vị nhất định là cần thiết để vận hành hệ thống AI phù hợp, nhưng sự thiên vị không

mong muốn có thể được đưa vào hệ thống AI một cách vô ý và có thể dẫn đến kết quả tạo ra một hệ thống không công bằng. Các nguồn gây ra sự thiên vị không mong muốn trong các hệ thống AI có liên quan lẫn nhau và bao gồm thiên vị về nhận thức của con người, thiên vị dữ liệu và thiên vị đưa ra bởi các quyết định kỹ thuật. Sự thiên vị trong dữ liệu huấn luyện là nguồn gốc chính gây ra sự thiên vị trong các hệ thống AI. Những thiên vị về nhận thức của con người có thể ảnh hưởng đến các quyết định về thu thập và xử lý dữ liệu, thiết kế hệ thống, đào tạo mô hình và các quyết định phát triển khác.

Giảm thiểu thiên vị không mong muốn trong các hệ thống AI là một mục tiêu đầy thách thức, nhưng việc phát hiện và xử lý nó là hoàn toàn khả thi [ISO/IEC TR 24027: 2021].

5.16 Xác minh và thẩm định trong AI

Xác minh là sự xác nhận rằng một hệ thống được xây dựng đúng và đáp ứng các yêu cầu cụ thể.

Thẩm định là sự xác nhận thông qua việc cung cấp bằng chứng khách quan rằng các yêu cầu cho một mục đích sử dụng hoặc ứng dụng cụ thể đã được đáp ứng. Các cân nhắc trong việc xác minh và xác nhận bao gồm những khía cạnh sau:

- Một số hệ thống hoàn toàn có thể xác minh được (ví dụ: tất cả các thành phần hệ thống có thể được xác minh riêng rẽ như một hệ thống hoàn chỉnh).
- Một số hệ thống có thể xác minh và thẩm định từng phần (ví dụ: ít nhất một thành phần hệ thống có thể được xác minh riêng và các thành phần còn lại của hệ thống có thể được thẩm định như một hệ thống hoàn chỉnh).
- Một số hệ thống không thể xác minh được nhưng lại có thể thẩm định được (ví dụ: không có thành phần hệ thống có thể được xác minh nhưng mọi thành phần hệ thống có thể được thẩm định như một hệ thống hoàn chỉnh).
- Một số hệ thống không thể xác minh được và có thể thẩm định từng phần (ví dụ: không có thành phần hệ thống có thể được xác minh, nhưng ít nhất một thành phần hệ thống có thể được thẩm định riêng).
- Một số hệ thống không thể xác minh và cũng không thể thẩm định được (ví dụ: không có thành phần hệ thống nào có thể được xác minh hoặc thẩm định).

5.17 Các vấn đề pháp lý

Hệ thống AI có thể được triển khai và vận hành ở các khu vực pháp lý khác với các khu vực mà hệ thống được thiết kế hoặc sản xuất. Các nhà phát triển và nhà sản xuất hệ thống AI cần nhận thức rõ các yêu cầu pháp lý hiện hành có thể khác nhau giữa các khu vực tài phán.

Ví dụ: một chiếc ô tô được sản xuất ở một khu vực pháp lý này có thể phải tuân thủ pháp yêu cầu pháp lý ở khu vực khác nếu muốn hoạt động ở đó.

Ngoài ra, các hệ thống AI thông thường yêu cầu dữ liệu phải được thu thập, xử lý và sử dụng trong các giai đoạn phát triển, vận hành của hệ thống AI và được khử bỏ khi ngừng sử dụng.

Các nhà phát triển, nhà sản xuất và người dùng hệ thống AI cần nhận thức rõ các yêu cầu pháp lý đối với việc thu thập, sử dụng và xử lý dữ liệu có thể khác nhau giữa các khu vực tài phán.

Để giảm thiểu tác động của các yêu cầu pháp lý khác nhau, các nhà phát triển và nhà sản xuất hệ thống AI có thể sử dụng một hoặc nhiều biện pháp giảm thiểu sau:

- Lưu ý các yêu cầu pháp lý có thể áp dụng cho hệ thống AI trong giai đoạn chuẩn bị. Điều này bao gồm các yêu cầu pháp lý liên quan đến việc thu thập, sử dụng và khử bỏ dữ liệu.
- Xây dựng kế hoạch tuân thủ các yêu cầu pháp lý trong (các) khu vực tài phán mà hệ thống AI dự kiến triển khai và vận hành.
- Xây dựng kế hoạch giám sát việc tuân thủ các yêu cầu pháp lý trong các giai đoạn phát triển, triển khai, vận hành và ngừng sử dụng của hệ thống AI.
- Xây dựng kế hoạch theo dõi và ứng phó với bất kỳ thay đổi nào về yêu cầu pháp lý.
- Áp dụng các phương pháp thiết kế, triển khai và vận hành linh hoạt.

5.18 Tác động xã hội

Hệ thống AI có một loạt rủi ro, được xác định bởi mức độ nghiêm trọng của tác động tiềm ẩn của một thất bại hoặc hành vi không mong muốn. Các khía cạnh liên quan để đánh giá mức độ rủi ro bao gồm:

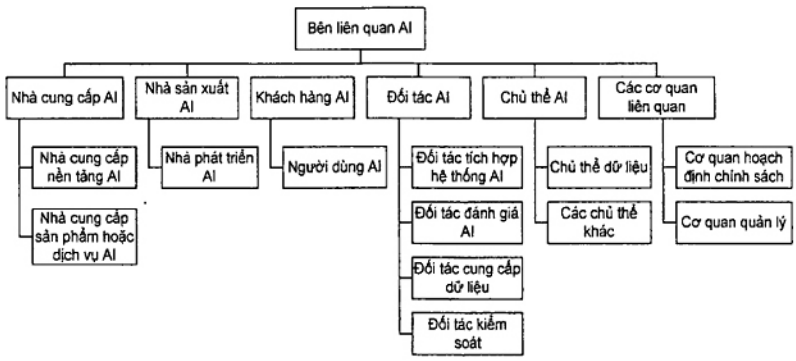
- Loại hình không gian hành động mà hệ thống đang vận hành (ví dụ: các khuyến nghị so với hành động trực tiếp trong một môi trường);
- Sự hiện diện hoặc vắng mặt giám sát bên ngoài;
- Loại hình giám sát bên ngoài (tự động hoặc thủ công);
- Sự liên quan đến khía cạnh đạo đức của tác vụ hoặc lĩnh vực;
- Mức độ minh bạch của các quyết định hoặc khâu xử lý;
- Mức độ tự động hóa của hệ thống.

Ví dụ một hệ thống chỉ đưa ra các khuyến nghị và không thể tự hành động trong một lĩnh vực không liên quan đến đạo đức có thể được coi là rủi ro thấp. Ngược lại, một hệ thống AI có thể được coi là tiêu cực và rủi ro cao nếu các hành động của nó có ảnh hưởng trực tiếp đến cuộc sống của con người, nó hoạt động mà không có sự giám sát từ bên ngoài và việc ra quyết định của nó là không rõ ràng.

CHÚ THÍCH: Đối với các lĩnh vực ứng dụng cụ thể, các yêu cầu pháp lý, chính sách và tiêu chuẩn áp dụng bổ sung có thể vượt ra ngoài phân tích tác động được mô tả trong điều này.

5.19 Vai trò của các bên liên quan đến AI

5.19.1 Yêu cầu chung



Hình 2 – Các vai trò và vai trò phụ của các bên liên quan

Như thể hiện trong Hình 2, AI có thể đòi hỏi vai trò hoặc vai trò phụ của một vài bên liên quan. Các vai trò và vai trò phụ này được mô tả trong các mục từ 5.19.2 đến 5.17.9.

CHÚ THÍCH: Một tổ chức hoặc đơn vị có thể đảm nhận nhiều hơn một vai trò hoặc vai trò phụ.

5.19.2 Nhà cung cấp AI

5.19.2.1 Yêu cầu chung

Nhà cung cấp AI là một tổ chức hoặc thực thể cung cấp các sản phẩm hoặc dịch vụ sử dụng một hoặc nhiều hệ thống AI. Các nhà cung cấp AI bao gồm các nhà cung cấp nền tảng AI và các nhà cung cấp sản phẩm hoặc dịch vụ AI.

5.19.2.2 Nhà cung cấp nền tảng AI

Nhà cung cấp nền tảng AI là một tổ chức hoặc thực thể cung cấp các dịch vụ cho phép các bên liên quan khác sản xuất các dịch vụ hoặc sản phẩm AI.

5.19.2.3 Nhà cung cấp dịch vụ hoặc sản phẩm AI

Nhà cung cấp dịch vụ hoặc sản phẩm AI là một tổ chức hoặc thực thể cung cấp các dịch vụ hoặc sản phẩm AI được khách hàng hoặc người dùng AI trực tiếp sử dụng, hoặc được tích hợp vào một hệ thống sử dụng AI cùng với các thành phần không phải AI.

5.19.3 Nhà sản xuất AI

5.19.3.1 Yêu cầu chung

Nhà sản xuất AI là một tổ chức hoặc thực thể thiết kế, phát triển, thử nghiệm và triển khai các sản phẩm hoặc dịch vụ sử dụng một hoặc nhiều hệ thống AI.

5.19.3.2 Nhà phát triển AI

Nhà phát triển AI là một tổ chức hoặc thực thể liên quan đến phát triển các dịch vụ và sản phẩm AI. Ví dụ về các nhà phát triển AI bao gồm, nhưng không giới hạn như dưới đây:

- Nhà thiết kế mô hình: thực thể nhận dữ liệu, đặc tả vấn đề và tạo mô hình AI;

- Người triển khai mô hình: thực thể nhận mô hình AI và chỉ định tính toán nào sẽ thực thi (triển khai sử dụng và dựa vào tài nguyên tính toán nào, ví dụ: CPU, GPU, ASIC, FPGA);
- Nhà xác minh tính toán: thực thể xác minh rằng tính toán đang được thực thi đúng như thiết kế;
- Nhà xác minh mô hình: thực thể xác minh rằng mô hình AI đang hoạt động đúng như thiết kế.

5.19.4 Khách hàng AI

5.19.4.1 Yêu cầu chung

Khách hàng AI là một tổ chức hoặc thực thể sử dụng sản phẩm hoặc dịch vụ AI trực tiếp hoặc cung cấp nó cho cho người dùng AI.

5.19.4.2 Người dùng AI

Người dùng AI là một tổ chức hoặc thực thể sử dụng các sản phẩm hoặc dịch vụ AI.

5.19.5 Đối tác AI

5.19.5.1 Yêu cầu chung

Đối tác AI là một tổ chức hoặc thực thể cung cấp dịch vụ trong bối cảnh AI. Các đối tác AI có thể thực hiện kỹ thuật phát triển các sản phẩm hoặc dịch vụ AI, tiến hành thử nghiệm và đánh giá các sản phẩm và dịch vụ AI, kiểm soát việc sử dụng AI, đánh giá các sản phẩm hoặc dịch vụ AI và thực hiện các tác vụ khác. Ví dụ về các loại đối tác AI được thảo luận trong các điều khoản dưới đây.

5.19.5.2 Đối tác tích hợp hệ thống AI

Đối tác tích hợp hệ thống AI là một tổ chức hoặc thực thể có liên quan đến việc tích hợp các thành phần AI vào các hệ thống lớn hơn, có khả năng bao gồm các cả các thành phần không phải AI.

5.19.5.3 Đối tác cung cấp dữ liệu

Đối tác cung cấp dữ liệu là một tổ chức hoặc thực thể liên quan đến cung cấp dữ liệu sử dụng bởi các sản phẩm hoặc dịch vụ AI.

5.19.5.4 Đối tác kiểm soát AI

Đối tác kiểm soát AI là một tổ chức hoặc đơn vị liên quan đến việc kiểm soát các tổ chức sản xuất, cung cấp hoặc sử dụng hệ thống AI bằng hoạt động đánh giá sự phù hợp với các tiêu chuẩn, chính sách hoặc yêu cầu pháp lý.

5.19.5.5 Đối tác đánh giá AI

Đối tác đánh giá AI là một tổ chức hoặc thực thể đánh giá hiệu năng của một hoặc nhiều hệ thống AI.

5.19.6 Chủ thể AI

5.19.6.1 Yêu cầu chung

Chủ thể AI là một tổ chức hoặc thực thể bị tác động bởi hệ thống, dịch vụ hoặc sản phẩm AI.

5.19.6.2 Chủ thể dữ liệu

Chủ thể dữ liệu là một tổ chức hoặc thực thể bị ảnh hưởng bởi các hệ thống AI với các khía cạnh sau:

- Chủ thể của dữ liệu huấn luyện: khi dữ liệu liên quan đến một tổ chức hoặc con người được sử dụng để đào tạo một hệ thống AI, có thể có những liên can đến bảo mật và quyền riêng tư, đặc biệt khi đối tượng đó là một cá nhân.

5.19.6.3 Chủ thể khác

Ví dụ, các tổ chức hoặc thực thể khác bị ảnh hưởng bởi hệ thống AI, dịch vụ hoặc sản phẩm có thể ở dạng cá nhân hoặc cộng đồng. Ví dụ: người tiêu dùng tương tác với mạng xã hội cung cấp các khuyến nghị dựa trên AI, người điều khiển phương tiện di chuyển có tự động hóa dựa trên AI.

5.19.7 Các cơ quan có liên quan

5.19.7.1 Yêu cầu chung

Các cơ quan có liên quan là các tổ chức hoặc thực thể có thể tác động ảnh hưởng đến hệ thống, dịch vụ hoặc sản phẩm AI.

5.19.7.2 Cơ quan hoạch định chính sách

Các tổ chức và thực thể có thẩm quyền thiết lập các chính sách trong phạm vi quốc tế, khu vực, quốc gia, lĩnh vực công nghiệp có thể tác động ảnh hưởng đến hệ thống, dịch vụ hoặc sản phẩm AI.

5.19.7.3 Cơ quan quản lý

Các tổ chức và thực thể có thẩm quyền thiết lập, thực hiện và thực thi các yêu cầu pháp lý theo chính sách do các nhà hoạch định chính sách đưa ra (5.17.9.2).

6 Vòng đời hệ thống AI

6.1 Mô hình vòng đời hệ thống AI

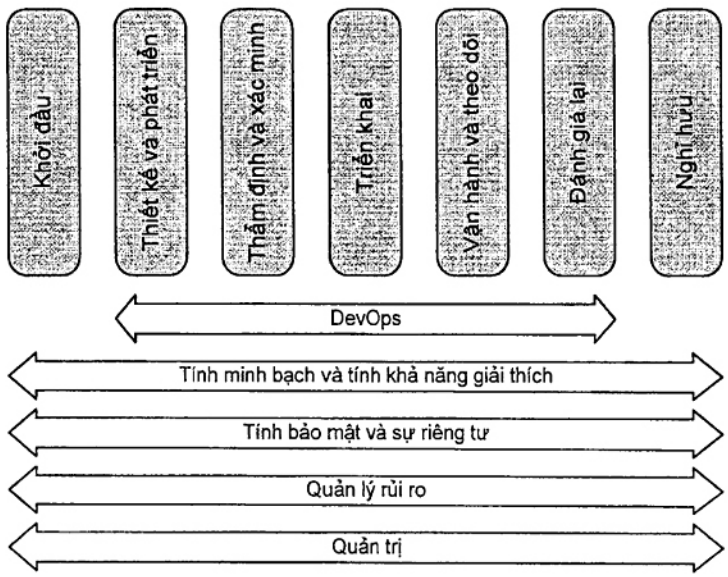
Mô hình vòng đời của hệ thống AI mô tả sự phát triển hệ thống AI từ khi bắt đầu cho đến khi ngừng sử dụng. Tiêu chuẩn này không quy định một mô hình vòng đời cụ thể nhưng nhấn mạnh một số quá trình đặc thù cho các hệ thống AI có thể xảy ra trong vòng đời của chúng. Các quá trình và tiến trình cụ thể có thể xảy ra trong một hoặc nhiều giai đoạn hoặc từng giai đoạn riêng của vòng đời và có thể lặp lại trong suốt quá trình tồn tại của hệ thống. Ví dụ các giai đoạn "thiết kế, phát triển" và "triển khai" được lặp lại nhiều lần cho hoạt động phát triển, sửa lỗi và cập nhật của một hệ thống.

Mô hình vòng đời hệ thống giúp các bên liên quan xây dựng hệ thống AI hiệu quả và cho phép tăng cường hơn nữa tính hiệu quả. Các tiêu chuẩn quốc tế như ISO/IEC 15288 cho hệ thống tổng thể, ISO/IEC TCVN 10539:2014 ISO/IEC 12207 cho phần mềm và ISO/IEC 15289 cho tài liệu hóa hệ thống sẽ rất hữu ích trong việc phát triển mô hình vòng đời. Các tiêu chuẩn quốc tế này mô tả các quy trình vòng đời cho các hệ thống nói chung và không chỉ định cụ thể cho các hệ thống AI.

Hình 3 đưa ra ví dụ về các giai đoạn và quá trình ở mức cao có thể được áp dụng trong vòng đời của hệ thống AI. Các giai đoạn và quy trình này có thể được tiến hành lặp đi lặp lại và thường được yêu cầu thực hiện trong quá trình phát triển và vận hành hệ thống AI. Có một vài cân nhắc khác nhau cần được tính toán trong toàn bộ vòng đời từ đầu đến cuối; ví dụ về những cân nhắc này bao gồm:

- Các hệ quả về quản trị phát sinh từ việc phát triển hoặc sử dụng các hệ thống AI;
- Các hệ quả về quyền riêng tư và bảo mật do việc sử dụng lượng lớn dữ liệu, một số dữ liệu trong đó có bản chất nhạy cảm;
- Các mối đe dọa bảo mật phát sinh từ việc phát triển hệ thống phụ thuộc vào dữ liệu;
- Các khía cạnh về tính minh bạch và tính diễn giải, bao gồm nguồn gốc dữ liệu và khả năng cung cấp các giải thích về cách xác định đầu ra của hệ thống AI.

Hình 3 cho thấy một ví dụ về các giai đoạn của mô hình vòng đời hệ thống AI và các quá trình ở mức cao. Phụ lục A trình bày cách mô hình vòng đời hệ thống AI này ánh xạ tới định nghĩa vòng đời hệ thống AI của OECD.



Hình 3 - Ví dụ về các giai đoạn của mô hình vòng đời hệ thống AI và các quá trình ở mức cao

Hệ thống AI khác với các loại hệ thống khác ở chỗ nó có thể tác động đến các quá trình của mô hình vòng đời, ví dụ như:

- Hầu hết các hệ thống phần mềm được lập trình để hoạt động theo những cách thức được xác định một cách chính xác và được điều khiển bởi các yêu cầu và thông số kỹ thuật của chúng. Các hệ thống AI dựa trên học máy sử dụng các phương pháp huấn luyện và tối ưu hóa dựa vào dữ liệu để xử lý các đầu vào biến đổi rộng.

- Các ứng dụng phần mềm truyền thống thường mang tính tiên đoán, điều này ít xảy ra hơn đối với các hệ thống AI.
- Các ứng dụng phần mềm truyền thống thường có thể xác minh được, trong khi việc đánh giá hiệu năng của các hệ thống AI thường yêu cầu các phương pháp thống kê và việc xác minh chúng có thể là những thách thức.
- Các hệ thống AI thường cần nhiều lần cải tiến để đạt được mức hiệu năng chấp nhận được.

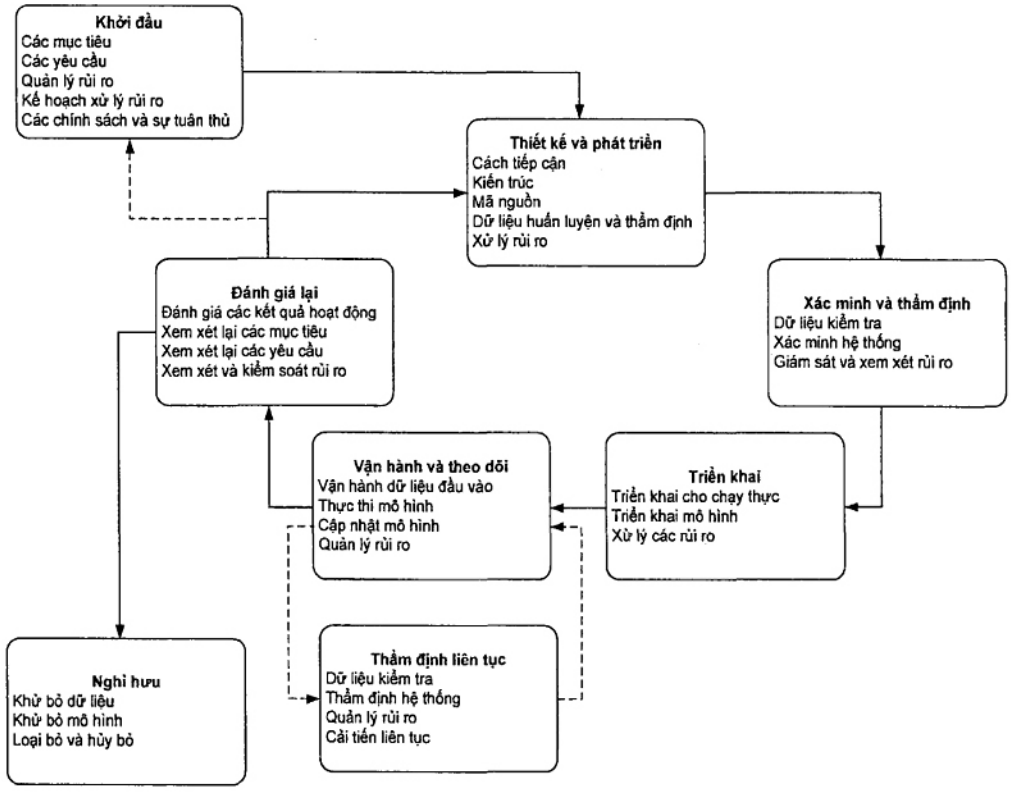
Quản lý dữ liệu (bao gồm các quá trình và công cụ để thu thập, chú giải, chuẩn bị, kiểm tra chất lượng, lấy mẫu, tăng cường dữ liệu) là một khía cạnh chính của hệ thống AI.

Các quá trình phát triển và thử nghiệm có thể khác nhau đối với các hệ thống AI vì các quá trình này được tạo lập dựa trên cơ sở dữ liệu. Điều này trở thành thách thức đối với các hệ thống AI sử dụng học liên tục (còn được gọi là học không ngừng hoặc học suốt đời), nơi mà hệ thống học ở trạng thái vận hành và yêu cầu tiến hành kiểm tra liên tục.

Quy trình quản lý phát hành cho các hệ thống AI khác với phần mềm truyền thống. Trong khi các ứng dụng phần mềm truyền thống xử lý các chức năng lập phiên bản và khác biệt mã nguồn thì các bản phát hành hệ thống AI bao gồm khác biệt mã nguồn và mô hình, cũng như sự khác biệt về dữ liệu tập huấn nếu sử dụng học máy.

Một số quy trình của vòng đời AI khác với những quy trình trong vòng đời phần mềm truyền thống được đề cập trong mục 6.2.

Hình 4 đưa ra ví dụ mô hình vòng đời cho một hệ thống AI; có thể có các mô hình vòng đời khác nhau dựa trên các kỹ thuật phát triển khác nhau. Hình 4 cho thấy chuỗi các giai đoạn của vòng đời và chỉ ra các quá trình với mỗi giai đoạn được coi là quan trọng đối với các hệ thống AI và cần được xem xét đặc biệt, ngoài những quan tâm về yêu cầu đối với phát triển các hệ thống không phải AI điển hình.



Hình 4 - Ví dụ về mô hình vòng đời hệ thống AI với các quá trình dành riêng cho hệ thống AI

Như thể hiện trong Hình 4, hoạt động phát triển và vận hành hệ thống AI có xu hướng lặp đi lặp lại nhiều hơn so với các hệ thống không sử dụng AI. Hệ thống AI có xu hướng ít có thể dự đoán được hơn và thường cần một số kinh nghiệm vận hành và điều chỉnh hệ thống AI để đáp ứng các mục tiêu của nó.

6.2 Các giai đoạn và quá trình trong vòng đời của hệ thống AI

6.2.1 Yêu cầu chung

Các quá trình được mô tả trong mỗi giai đoạn là những ví dụ đại diện vì các quá trình cụ thể phụ thuộc vào hệ thống AI. Các quá trình có thể được thực hiện theo các trình tự khác nhau và trong một số trường hợp chúng được thực hiện song song.

Bản thân các quy trình này không nhất thiết phải dành riêng cho AI, nhưng các vấn đề liên quan đến AI khiến chúng có tầm quan trọng đặc biệt trong bối cảnh này.

6.2.2 Khởi đầu

Khởi đầu xảy ra khi một hoặc nhiều bên liên quan quyết định biến một ý tưởng thành một hệ thống hữu hình. Giai đoạn khởi đầu có thể bao gồm một số quá trình và quyết định dẫn đến việc quyết định tiếp tục giai đoạn thiết kế và phát triển. Giai đoạn khởi đầu có thể được xem lại trong vòng đời khi thông tin mới

được phát hiện ở các giai đoạn sau. Ví dụ, có thể phát hiện ra rằng hệ thống không khả thi về mặt kỹ thuật hoặc tài chính. Ví dụ về các quá trình có thể xảy ra trong giai đoạn khởi đầu bao gồm:

Mục tiêu: Các bên liên quan nên xác định lý do tại sao cần phát triển một hệ thống AI. Hệ thống giải quyết vấn đề gì? Hệ thống giải quyết nhu cầu khách hàng hoặc cơ hội kinh doanh nào? Các thước đo cho sự thành công là gì?

Yêu cầu: Các bên liên quan cần tập hợp các yêu cầu đối với hệ thống AI trải dài trong suốt vòng đời của hệ thống AI. Việc không xem xét đầy đủ các yêu cầu về triển khai, vận hành và ngừng sử dụng có thể dẫn đến các vấn đề trong tương lai. Phương pháp tiếp cận đa bên liên quan bao gồm chuyên môn đa dạng về các chủ đề có thể giúp xác định các rủi ro tiềm ẩn và hậu quả không mong muốn của hệ thống. Các bên liên quan phải đảm bảo các yêu cầu của hệ thống AI đáp ứng các mục tiêu của hệ thống AI. Các yêu cầu cần tính đến rằng nhiều hệ thống AI không thể dự đoán được và các ảnh hưởng tới việc hệ thống đạt được các mục tiêu đề ra. Các bên liên quan nên xem xét các yếu tố về quy định và đảm bảo sự phát triển và vận hành của hệ thống AI tuân thủ các chính sách bắt buộc một cách tương xứng.

Quản lý rủi ro: Các tổ chức nên đánh giá rủi ro liên quan đến AI trong toàn bộ vòng đời của hệ thống AI. Đầu ra của hoạt động này phải là một kế hoạch xử lý rủi ro. Quản lý rủi ro, bao gồm việc xác định, đánh giá và xử lý rủi ro liên quan đến AI được mô tả trong ISO/IEC 23894.

Các tổ chức nên xác định những tác hại và lợi ích tiềm ẩn liên quan đến hệ thống AI bao gồm cả việc hỏi ý kiến đại diện người dùng. Quá trình này có thể mang lại những thông tin có giá trị dùng để hướng dẫn phát triển các phần của hệ thống bao gồm các tính năng, giao diện người dùng, tài liệu và cách sử dụng. Các tổ chức nên nghiên cứu sâu hơn và căn chỉnh các giá trị đến mức độ chúng có thể trở thành một phần các yêu cầu cho hệ thống. Các khuôn khổ pháp lý, quyền con người, trách nhiệm xã hội và môi trường có thể giúp căn chỉnh và mô tả các giá trị của hệ thống.

Ngoài các rủi ro điển hình được xem xét đối với hệ thống chẳng hạn như bảo mật và quyền riêng tư, các rủi ro liên quan đến các giá trị đã được xác định cũng cần được lập kế hoạch giải quyết.

Tính minh bạch và trách nhiệm giải trình: Các bên liên quan phải đảm bảo rằng các mối quan tâm trong suốt vòng đời như nguồn gốc dữ liệu, tính hợp lệ của nguồn dữ liệu, nỗ lực giảm thiểu rủi ro, các quá trình và quyết định được thực thi đều phải được ghi lại để hỗ trợ hiểu biết mang tính toàn diện về cách thức có được kết quả đầu ra của hệ thống AI cũng như cho mục đích thực thi trách nhiệm giải trình.

Chi phí và nguồn vốn: Các bên liên quan nên dự toán chi phí của hệ thống AI cho suốt vòng đời và đảm bảo sự sẵn sàng của nguồn vốn.

Nguồn lực: Các bên liên quan nên xác định những nguồn lực nào được yêu cầu để thực hiện và hoàn thành từng giai đoạn của vòng đời và họ cần đảm bảo các nguồn lực sẽ luôn sẵn sàng khi cần thiết.

Cần đưa ra mối quan tâm về dữ liệu có thể yêu cầu cho phát triển hoặc đánh giá một hệ thống AI. Đối với hệ thống AI dựa trên ML, cần đặc biệt xem xét việc huấn luyện, thẩm định và kiểm tra dữ liệu.

Tính khả thi: Giai đoạn khởi đầu dẫn đến quyết định liệu hệ thống AI có khả thi hay không. Có thể tiến hành quá trình chứng minh tính khả thi để xác định xem hệ thống có đáp ứng các mục tiêu và yêu cầu đề ra hay không. Ví dụ về các mục tiêu và yêu cầu có thể bao gồm:

- Giải quyết vấn đề đã xác định,
- Giải quyết cơ hội kinh doanh hoặc hoàn thành một sứ mệnh,
- Đáp ứng các thuộc tính và năng lực được chỉ định.

Nếu hệ thống AI được coi là khả thi, các bên liên quan có thể quyết định tiến hành giai đoạn phát triển.

6.2.3 Thiết kế và phát triển

Giai đoạn thiết kế và phát triển tạo ra hệ thống AI và khẳng định một hệ thống AI đã sẵn sàng để xác minh và thẩm định. Trong giai đoạn này và đặc biệt là trước khi kết thúc, các bên liên quan phải đảm bảo hệ thống AI đáp ứng các mục tiêu, yêu cầu ban đầu và các mục tiêu khác được xác định trong giai đoạn khởi đầu. Ví dụ về các quá trình có thể xảy ra trong giai đoạn thiết kế và phát triển bao gồm:

Tiếp cận: Các bên liên quan nên xác định một cách tiếp cận tổng thể để thiết kế hệ thống AI, thử nghiệm hệ thống đó và chuẩn bị để hệ thống sẵn sàng được chấp nhận và triển khai. Giai đoạn tiếp cận có thể bao gồm việc cân nhắc xem có cần cả phần cứng và phần mềm hay không, nguồn linh kiện ở đâu (ví dụ: phát triển từ đầu, mua phần cứng bán sẵn, sử dụng phần mềm nguồn mở).

Kiến trúc: Các bên liên quan nên xác định và ghi lại kiến trúc tổng thể của hệ thống AI. Kiến trúc và quá trình tiếp cận có liên quan với nhau và có thể cần phải lặp lại các hoạt động giữa hai quá trình này.

Mã nguồn: Mã nguồn phần mềm cho hệ thống AI được phát triển hoặc mua lại.

Dữ liệu tập huấn: Hệ thống AI bao gồm cả tri thức thu được. Xử lý dữ liệu tập huấn là một phần cơ bản của việc phát triển các hệ thống AI dựa trên học máy (xem 5.10).

Xử lý rủi ro: Các tổ chức nên thực hiện các quá trình và hoạt động kiểm soát được mô tả trong kế hoạch xử lý rủi ro (xem ISO/IEC 23894).

6.2.4 Xác minh và thẩm định

Xác minh và thẩm định kiểm tra xem hệ thống AI từ giai đoạn thiết kế đến phát triển có hoạt động theo đúng yêu cầu và đáp ứng mục tiêu đề ra hay không.

Ví dụ về các quá trình có thể tạo ra từng khâu của xác minh và thẩm định bao gồm:

Xác minh: Phần mềm được kiểm tra chức năng và lỗi đối với bất kỳ phần cứng nào. Kiểm tra tích hợp hệ thống cũng có thể được thực hiện. Có thể tiến hành kiểm tra hiệu năng, kiểm tra thời gian phản hồi, độ trễ hoặc bất kỳ đặc tính hiệu năng liên quan khác của hệ thống AI xem liệu có đáp ứng các yêu cầu cụ thể hay không.

Một khía cạnh quan trọng của hệ thống AI là cần phải xác minh rằng các năng lực của AI có hoạt động

như thiết kế hay không. Điều này yêu cầu thu thập, chuẩn bị và sử dụng dữ liệu kiểm tra. Dữ liệu kiểm tra cần tách biệt với bất kỳ dữ liệu nào khác được sử dụng trong quá trình thiết kế và phát triển và nó cũng cần phải đại diện cho dữ liệu đầu vào mà hệ thống AI dự kiến sẽ xử lý.

Chấp nhận: Các bên liên quan cho rằng hệ thống AI đã hoàn thiện về mặt chức năng và ở mức chất lượng có thể chấp nhận được và sẵn sàng để triển khai.

Giám sát và xem xét rủi ro: Các tổ chức nên xem xét các kết quả xác minh, kiểm tra và thẩm định để nhận biết các sự kiện và điều kiện dẫn đến rủi ro theo kế hoạch xử lý rủi ro (xem ISO/IEC 23894).

6.2.5 Triển khai

Hệ thống AI được cài đặt, phát hành hoặc cấu hình để hoạt động trong môi trường theo mục tiêu đề ra. Các ví dụ về các quá trình của giai đoạn triển khai có thể bao gồm:

Mục tiêu: Hệ thống AI có thể được phát triển trong một môi trường và sau đó triển khai sang môi trường khác. Ví dụ, một hệ thống tự lái có thể phát triển trong phòng thí nghiệm và sau đó được triển khai trên hàng triệu ô tô. Hệ thống AI khác có thể phát triển trên các thiết bị khách và sau đó được triển khai lên đám mây. Đối với một số hệ thống AI, điều quan trọng là phải phân biệt giữa các thành phần phần mềm được triển khai và mô hình có thể triển khai riêng biệt và được phần mềm sử dụng theo thời gian. Phần mềm và mô hình có thể được triển khai mang tính độc lập với nhau.

Xử lý rủi ro: Các tổ chức nên xem xét và cải tiến các quá trình và hoạt động kiểm soát để quản lý rủi ro và nên có chức năng cập nhật kế hoạch xử lý rủi ro (xem ISO/IEC 23894).

6.2.6 Vận hành và theo dõi

Trong giai đoạn vận hành và theo dõi, hệ thống AI hoạt động và sẵn sàng cho sử dụng.

Ví dụ về các quá trình có thể thực thi trong giai đoạn vận hành và theo dõi bao gồm:

Theo dõi: Hệ thống AI được theo dõi trong trạng thái hoạt động bình thường và trạng thái xuất hiện các sự cố bao gồm sự kiện không khả dụng, lỗi chạy chương trình, hỏng hóc. Các sự kiện này được thông báo cho các nhà cung cấp AI liên quan để có các hành động xử lý tương ứng.

Sửa chữa: Nếu hệ thống AI bị lỗi hoặc có thể đang bị lỗi thì việc sửa chữa hệ thống là cần thiết.

Cập nhật: Phần mềm, mô hình và phần cứng hệ thống AI có thể được cập nhật để đáp ứng các yêu cầu mới cũng như cải thiện hiệu năng và độ tin cậy.

Hỗ trợ: Người dùng hệ thống AI được cung cấp bất kỳ hỗ trợ cần thiết nào cần thiết để sử dụng hệ thống.

Theo dõi và xem xét rủi ro: Các tổ chức nên theo dõi các hệ thống AI trong quá trình vận hành để đảm bảo và cải thiện chất lượng và hiệu quả của quá trình quản lý rủi ro (xem ISO/IEC 23894).

6.2.7 Thẩm định liên tục

Nếu hệ thống AI sử dụng tính năng học liên tục thì giai đoạn vận hành và theo dõi được mở rộng thành một giai đoạn xác nhận bổ sung liên tục. Trong giai đoạn này, huấn luyện gia tăng diễn ra liên tục khi hệ thống đang chạy trong quá trình sản xuất. Hoạt động của hệ thống AI liên tục được kiểm tra để đảm bảo sự làm việc một cách chính xác bằng sử dụng dữ liệu kiểm tra. Trong một số trường hợp, dữ liệu kiểm tra tự nó có thể yêu cầu được cập nhật để có tính đại diện cho dữ liệu sản xuất hiện tại và do đó cung cấp đánh giá trung thực hơn về năng lực của hệ thống AI.

Cải tiến liên tục trong quản lý rủi ro: Việc thẩm định liên tục cũng nên sử dụng để cải tiến liên tục các quá trình quản lý rủi ro (xem ISO/IEC 23894).

6.2.8 Đánh giá lại

Sau giai đoạn vận hành và theo dõi, dựa trên kết quả hoạt động của hệ thống AI việc đánh giá lại có thể phát sinh tùy thuộc nhu cầu. Ví dụ về các quá trình có thể thực thi trong giai đoạn đánh giá lại bao gồm:

Đánh giá kết quả hoạt động: Kết quả của hệ thống đang hoạt động cần được đánh giá và định lượng dựa trên các mục tiêu và rủi ro đã được xác định đối với hệ thống AI.

Hoàn thiện các mục tiêu: Nếu hệ thống AI không thể đạt được các mục tiêu ban đầu hoặc các mục tiêu đó cần được sửa đổi dựa trên kinh nghiệm vận hành hệ thống thì sẽ dẫn đến việc phải điều chỉnh các mục tiêu.

Sàng lọc các yêu cầu: Kinh nghiệm hoạt động của hệ thống có thể cung cấp bằng chứng rằng một số yêu cầu ban đầu không phù hợp, điều này có thể dẫn đến việc sàng lọc lại các yêu cầu bằng việc có thể bổ sung các yêu cầu mới hoặc loại bỏ một số yêu cầu ban đầu.

Theo dõi và xem xét rủi ro: Các tổ chức nên theo dõi các sự kiện và điều kiện dẫn đến rủi ro như mô tả trong kế hoạch xử lý rủi ro (xem ISO/IEC 23894).

6.2.9 Ngừng sử dụng

Tại một thời điểm nào đó, hệ thống AI có thể trở nên lỗi thời đến mức việc sửa chữa và cập nhật không đủ tốt để đáp ứng các yêu cầu mới. Ví dụ về các quá trình có thể thực thi trong giai đoạn ngừng sử dụng bao gồm:

Ngừng hoạt động và loại bỏ: Nếu mục đích sử dụng của hệ thống AI không còn tồn tại hoặc xuất hiện một cách tiếp cận tốt hơn, thì hệ thống AI có thể ngừng hoạt động và loại bỏ hoàn toàn, điều này bao gồm đối với cả dữ liệu liên kết với hệ thống.

Thay thế: Nếu mục đích sử dụng của hệ thống AI còn phù hợp, nhưng xuất hiện một cách tiếp cận tốt hơn thì hệ thống AI (hoặc các thành phần của hệ thống AI) có thể được thay thế.

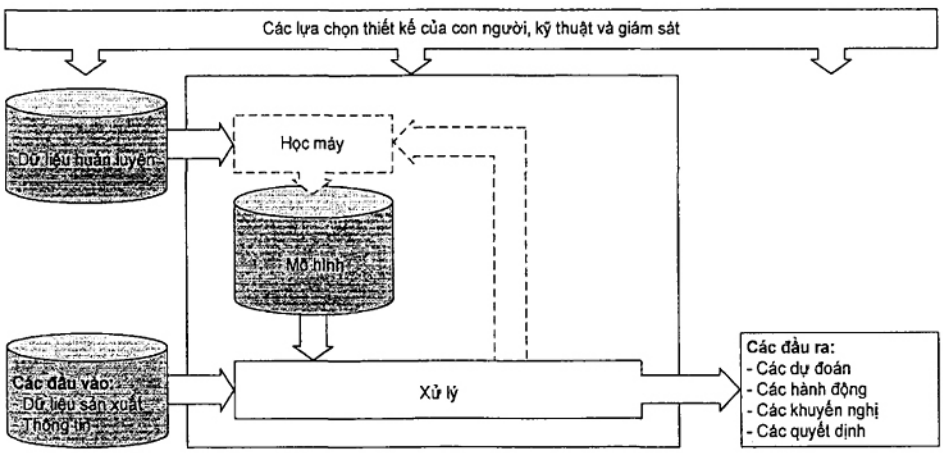
7 Tổng quan về chức năng của hệ thống AI

7.1 Yêu cầu chung

Tiêu chuẩn này AI là hệ thống được thiết kế để tạo ra các kết quả đầu ra như nội dung, dự báo, khuyến nghị hoặc quyết định cho một tập các mục tiêu nhất định do con người đề ra. Hệ thống AI cần con người lựa chọn thiết kế, xây dựng và giám sát. Mức độ giám sát tùy thuộc vào trường hợp sử dụng. Ở mức tối thiểu, sự giám sát thường xuất hiện trong quá trình huấn luyện và thẩm định.

Hoạt động giám sát rất hữu ích để đảm bảo rằng hệ thống AI được phát triển và sử dụng như dự kiến, và các tác động đối với các bên liên quan được xem xét một cách thích đáng trong suốt vòng đời của hệ thống.

Hình 5 mô tả một cách nhìn về chức năng của một hệ thống AI. các đầu vào được xử lý bằng một mô hình để tạo ra đầu ra và mô hình đó có thể được xây dựng trực tiếp hoặc từ việc học trên dữ liệu tập huấn. Các phần được vẽ bằng đường đứt nét dành cho hệ thống AI dựa trên ML.



Hình 5 – Sơ đồ chức năng của hệ thống AI

Mục đích của sơ đồ này nhằm cung cấp mô tả phi kỹ thuật về những gì hệ thống AI làm để đạt được kết quả. Tóm lại, các hệ thống AI chứa một mô hình mà chúng sử dụng để đưa ra các dự đoán và các dự đoán lại được sử dụng để đưa ra các đề xuất, quyết định và hành động có tính kế tiếp nhau, có thể là toàn bộ hoặc một phần tùy thuộc vào hệ thống hoặc con người.

7.2 Dữ liệu và thông tin

Dữ liệu đưa vào hệ thống AI trong quá trình sản xuất và nó được gọi là dữ liệu sản xuất. Dữ liệu đầu vào có thể yêu cầu chuẩn bị trước khi nó hiện diện trong hệ thống AI, chẳng hạn như việc trích xuất các thuộc tính liên quan.

Đầu vào cho hệ thống AI cũng có thể là thông tin thay vì dữ liệu và thường sử dụng cho các tác vụ tối ưu hóa, trong đó đầu vào duy nhất cần thiết là thông tin về những gì sẽ được tối ưu hóa. Một số hệ thống AI hoàn toàn không yêu cầu bất kỳ đầu vào nào, thay vào đó nó thực hiện một tác vụ nhất định theo yêu cầu (ví dụ: tạo các hình ảnh tổng hợp).

Đối với ML, dữ liệu tập huấn được sử dụng để thu thập thông tin về lĩnh vực nó quan tâm và tác vụ cần giải quyết.

Dữ liệu còn có các mục đích sử dụng khác để phát triển và đánh giá các hệ thống AI (xem mục 5.10).

7.3 Tri thức và học tập

Mô hình được hệ thống AI sử dụng để xử lý và giải quyết vấn đề là một biểu diễn tri thức có thể đọc được bằng máy.

Có hai loại tri thức chính: khai báo và thủ tục.

- Tri thức khai báo là về cái gì. Nó rất dễ để diễn đạt bằng lời và sẽ chuyển thành các câu nói. Ví dụ, "nấm mũ tử thần có độc" là tri thức khai báo.

- Tri thức thủ tục là cách thức thực hiện một điều gì đó. Nó thường khó diễn đạt thành lời mà nó được chuyển thành dạng các các thủ tục. Ví dụ, để biết nấm có độc hay không, bạn có thể áp dụng tri thức thủ tục: "Nếu bạn có sổ nấm, hãy tra vào sổ để nhận biết loại nấm của mình. Nếu nhận biết được, cuốn sách sẽ cho bạn biết câu trả lời. Nếu không thể nhận biết được, hãy đến gặp dược sĩ".

Tri thức có nhiều cách thể hiện khác nhau, từ ngầm định cho đến tường minh.

Tri thức cũng có thể đến từ nhiều nguồn khác nhau tùy thuộc vào các thuật toán được sử dụng: nó có thể tồn tại từ trước, nó có thể được thu nhận thông qua quá trình cảm biến và học hỏi, hoặc là sự kết hợp của cả hai quá trình nêu trên.

Hệ thống heuristic: Các hệ thống AI không liên quan đến việc học được gọi là heuristic. Hệ thống chuyên gia cổ điển hoặc hệ thống suy diễn được trang bị một cơ sở tri thức cố định là những ví dụ điển hình. Trong những trường hợp này, các nhà phát triển hệ thống tận dụng kiến thức của con người để đưa ra các quy tắc hợp lý cho hành vi của hệ thống AI.

Hệ thống AI dựa trên ML: Các hệ thống AI liên quan đến việc học được cho là dựa trên học máy. Học tập đòi hỏi các phân tích tính toán tập dữ liệu huấn luyện để phát hiện các kiểu mẫu, xây dựng mô hình và so sánhđầu ra kết quả của mô hình với các hành vi được kỳ vọng. Nó còn được gọi là huấn luyện. Cơ sở tri thức thu nạp là một mô hình được đào tạo dựa trên hàm toán học và tập huấn luyện, thể hiện sự ước lượng tốt nhất hành vi dựa trên môi trường nhất định.

Học liên tục: Các hệ thống AI khác nhau về thời điểm và cách thức thu thập dữ liệu. Trong một số trường hợp cơ sở tri thức là tĩnh và được cung cấp ngay từ đầu cùng với hệ thống có các thành phần được lập trình sẵn. Trong các trường hợp khác cơ sở tri thức thay đổi hoặc thích ứng theo thời gian với thông tin được cập nhật trong quá trình hoạt động của nó. Hệ thống học máy có thể được đặc trưng bởi thời điểm, trong suốt thời gian tồn tại của chúng mà việc học tập xảy ra. Trong nhiều trường hợp, giai đoạn huấn luyện ban đầu mang lại một vài ước lượng gần đúng cho hàm mục tiêu thực tế và hệ thống vẫn tiếp tục hoạt động như cũ mà không cập nhật đại diện nội bộ dựa trên các mẫu mới. Một cách tiếp

cận thay thế khác được gọi là học suốt đời hoặc học liên tục, trong đó quá trình học được tiến hành dàn trải theo thời gian; việc cập nhật cho mô hình được lặp lại mỗi khi có dữ liệu mới. Trên thực tế các mô hình sử dụng học suốt đời thường thực hiện kết hợp cả hai cách tiếp cận; sau giai đoạn tập huấn ban đầu trong đó phần lớn việc học diễn ra, mô hình sẽ được tinh chỉnh với nhiều dữ liệu hơn theo thời gian.

7.4 Từ dự đoán đến hành động

7.4.1 Yêu cầu chung

Kết quả của quá trình xử lý đầu vào bởi hệ thống AI thể hiện ở nhiều dạng khác nhau tùy thuộc mức độ tự động hóa của hệ thống. Tùy vào trường hợp sử dụng hệ thống AI có thể chỉ tạo ra đầu ra thô thuần túy về mặt kỹ thuật (dự đoán), hoặc nó có thể thực hiện các bước hiệu quả hơn bằng việc đưa ra đề xuất hoặc hành động trên môi trường hoạt động của nó (khuyến nghị, quyết định và cuối cùng là hành động).

Trong trường hợp phân loại, kết quả có sai sót thường được phân loại là sai số dương tính giả hoặc sai số âm tính giả. Dương tính giả được mô tả là một dự đoán dương tính khi kết quả thực là âm tính. Âm tính giả là kết quả mà mô hình dự đoán không chính xác một kết quả thực mang tính tiêu cực. Người sử dụng hệ thống AI cần hiểu tác động của một kết quả sai bao gồm cả khả năng dự đoán thiên vị. Những vấn đề như vậy có thể phản ánh trực tiếp đặc điểm của các công cụ, quá trình hoặc dữ liệu được sử dụng để phát triển hệ thống.

Một điểm chính là đầu ra AI dễ bị lỗi. Kết quả đầu ra thể hiện xác suất đúng thay vì hoàn toàn đúng. Cả nhà thiết kế hệ thống và người sử dụng hệ thống AI cần phải hiểu rằng các hệ thống như vậy có thể tạo ra kết quả đầu ra không chính xác và những hàm ý về trách nhiệm giải trình khi sử dụng kết quả đầu ra không chính xác.

7.4.2 Dự đoán

Thuật ngữ "dự đoán" đề cập đến đầu ra trước tiên của một hệ thống AI.

Hệ thống AI đưa ra dự đoán bằng việc áp dụng mô hình với dữ liệu hoặc tình huống mới. Ở đây sử dụng lại kịch bản ở mục 7.4.3 mô tả một hệ thống AI đã được phát triển bằng cách sử dụng các hồ sơ khoản vay trước đó; khi một người mới đăng ký một khoản vay, thông tin của họ được cung cấp cho mô hình và nó sẽ đưa ra ước tính một khoản tiền cho người đó vay.

CHÚ THÍCH: Trong việc sử dụng trí thông minh nhân tạo, dự đoán không nhất thiết bao hàm một tuyên bố về tương lai — nó chỉ đề cập đến kết quả đầu ra của hệ thống AI, có thể là một loại hoa trong một hình ảnh hoặc một bản dịch sang một ngôn ngữ khác.

7.4.3 Quyết định

Quyết định tương ứng với việc lựa chọn một quá trình hành động cụ thể cho mục đích áp dụng nó.

Quyết định có thể được thực hiện bởi chính hệ thống hoặc bởi con người dựa vào các kết quả của hệ thống. Chúng có thể được thực hiện dựa trên các khuyến nghị hoặc trực tiếp dựa vào các dự đoán.

Ví dụ: một người được dự đoán là có rủi ro tín dụng thấp, nhân viên cho vay là con người có thể phân tích kết quả đó cùng với thông tin khác của người này và tình hình tài chính của bên cho vay và sau đó quyết định phê duyệt đơn xin vay của người đó. Ngoài ra, hệ thống có thể tự đưa ra khuyến nghị phê duyệt khoản vay và ước lượng mang tính xác suất về hành động tốt nhất liên quan đến kỳ vọng của bên cho vay, nhân viên cho vay căn cứ vào đó cho rằng xác suất đó có thể chấp nhận được và sẽ quyết định phê duyệt khoản vay. Hoặc là đơn xin vay có thể được phê duyệt tự động bằng cách áp dụng các ngưỡng quyết định của hệ thống đối với các khuyến nghị đó.

Sự phán xét và giám sát của con người có liên quan đến nhiều cách thức khác nhau trong quá trình ra quyết định. Các ngưỡng do con người xác định thường được đặt ra bằng việc xem xét các rủi ro liên quan đến việc tự động hóa các quyết định. Ngay cả khi các quyết định hoàn toàn tự động, con người vẫn có thể sử dụng các dự đoán để theo dõi các quyết định đưa ra.

7.4.4 Hành động

Các hành động tuân theo các quyết định, đây là khi kết quả của hệ thống AI bắt đầu tác động đến thế giới thực (dù là vật lý hay ảo).

Thực hiện một hành động là bước cuối cùng của việc áp dụng thông tin trong hệ thống AI. Ví dụ: trong tình huống xử lý đơn đăng ký vay tín dụng trong mục 7.4.2, khi khoản vay của một người được chấp thuận, các hành động có thể bao gồm chuẩn bị các tài liệu cho vay, lấy chữ ký và phát hành thanh toán. Hãy xem xét hoạt động của một người máy, một hành động có thể là các lệnh chỉ dẫn cho bộ truyền động của người máy để định vị cánh tay và bàn tay của nó. Tùy thuộc vào hệ thống AI, hành động có thể diễn ra trong ranh giới hệ thống AI hoặc bên ngoài ranh giới hệ thống AI.

8 Hệ sinh thái AI

8.1 Yêu cầu chung

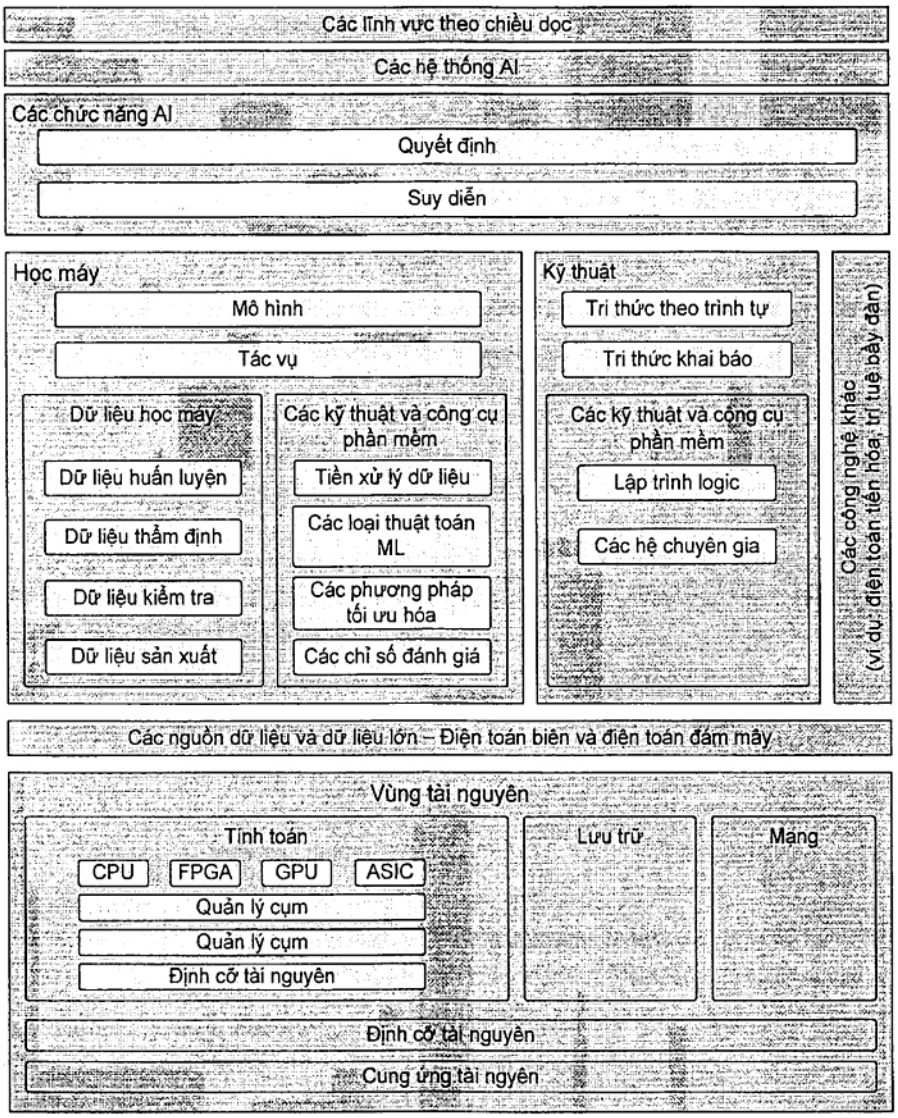
Hình 6 thể hiện hệ sinh thái AI theo các lớp chức năng. Các hệ thống AI lớn không dựa trên một công nghệ duy nhất mà dựa trên sự kết hợp của các công nghệ phát triển theo thời gian. Các hệ thống như vậy có thể sử dụng đồng thời nhiều công nghệ khác nhau, ví dụ: mạng nơ-ron, mô hình biểu trưng và suy diễn xác suất.

Mỗi lớp của Hình 6 sử dụng tài nguyên của các lớp dưới để thực hiện các chức năng của nó. Các hộp màu sáng hơn là các thành phần phụ của một lớp hoặc một chức năng. Kích thước của các lớp và thành phần phụ không thể hiện tính quan trọng của nó.

Tạo lập các hệ thống AI vẫn là một chủ đề được tiến hành nghiên cứu. Trong khi đó việc sử dụng công nghệ AI đang trở thành một phần cố hữu của nhiều ngành công nghiệp, mỗi ngành đều có những nhu cầu, giá trị và ràng buộc pháp lý riêng.

Các ứng dụng AI chuyên biệt, chẳng hạn như ứng dụng sử dụng cho thị giác máy tính hoặc để xử lý ngôn ngữ tự nhiên, tự chúng đã trở thành các khối tạo lập chức năng để triển khai các sản phẩm và dịch

vụ khác nhau. Các ứng dụng này đang thúc đẩy thiết kế các hệ thống AI chuyên biệt và đang được ưu tiên nghiên cứu và phát triển.



Hình 6 - Hệ sinh thái AI

Công nghệ AI thường yêu cầu sử dụng số lượng đáng kể các máy tính, bộ lưu trữ và tài nguyên mạng. Ví dụ trong giai đoạn tập huấn cho hệ thống học máy, các tài nguyên như trong Hình 6 có thể được cung cấp một cách hiệu quả bằng sử dụng công nghệ tính toán đám mây.

Các điều khoản dưới đây mô tả các thành phần chính của hệ sinh thái AI như thể hiện trong Hình 6.

8.2 Hệ thống AI

Hệ thống AI có thể được sử dụng trong nhiều ứng dụng và để giải quyết vô số tác vụ. Điều 9 mô tả ví dụ về các ứng dụng sử dụng AI như nhận dạng hình ảnh, xử lý ngôn ngữ tự nhiên và bảo trì mang tính dự đoán. Điều 5 liệt kê các danh mục tác vụ mà hệ thống AI có thể giải quyết.

Các hệ thống AI tuân theo những chỉ dẫn thực thi các chức năng ở phạm vi toàn cục. Thông tin được thu thập bằng cách mã hóa cứng (bằng kỹ thuật phù hợp) hoặc bằng học máy để xây dựng mô hình cho miền ứng dụng. Thông tin sau đó được mã hóa dưới dạng mô hình và áp dụng ở cấp độ suy diễn mà tại đó các giải pháp tiềm năng được tính toán. Tiếp theo là cấp độ quyết định mà tại đó các hành động tiềm năng được lựa chọn thực thi để có thể đạt được mục tiêu đề ra. Cấp độ suy diễn bao gồm suy diễn không gian, suy diễn thời gian, suy diễn thông thường theo tri giác, ứng dụng chính sách được tính toán hoặc bất kỳ hình thức suy diễn nào có thể được mã hóa. Mức độ quyết định bao gồm sự lựa chọn dựa vào sở thích hoặc tiện ích theo hành động.

8.3 Chức năng AI

Khi mô hình được xây dựng, chức năng AI có vai trò tính toán để đưa ra các dự đoán, đề xuất hoặc quyết định nói chung để có thể đạt được mục tiêu hiện tại của hệ thống AI.

Suy diễn chỉ là áp dụng các dữ liệu có sẵn trong tình huống hiện tại vào trong mô hình và hỏi mô hình các tùy chọn khả thi là gì.

Một số ví dụ về công nghệ thực hiện các hình thức suy diễn bao gồm lập kế hoạch, suy diễn Bayes, chứng minh các định lý một cách tự động, suy luận thời gian và không gian và các bộ suy diễn về bản thể học.

Trong số các tùy chọn khả thi có thể đạt được mục tiêu, hệ thống vẫn cần quyết định lựa chọn nào là tốt nhất.

Các ứng dụng và tiện ích sắp tới có thể là: taxi tự động sẽ tối đa hóa sự tiện nghi và thoải mái cho khách hàng, chương trình chơi bài để tối đa hóa lợi nhuận cho người chơi.

8.4 Học máy

8.4.1 Yêu cầu chung

Học máy là quá trình sử dụng các kỹ thuật điện toán cho phép các hệ thống học hỏi từ dữ liệu hoặc kinh nghiệm. Nó sử dụng một tập hợp các phương pháp thống kê để tìm các kiểu mẫu trong dữ liệu hiện có và sau đó sử dụng các kiểu mẫu để đưa ra dự đoán đối với dữ liệu sản xuất.

Trong lập trình máy tính truyền thống, một lập trình viên xác định các quan hệ logic để giải quyết một vấn đề nhất định bằng cách chỉ định các bước tính toán chính xác bằng ngôn ngữ lập trình. Ngược lại, logic của mô hình học máy một phần phụ thuộc vào dữ liệu được sử dụng để huấn luyện mô hình. Do đó các phép tính hoặc các bước cần thiết để giải quyết vấn đề không được xác định theo phương pháp tiên nghiệm.

Ngoài ra, trái ngược với lập trình máy tính truyền thống, các mô hình học máy có thể cải thiện theo thời gian mà không cần phải viết lại chương trình mà bằng cách tái huấn luyện trên dữ liệu mới, dữ liệu bổ sung và bằng cách sử dụng các kỹ thuật để tối ưu hóa các thông số mô hình và đặc trưng dữ liệu.

8.5 Kỹ thuật

8.5.1 Yêu cầu chung

Trong phương pháp tiếp cận kỹ thuật bằng chuyên gia là con người, việc xử lý chỉ dựa vào chuyên môn của nhà phát triển và sự hiểu biết của họ về tác vụ. Tri thức không được học từ dữ liệu mà thông qua mã hóa cứng của nhà phát triển dựa trên kinh nghiệm của họ trong một lĩnh vực cụ thể.

Có hai loại kiến thức chính: khai báo và thủ tục. Xem 7.3 để biết thêm chi tiết về cả hai loại tri thức này.

8.5.2 Hệ chuyên gia

Như tên gọi của nó, hệ chuyên gia là một hệ thống AI đóng gói tri thức cung cấp bởi một chuyên gia là con người trong một lĩnh vực cụ thể để suy diễn giải pháp cho các vấn đề.

Hệ chuyên gia bao gồm một cơ sở tri thức, một công cụ suy luận và một giao diện người dùng. Cơ sở tri thức lưu trữ tri thức khai báo của một lĩnh vực cụ thể, bao gồm cả thông tin thực tế và thông tin heuristic. Công cụ suy diễn nắm giữ tri thức thủ tục: tập hợp các quy tắc và phương pháp luận để lập luận. Nó kết hợp các dữ kiện do người dùng cung cấp với thông tin từ cơ sở tri thức.

Suy luận được thực hiện bằng cách sử dụng các quy tắc xác định trước bởi chuyên gia và các đánh giá câu lệnh logic. Các loại vấn đề có thể được giải quyết bằng cách sử dụng hệ thống chuyên gia bao gồm phân loại, chẩn đoán, theo dõi và dự đoán.

8.5.3 Lập trình logic

Lập trình logic là hình thức lập trình dựa trên các ngôn ngữ lập trình thể hiện bằng logic hình thức. Prolog là một ví dụ về ngôn ngữ lập trình logic.

Logic hình thức cho AI là một trọng tâm nghiên cứu hiện nay. Nhiều loại logic học hình thức hướng đến mục tiêu mô hình hóa suy diễn của con người trong các tình huống khác nhau. Lập trình logic cung cấp khuôn khổ thực hiện các mô hình suy diễn của con người. Các tác nhân AI cần phải có khả năng tái tạo các loại suy diễn khác nhau theo cách được chỉ định rõ ràng, minh bạch và có thể giải thích được.

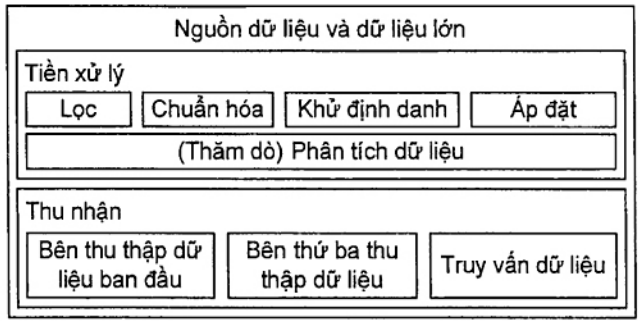
Lập trình logic với các câu lệnh khai báo phối ghép với nhau để xử lý ngôn ngữ tự nhiên một cách mạnh mẽ cho phép các thực thể suy diễn bằng phép loại suy, rút ra kết luận và khái quát hóa được các đối tượng và môi trường.

VÍ DỤ: Apache Jena [19] là một khuôn khổ web ngữ nghĩa cung cấp một công cụ suy luận.

8.6 Dữ liệu lớn và nguồn dữ liệu - tính toán đám mây và tính toán biên

8.6.1 Dữ liệu lớn và nguồn dữ liệu

Mọi hệ thống ML đều sử dụng dữ liệu. Dữ liệu đó có thể ở nhiều dạng khác nhau. Trong một số trường hợp dữ liệu được sử dụng bởi hệ thống ML là dữ liệu lớn. Hình 6 thể hiện khối dữ liệu lớn này đại diện cho nguồn, định dạng và quá trình xử lý điển hình của dữ liệu lớn cho bất kể mục đích sử dụng nào của nó. Mục này mô tả thêm các thành phần chính của nó trong Hình 7.



Hình 7 - Dữ liệu lớn và nguồn dữ liệu

Dữ liệu lớn là những tập dữ liệu mở rộng có các đặc tính được hiểu trong khái niệm về khối lượng, tốc độ và sự đa dạng của dữ liệu được tạo ra. Nó đòi hỏi các công nghệ và kỹ thuật chuyên biệt để xử lý và nhận ra giá trị trong đó. Ví dụ các công nghệ đã được phát triển đặc biệt cho phép xử lý phân tán các bộ dữ liệu lớn bằng máy tính cụm thông qua các mô hình lập trình đơn giản. Ngoài ra, các công nghệ lưu trữ và quản trị cơ sở dữ liệu cũng đã được phát triển riêng để quản lý khối lượng dữ liệu lớn được tập hợp từ các khối lượng dữ liệu lớn khác.

Dữ liệu lớn đã trở nên quan trọng khi các tổ chức tăng cường độ rộng và chiều sâu của việc thu thập dữ liệu và do đó yêu cầu các công nghệ và kỹ thuật chuyên biệt để nhận biết chi tiết hơn về dữ liệu của mình.

Xem trong TCVN 13238:2020 ISO/IEC 20546 và ISO/IEC 20547-3 để biết thêm thông tin về dữ liệu lớn.

Dữ liệu lớn sử dụng nhiều trong các hệ thống AI và nó là yếu tố thúc đẩy việc phát triển các hệ thống như vậy. Sự sẵn có của các bộ sưu tập lớn dữ liệu phi cấu trúc trong các lĩnh vực ứng dụng khác nhau cung cấp những hiểu biết mới về dữ liệu từ việc sử dụng các kỹ thuật AI, chẳng hạn như khám phá tri thức và nhận dạng kiểu mẫu. Sự sẵn có của một lượng lớn dữ liệu huấn luyện dẫn đến việc các mô hình học máy được cải tiến để có khả năng sử dụng được trong nhiều loại hình ứng dụng.

Dữ liệu có thể được thu thập bởi tổ chức sử dụng nó (thu thập của bên thứ nhất). Ví dụ nhà bán lẻ sử dụng dữ liệu giao dịch mà họ có được từ các hệ thống điểm bán hàng mà họ sở hữu. Dữ liệu cũng có thể được mua từ bên thứ ba, chẳng hạn như các tổ chức nghiên cứu và các nhà cung cấp dữ liệu khác, họ thu thập dữ liệu và sau đó bán hoặc chia sẻ dữ liệu với các tổ chức khác sử dụng. Ngoài ra, dữ liệu có thể được thu thập bằng cách truy vấn và kết hợp dữ liệu từ các bộ dữ liệu khác nhau, chẳng hạn như kết hợp dữ liệu của cả bên thứ nhất và bên thứ ba.

Dữ liệu có thể đến từ nhiều nguồn như:

- Điểm bán hàng và các giao dịch khác;
- Các cuộc thăm dò hoặc khảo sát;
- Nghiên cứu thống kê;
- Các quan sát được ghi lại;
- Các bộ cảm biến;
- Hình ảnh;
- Bản ghi âm;
- Các tài liệu;
- Các tương tác với các hệ thống.

8.6.2 Tính toán đám mây và tính toán biên

Tính toán đám mây là một mô hình cho phép truy cập mạng vào một vùng chứa tài nguyên vật lý hoặc ảo có thể mở rộng và có tính co dãn với khả năng tự cung ứng và quản trị các dịch vụ theo yêu cầu, xem trong TCVN 12480:2019 ISO/IEC 17788 và TCVN 12481:2019 ISO/IEC 17789.

Tính toán đám mây thường kết hợp với các trung tâm dữ liệu tập trung, lớn để có khả năng cung cấp dung lượng xử lý và lưu trữ dữ liệu rất lớn. Năng lực lớn như vậy có thể là điều cần thiết đối với một số thành phần trong vòng đời AI, đặc biệt là khi phải xử lý các tập dữ liệu lớn để huấn luyện các hệ thống AI và xây dựng các mô hình mà chúng sử dụng.

Tính toán biên là tính toán phân tán, trong đó quá trình xử lý và lưu trữ dữ liệu diễn ra tại biên hoặc gần với biên của mạng, nơi mà mức độ gần được xác định theo yêu cầu của hệ thống. Biên là ranh giới giữa các thực thể vật lý và và số liên quan, được phân định bằng các cảm biến và thiết bị truyền động kết nối mạng (xem ISO/IEC 23188).

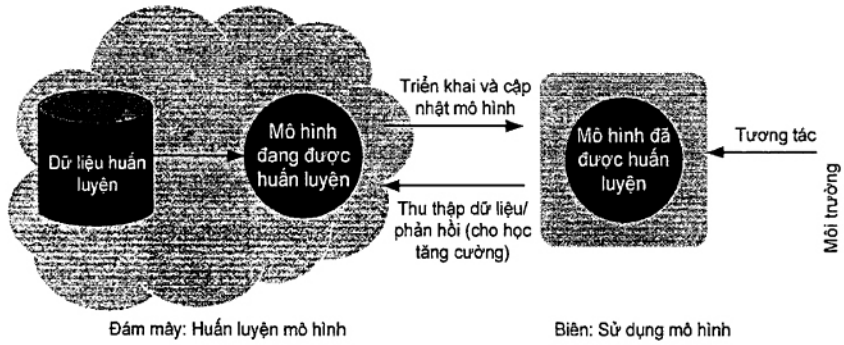
Tính toán biên phần lớn đề cập đến bố trí và vận hành các phần mềm thành phần và lưu trữ dữ liệu. Các phần mềm thành phần, chẳng hạn như phần mềm liên kết với hệ thống AI xử lý các thiết bị IoT (cảm biến và thiết bị truyền động) thường yêu cầu giảm thiểu độ trễ và tạo ra kết quả với giới hạn đáng kể về thời gian (thường được gọi là thời gian thực), hoặc cần khả năng phục hồi để hệ thống vẫn có thể hoạt động nếu đường truyền thông bị gián đoạn, hoặc cần bảo vệ quyền riêng tư của dữ liệu cá nhân được thu thập từ các thiết bị biên. Để đạt được điều này có thể yêu cầu quá trình xử lý và lưu trữ dữ liệu được thực hiện tại biên hoặc gần biên. Xem ISO/IEC 23188 để biết thêm chi tiết.

Tuy nhiên, điều quan trọng là phải hiểu rằng tính toán đám mây có thể được triển khai ở nhiều nơi trong môi trường điện toán phân tán, bao gồm cả những nơi không tập trung và gần biên. Ở dạng này tính toán đám mây có thể triển khai một cách mềm dẻo và linh hoạt với cả phần mềm và dữ liệu, sử dụng ảo hóa trong xử lý và lưu trữ dữ liệu, kết hợp với tạo vùng tài nguyên có khả năng co giãn và mở rộng nhanh chóng để bố trí và vận hành các thành phần của hệ thống AI một cách phù hợp.

Thông thường, các hệ thống tính toán biên được kết hợp với các hệ thống tập trung để tạo ra các giải pháp hoàn chỉnh nhằm tận dụng năng lực của từng loại hệ thống.

Có ba nguyên tắc chính cho thiết kế hệ thống ML kết hợp với đám mây và biên là: đào tạo mô hình trên đám mây, đào tạo mô hình tại biên và đào tạo mô hình trong đám mây và tại biên.

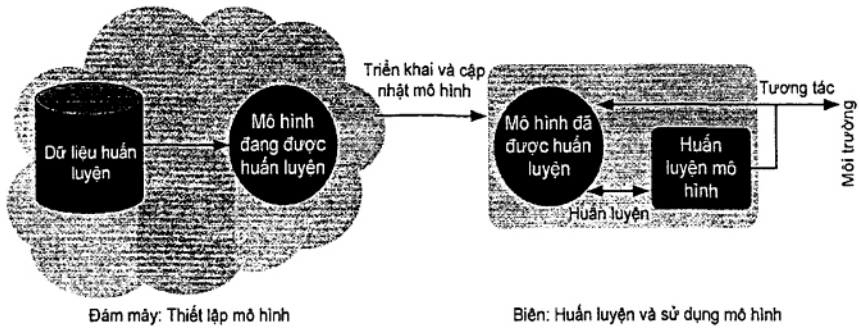
a) Các dịch vụ đám mây có thể được sử dụng như một nền tảng tập trung để đào tạo các mô hình ML (Hình 8). Do các hạn chế về tài nguyên của các thiết bị biên nên các tác vụ chuyên sâu về tính toán và lưu trữ liên quan đến huấn luyện, thẩm định và bảo trì các mô hình được thực hiện bằng cơ sở hạ tầng đám mây. Sau mỗi lần được huấn luyện, mô hình sẽ được triển khai áp dụng, nếu cần sẽ được cập nhật trên các thiết bị biên. Dữ liệu từ các thiết bị biên có thể được sử dụng thêm để thực hiện các hoạt động huấn luyện, hoặc như trong trường hợp học tăng cường nó cung cấp phản hồi về chất lượng mô hình.



Hình 8 - Ví dụ về huấn luyện mô hình trong đám mây

Ví dụ về các ứng dụng sử dụng thiết kế này là phát hiện tấn công trên các bộ định tuyến biên (tường lửa thông minh), phát hiện và ngăn chặn lỗi trong các ứng dụng điều khiển công nghiệp (bảo trì chủ động), nhận dạng biển báo giao thông cho ô tô tự lái.

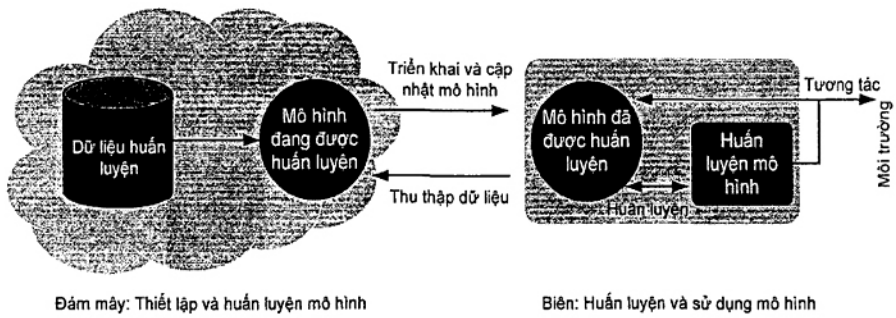
b) Trong trường hợp nguyên tắc thiết kế tập trung không tối ưu cho việc huấn luyện các mô hình đã được cá nhân hóa hoặc các mô hình được sử dụng trong các bối cảnh ứng dụng cụ thể thì có thể áp dụng một sơ đồ khác (Hình 9). Nó dựa trên việc huấn luyện mô hình được thực hiện trực tiếp tại các thiết bị biên (được cung cấp và có đủ tài nguyên trên các thiết bị đó).



Hình 9 - Ví dụ về huấn luyện mô hình tại biên

Trong thiết kế này, chỉ một mô hình khởi đầu (tổng quát) được thiết lập và huấn luyện trong môi trường đám mây. Huấn luyện theo ngữ cảnh hoặc cá nhân hóa được thực hiện tại biên và sử dụng dữ liệu trong thế giới thực. Loại hình huấn luyện mô hình này rất phù hợp đối với các hệ thống hoàn toàn tự động sử dụng các phương pháp tiếp cận học máy tăng cường hoặc không giám sát.

c) Phương pháp kết hợp thực hiện huấn luyện mô hình cả trong đám mây và tại biên (Hình 10). Điều này có thể cần thiết khi thiết kế hệ thống có các thiết bị biên. Trong một số trường hợp, các dịch vụ đám mây chuẩn bị mô hình được huấn luyện ban đầu, sau đó được triển khai tại biên. Trong các trường hợp khác các hệ thống biên huấn luyện các mô hình nội bộ của chúng dựa trên dữ liệu tại chỗ và không truyền dữ liệu cho nhau và lên các dịch vụ đám mây. Các dịch vụ đám mây cũng có thể hoạt động như một máy chủ cung cấp các tham số để đồng bộ hóa các bản cập nhật mô hình của các hệ thống biên khác nhau, sau đó trả lại các bản cập nhật mô hình đã đồng bộ hóa cho các hệ thống biên để cập nhật từng mô hình riêng rẽ của chúng.



Hình 10 - Ví dụ về huấn luyện mô hình trong đám mây và tại biên

Ví dụ này được áp dụng cho dịch vụ dữ liệu dựa trên khu vực cục bộ (ví dụ: dịch vụ thu thập hình ảnh dựa trên máy bay không người lái từ cho các khu vực và lĩnh vực ứng dụng khác nhau, hoặc dịch vụ cho thiết bị gia dụng), nó cho phép cung cấp chất lượng dịch vụ tốt hơn với các mô hình huấn luyện được cập nhật so với các mô hình được đào tạo ban đầu.

Cách tiếp cận khác bao gồm như tải lên mô hình cũng có thể được quan tâm. Với phương pháp này, một mô hình được huấn luyện bởi hệ thống biên sẽ được gửi đến kho lưu trữ mô hình trong môi trường đám mây và phân phối đến các hệ thống biên khác hoạt động trong cùng một môi trường hoặc môi trường tương tự nếu mô hình đó có hiệu năng tốt hơn mô hình trước đó. Cách tiếp cận này có thể được áp dụng để học chuyển giao và các kỹ thuật nén mô hình. Ví dụ về học chuyển giao chẳng hạn như một mô hình được huấn luyện để nhận dạng số nhà ở chế độ xem phố có thể được sử dụng để nhận dạng số viết tay. Mô hình ban đầu hoặc mô hình đã được huấn luyện để giải quyết một vấn đề cụ thể nào đó có thể áp dụng cho những vấn đề tương tự khác. Kỹ thuật nén mô hình có thể sử dụng cho một thiết bị biên có năng lực tính toán thấp. Một mô hình được huấn luyện đầy đủ trong dịch vụ đám mây với tài nguyên máy tính dồi dào có thể được nén trước khi được sử dụng ở một hệ thống biên có tài nguyên ít hơn.

8.7 Vùng tài nguyên

8.7.1 Yêu cầu chung

Hình 6 cho thấy các nguồn lực cần thiết để hỗ trợ hệ sinh thái AI. Các tài nguyên như máy tính, mạng và lưu trữ là rất cần thiết để hỗ trợ các hệ thống AI.

Phát triển và triển khai các hệ thống AI có thể diễn ra với các nguồn lực ở nhiều quy mô khác nhau, từ dịch vụ đám mây tập trung và trung tâm dữ liệu phi đám mây đến các máy chủ (hoặc cụm máy chủ), hệ thống tính toán biên, thiết bị di động và thiết bị IoT. Một số hệ thống nói trên có thể bị hạn chế về tài nguyên như khả năng xử lý và lưu trữ dữ liệu, băng thông mạng và độ trễ. Điều này đặc biệt đúng đối với các hệ thống và thiết bị biên. Tài nguyên xử lý cho các hệ thống AI có thể là bất kỳ sự bố trí và kết hợp nào giữa một hoặc nhiều GPU, NPU, CPU với các loại hình bộ xử lý khác có trong một hoặc nhiều hệ thống để có thể tạo thành các cụm máy tính.

Các yêu cầu của hệ thống AI đối với tài nguyên máy tính có thể thay đổi dựa trên việc sử dụng học máy hoặc học sâu cũng như các loại hình khối lượng công việc (ví dụ: tập huấn và suy luận với các cấu trúc liên kết khác nhau). Vì vậy các giải pháp điện toán không đồng nhất có thể cần thiết để phù hợp với khối lượng công việc và hệ thống AI cụ thể. Ví dụ như các bộ tăng tốc phần cứng (GPU, NPU, FPGA, DSP, ASIC v.v..) được sử dụng cho máy tính xử lý khối lượng công việc liên quan đến AI, chẳng hạn như huấn luyện các cấu trúc liên kết mạng nơ-ron nào đó.

Để đáp ứng nhu cầu của các hệ thống AI khác nhau, việc cung ứng tài nguyên cần phải quản lý tự động, bao gồm việc cung cấp theo yêu cầu và điều phối các tài nguyên không đồng nhất (ví dụ: cung cấp tài nguyên tại chỗ, đám mây và biên).

8.7.2 Mạch tích hợp dành riêng cho ứng dụng

ASIC là một loại mạch tích hợp được tùy chỉnh cho mục đích sử dụng cụ thể. ASIC là một tùy chọn để cung cấp chức năng dành riêng cho AI.

ASIC có thể được tùy chỉnh như một bộ tăng tốc để tăng tốc quá trình AI bằng cách cung cấp các chức năng như khối tích lũy nhân song song chuyên dụng, phân bổ bộ nhớ được tối ưu hóa và bộ tính toán số học có độ chính xác thấp. ASIC cũng có thể được định cấu hình như một bộ đồng xử lý để cung cấp các chức năng xử lý dữ liệu trước và sau cho các tác vụ AI như cắt và thay đổi kích thước hình ảnh, chuyển đổi, giảm nhiễu và tổng hợp dữ liệu cảm biến.

So với các bộ xử lý phổ thông (ví dụ: CPU và GPU) thì ASIC thường được thiết kế, sản xuất và sử dụng chỉ cho các tình huống cụ thể như triển khai cấu trúc mạng nơ-ron nào đó. ASIC cung cấp năng lực tính toán cao hơn cho AI với khối lượng không gian thấp hơn, chi phí và tiêu thụ năng lượng cũng thấp hơn.

ASIC cho phép AI được triển khai trong các thiết bị hạn chế về không gian và năng lượng như điện thoại di động. ASIC cũng cho phép AI sử dụng trong các thiết bị IoT trong các lĩnh vực rất đa dạng như sản xuất công nghiệp, y học, vệ sinh, an ninh và nhà thông minh.

9 Các lĩnh vực của AI

9.1 Thị giác máy tính và nhận dạng hình ảnh

Tiêu chuẩn này định nghĩa thị giác máy tính là "khả năng của một khối chức năng thu nhận, xử lý và biên dịch dữ liệu đại diện cho hình ảnh hoặc video" (3.7.1). Thị giác máy tính liên quan chặt chẽ đến nhận dạng hình ảnh, ví dụ như xử lý ảnh số. Dữ liệu trực quan thường bắt nguồn từ cảm biến ảnh kỹ thuật số, hình ảnh tương tự được số hóa bằng kỹ thuật quét hoặc bằng các thiết bị đồ họa khác. Trong phạm vi của tiêu chuẩn này thì ảnh số bao gồm cả ảnh tĩnh và ảnh động.

Ảnh số tồn tại dưới dạng ma trận các con số đại diện cho các thang màu xám hoặc màu sắc trong hình ảnh được chụp hoặc trong các trường hợp khác là một tập các véc-tơ. Ảnh số có thể bao gồm siêu dữ liệu mô tả các đặc điểm và thuộc tính liên quan đến hình ảnh. Ảnh số có thể được nén để tiết kiệm không gian lưu trữ và cải thiện hiệu suất truyền tải của chúng trên mạng.

Dưới đây là các ví dụ về ứng dụng AI dựa trên thị giác máy tính và nhận dạng hình ảnh:

- Xác định các ảnh cụ thể từ một tập các ảnh (ví dụ: ảnh của những con chó trong tập các hình ảnh về động vật);
- Xe tự lái: phát hiện và nhận dạng tín hiệu giao thông, đối tượng trên xe tự động;
- Chẩn đoán y tế: phát hiện bệnh và tình trạng bất thường trong hình ảnh y tế;
- Kiểm soát chất lượng (ví dụ: phát hiện các bộ phận bị lỗi trên dây chuyền lắp ráp);
- Nhận dạng khuôn mặt.

Các tác vụ cơ bản đối với thị giác máy tính bao gồm thu nhận hình ảnh, lấy mẫu lại, chia tỷ lệ, giảm nhiễu, tăng cường độ tương phản, trích xuất thuộc tính, phân đoạn, phát hiện và phân loại đối tượng.

Hiện đã có một số phương pháp thực hiện các tác vụ thị giác máy tính trong các hệ thống AI. Sử dụng mạng nơ-ron tích hợp sâu (5.12.1.4) đã được áp dụng trong những năm gần đây do độ chính xác cao

đối với tác vụ phân loại hình ảnh, hiệu suất huấn luyện và dự đoán.

9.2 Xử lý ngôn ngữ tự nhiên

9.2.1 Yêu cầu chung

Xử lý ngôn ngữ tự nhiên là xử lý thông tin dựa trên sự hiểu biết ngôn ngữ tự nhiên và tạo ra ngôn ngữ tự nhiên. Điều này bao gồm việc phân tích và tạo ngôn ngữ tự nhiên trên cơ sở văn bản hoặc lời nói. Bằng cách sử dụng các tính năng của NLP, máy tính có thể phân tích văn bản viết bằng ngôn ngữ của con người và xác định các khái niệm, thực thể, từ khóa, quan hệ, cảm xúc, tình cảm và các đặc điểm khác để sau đó cho phép người dùng hiểu rõ về nội dung. Với những khả năng đó, máy tính cũng có thể tạo ra văn bản hoặc giọng nói để giao tiếp với người dùng. Hệ thống xử lý ngôn ngữ tự nhiên có khả năng sử dụng ngôn ngữ tự nhiên cho đầu vào hoặc xuất nó ở đầu ra ở dạng văn bản hoặc lời nói và có khả năng xử lý chúng bằng các phần tử xử lý ngôn ngữ tự nhiên. Ví dụ về hệ thống nêu trên có thể là hệ thống đặt vé tự động cho một công ty hàng không, nó có thể nhận cuộc gọi từ khách hàng và đặt chuyến bay cho họ. Một hệ thống như vậy cần có các phần tử hiểu và tạo được ngôn ngữ tự nhiên.

Dưới đây là các ví dụ về ứng dụng AI dựa trên xử lý ngôn ngữ tự nhiên:

- Nhận dạng chữ viết tay (ví dụ: chuyển đổi các ghi chú viết tay thành văn bản số hóa);
- Nhận dạng lời nói (ví dụ: hiểu nội dung, ý nghĩa của lời nói của con người);
- Phát hiện thư rác (ví dụ: sử dụng ý nghĩa của các từ trong thư email để xác định xem thư đó có được phân loại là không mong muốn hay không);
- Trợ lý số cho cá nhân và trò chuyện trực tuyến, sử dụng khả năng hiểu và tạo ngôn ngữ tự nhiên (bao gồm nhận dạng và tạo lời nói) để cung cấp giao diện trò chuyện đối với người dùng;
- Tóm tắt, tổng hợp;
- Tạo văn bản;
- Tìm kiếm nội dung.

NLP cũng được sử dụng trong nhiều hệ thống ứng dụng như chatbot, hệ thống quảng cáo dựa trên nội dung, hệ thống dịch giọng nói và hệ thống học tập điện tử.

9.2.2 Các thành phần xử lý ngôn ngữ tự nhiên

9.2.2.1 Yêu cầu chung

Các thành phần NLP giải quyết các tác vụ khác nhau. Các tác vụ phổ biến nhất bao gồm:

NLU: Thành phần NLU chuyển đổi văn bản hoặc lời nói thành một mô tả nội bộ có ngữ nghĩa của nội dung đầu vào. Khó khăn đến từ sự mơ hồ mang tính cố hữu của các ngôn ngữ tự nhiên: các từ và câu về bản chất là đa nghĩa, do đó kết quả của NLU dễ xảy ra sai sót.

NLG: Thành phần NLG chuyển đổi mô tả nội bộ thành văn bản hoặc lời nói mà con người có thể hiểu được. Tác vụ này có thể liên quan đến việc điều chỉnh các cụm từ để người dùng cảm thấy sự diễn đạt

có vẻ tự nhiên hơn.

POS: Thành phần gắn nhãn POS sử dụng để phân loại từ ở đầu vào thành các đối tượng ngữ pháp như là danh từ, tính từ, động từ v.v.. Việc gắn thẻ POS cũng bị ảnh hưởng bởi tính đa nghĩa của các từ.

NER: Thành phần NER thực hiện nhận dạng và gắn nhãn các tên biểu thị của người, vị trí, tổ chức hoặc thực thể khác cho các chuỗi từ trong một luồng văn bản hoặc lời nói. Tùy thuộc vào thực thể mà nhiều thông tin khác có thể được trích xuất. Ví dụ đối với một cá nhân có thể bổ sung nhãn thông tin về chức danh, chức vụ, chuyên môn của họ là rất hữu ích.

Trả lời câu hỏi: *Thành phần trả lời câu hỏi sẽ cố gắng đưa ra câu trả lời thích hợp nhất cho câu hỏi của* con người. Người dùng hỏi điều gì đó bằng ngôn ngữ tự nhiên và hệ thống cung cấp câu trả lời cũng bằng ngôn ngữ tự nhiên.

MT: Thành phần MT tự động dịch nội dung ngôn ngữ tự nhiên từ ngôn ngữ này sang ngôn ngữ khác. Kịch bản tự động dịch có thể từ văn bản sang văn bản, lời nói sang văn bản, lời nói sang lời nói hoặc *văn bản sang lời nói. Hoạt động dịch tự động sẽ gặp khó khăn do tính đa nghĩa của từ ngữ, một từ có* thể có nhiều nghĩa cũng như sự biến thể của ngôn ngữ, chẳng hạn việc tham chiếu giữa các câu hoặc bên trong các câu, các ý không thành văn. Trong nhiều trường hợp có thể thực hiện dịch đa ngôn ngữ.

OCR: Thành phần OCR chuyển đổi các tài liệu viết dưới dạng hình ảnh (có thể được quét) thành một dạng mô tả được mã hóa bằng kỹ thuật số về nội dung của chúng, chẳng hạn như văn bản, bảng biểu, số liệu, tiêu đề và các mối quan hệ của chúng.

Trích xuất quan hệ: Thành phần trích xuất quan hệ thực hiện tác vụ trích xuất quan hệ giữa các thực thể được đặt tên hoặc thậm chí giữa bất kỳ thực thể nào ở đầu vào. Ví dụ thành phần này có thể xác định trong văn bản đầu vào là diễn viên "Al Pacino" "đóng vai chính" trong bộ phim "Serpico".

IR: IR hoặc thành phần tìm kiếm giải quyết nhu cầu thông tin của người dùng bằng cách tìm kiếm thông tin trong một tập hợp nội dung không có cấu trúc. Truy vấn người dùng thể hiện nhu cầu thông tin của họ được so sánh theo thuật toán với mỗi phần tử trong tập hợp thông tin để đưa ra dự đoán mức độ liên quan của phần tử đó với nhu cầu thông tin của người dùng. Đầu ra của thành phần này thường được trình bày cho người dùng dưới dạng danh sách các phần tử đã chọn được xếp hạng theo thứ tự giảm dần về mức độ liên quan của chúng. Các thành phần truy xuất thông tin có thể phát triển cho nhiều loại phần tử thông tin khác nhau bao gồm văn bản tự do, tài liệu bán cấu trúc, tài liệu có cấu trúc, âm thanh, hình ảnh và video và bằng các ngôn ngữ tự nhiên khác nhau.

Phân tích cảm xúc: Thành phần phân tích cảm xúc dùng để xác định và phân loại một cách tính toán các ý kiến được thể hiện trong một đoạn văn bản, bài phát biểu hoặc hình ảnh. Nó còn được gọi là khai thác các ý kiến. Ví dụ về các khía cạnh chủ quan có thể bao gồm cảm xúc tích cực hoặc tiêu cực.

Tóm tắt tự động: *Thành phần tóm tắt tự động thực hiện truyền đạt một cách súc tích hơn thông tin quan* trọng từ một phần tử nội dung bằng một trong hai cách tiếp cận sau đây (hoặc kết hợp chúng). Tóm tắt

trích xuất là chọn nội dung chính có liên quan từ nội dung nguồn để tạo ra một phiên bản tóm tắt rút gọn. Tóm tắt trừu tượng là tổng hợp một văn bản mới ngắn hơn để truyền đạt thông tin có liên quan. Tóm tắt trừu tượng có liên quan đến việc tạo ra ngôn ngữ tự nhiên.

Quản lý đối thoại: Thành phần quản lý đối thoại sẽ giúp quản lý một loạt các tương tác giữa người dùng và hệ thống với mục đích cải thiện trải nghiệm người dùng theo cách thức giống như một cuộc trò chuyện bằng ngôn ngữ tự nhiên. Quản lý đối thoại sử dụng một loạt các phương pháp tiếp cận khác nhau, bao gồm các quy tắc khai báo chỉ định phản hồi cho các trình kích hoạt đầu vào cụ thể, các phương pháp tiếp cận dựa vào học máy. Quản lý đối thoại có thể thúc đẩy các tương tác dựa trên văn bản, chẳng hạn như cung cấp trải nghiệm trò chuyện nhiều hơn với các thành phần trả lời câu hỏi, hoặc tích hợp với các thành phần tổng hợp và nhận dạng giọng nói để hỗ trợ các ứng dụng trong trợ lý cá nhân, đại lý dịch vụ khách hàng trực tuyến hoặc người máy chăm sóc cá nhân.

9.2.2.2 Dịch máy

Dịch máy là một tác vụ của NLP, trong đó một hệ thống máy tính được sử dụng để tự động dịch văn bản hoặc lời nói từ ngôn ngữ tự nhiên này sang ngôn ngữ tự nhiên khác.

Nói chung quá trình dịch thuật của con người diễn ra theo hai bước. Bước đầu tiên là giải mã ý nghĩa của ngôn ngữ nguồn. Bước thứ hai là mã hóa ý nghĩa đó sang ngôn ngữ đích. Quá trình này đòi hỏi kiến thức chuyên sâu về ngữ pháp, ngữ nghĩa, cú pháp, thành ngữ, nền tảng văn hóa và các lĩnh vực khác.

Thách thức kỹ thuật đối với dịch máy bao gồm từ đa nghĩa, ngữ cảnh quan tâm, sự khác biệt về ngữ pháp và ngôn ngữ sử dụng hệ thống viết theo kiểu tốc ký. Đã có nhiều cách tiếp cận để dịch máy, chẳng hạn như dựa trên quy tắc, dựa trên ví dụ, thống kê, nơ-ron hoặc sự kết hợp của chúng.

Trong những năm gần đây mạng nơ-ron được sử dụng để thực hiện dịch máy, điều này đã dẫn đến những cải tiến đáng kể về độ trôi chảy và độ chính xác của bản dịch. Thông qua học sâu mô hình có thể được huấn luyện và tùy chỉnh để diễn đạt trong các lĩnh vực cụ thể để đạt được mức độ chính xác cao.

9.2.2.3 Tổng hợp lời nói

Một hệ thống chuyển đổi văn bản ngôn ngữ tự nhiên thành lời nói được gọi là hệ thống chuyển văn bản thành lời nói.

Nói chung quá trình TTS có ba giai đoạn: 1) phân tích, 2) mô hình hóa và 3) tổng hợp. Tính tự nhiên và tính dễ hiểu là những đặc điểm quan trọng của hệ thống TTS. Tính tự nhiên mô tả âm thanh đầu ra gần với giọng nói của con người gần như thế nào. Trong khi tính dễ hiểu là mức độ rõ ràng và dễ hiểu của đầu ra đối với con người. Hệ thống tổng hợp lời nói thường cố gắng tối đa hóa cả hai đặc điểm nêu trên.

Các phương pháp tiếp cận sử dụng trong tổng hợp giọng nói bao gồm tổng hợp nối, tổng hợp formant, tổng hợp khớp, tổng hợp dựa trên HMM, tổng hợp sóng hình sin và DNN. Mỗi cách tiếp cận đều có điểm mạnh và điểm yếu riêng. Một số bộ tổng hợp giọng nói dựa trên DNN đang đạt đến chất lượng của giọng nói con người.

9.2.2.4 Nhận dạng lời nói

Tiêu chuẩn này định nghĩa nhận dạng lời nói là chuyển đổi tín hiệu lời nói bởi một khối chức năng thành dạng trình bày nội dung của lời nói. Lời nói được số hóa là một dạng dữ liệu tuần tự, do đó các kỹ thuật có thể kiểm soát dữ liệu trong một khoảng thời gian đủ để xử lý âm vị từ lời nói.

Một số cách tiếp cận sử dụng mạng nơ-ron đã được sử dụng để nhận dạng lời nói. Một cách tiếp cận khác liên quan đến việc sử dụng LSTM [18]. Phương pháp này cho phép mạng nơ-ron được huấn luyện và triển khai như một giải pháp nhận dạng lời nói mà không cần kết hợp với các quá trình khác như HMM và cho hiệu suất nhận dạng hợp lý.

Dưới đây là các ví dụ về ứng dụng AI dựa trên nhận dạng lời nói:

- Hệ thống ra lệnh bằng lời nói;
- Đọc chính tả;
- Trợ lý cá nhân.

9.2.2.5 Trả lời câu hỏi

Hệ thống trả lời câu hỏi có thể nhập một số lượng lớn các trang văn bản và áp dụng công nghệ trả lời câu hỏi để trả lời các câu hỏi do con người đặt ra bằng ngôn ngữ tự nhiên. Cách tiếp cận này cho phép mọi người "hỏi" và nhận được câu trả lời gần như ngay lập tức kể cả đối với những câu hỏi phức tạp. Công nghệ trả lời câu hỏi kết hợp với các API khác và với phân tích nâng cao cho phép nó tự phân biệt với công nghệ tìm kiếm thông thường (được kích hoạt bởi từ khóa) bằng việc cung cấp nhiều hơn trải nghiệm tương tác đối với người dùng.

9.3 Khai phá dữ liệu

Khai thác dữ liệu đề cập đến việc áp dụng các thuật toán để khám phá thông tin hợp lệ, mới lạ và hữu ích từ dữ liệu. Khai phá dữ liệu trở nên nổi tiếng vào cuối những năm 1990 và được công nhận là khác biệt so với các phương pháp thống kê trước đó. Thống kê truyền thống tập trung vào việc thu thập dữ liệu cần thiết và đủ để trả lời dứt khoát một câu hỏi cụ thể. Khai thác dữ liệu thường áp dụng cho dữ liệu được tái sử dụng cho nhiều mục đích để tìm câu trả lời gần đúng hoặc xác suất phù hợp với các kiểu mẫu cần thiết. Khai phá dữ liệu được coi là bước mô hình hóa thuật toán trong quy trình KDD hoàn chỉnh. Nổi lên nhờ những nỗ lực khai thác dữ liệu ban đầu, một tập đoàn đã mô tả chi tiết các bước khai phá dữ liệu và đưa vào trong tiêu chuẩn công nghiệp CRISP-DM xuất bản năm 2000 [28]. Khai phá dữ liệu bao gồm các kỹ thuật như cây ra quyết định, phân cụm và phân loại. Khi công nghệ dữ liệu lớn xuất hiện vào giữa những năm 2000 thì việc áp dụng các thuật toán là không thể tách rời với lưu trữ dữ liệu, việc lấy mẫu mang tính nghiêm ngặt đã nhường chỗ cho việc xử lý dữ liệu chuyên sâu nhanh hơn. Những thay đổi này dẫn đến cách thức mới về mô tả dữ liệu cho phiên bản dữ liệu lớn của quá trình vòng đời KDD như một phần của khoa học dữ liệu. Mặc dù KDD và khai phá tri thức là những thuật ngữ phổ biến trong AI, nhưng thứ mà máy tính tạo ra không phải là tri thức mà là thông tin.

9.4 Lập kế hoạch

Lập kế hoạch là một lĩnh vực phụ của AI. Nó rất quan trọng đối với các ứng dụng công nghiệp và được coi là quan trọng trong nhiều lĩnh vực kinh doanh, chẳng hạn như quản lý rủi ro, chăm sóc sức khỏe, người máy công tác trong công nghiệp, an ninh mạng, trợ lý nhận thức và quốc phòng.

Việc lập kế hoạch cho phép máy tự động đưa ra trình tự các hoạt động để đạt được các mục tiêu nhất định và tối ưu hóa hiệu năng cho các biện pháp thực thi đã được xác định. Xét ở góc độ lập kế hoạch, một hệ thống hoạt động ở một trạng thái nhất định. Việc thực hiện một hành động có thể thay đổi trạng thái hệ thống và trình tự thực thi các hoạt động do công tác lập kế hoạch đề xuất. Điều này có thể chuyển dịch hệ thống từ trạng thái ban đầu đến gần hơn với trạng thái mục tiêu.

10 Các ứng dụng của hệ thống AI

10.1 Tổng quan

Các hệ thống AI có thể hỗ tự động hóa ra quyết định hoặc đưa ra các đề xuất và hỗ trợ tự động hóa các tác vụ nhất định. Nó có các ứng dụng trong các ngành công nghiệp khác nhau bao gồm:

- Nông nghiệp và trồng trọt;
- Ô tô;
- Công nghệ tài chính và ngân hàng;
- Quốc phòng;
- Giáo dục;
- Năng lượng;
- Chăm sóc sức khỏe;
- Pháp luật;
- Chế tạo;
- Truyền thông và giải trí;
- Thực tế hỗn hợp;
- Khu vực công;
- Bán lẻ và tiếp thị;
- An ninh;
- Công nghệ không gian;
- Viễn thông.

Ví dụ về các ứng dụng AI được đưa ra trong mục 10.2 đến 10.4.

10.2 Phát hiện gian lận

Gian lận là lừa dối để đạt được lợi ích. Gian lận xuất hiện trong nhiều lĩnh vực bao gồm:

- Làm giả tiền và tài liệu;
- Đánh cắp thẻ tín dụng và tài liệu;
- Thông tin liên lạc cá nhân như thư điện tử;
- Giả mạo hoặc đánh cắp danh tính

Dưới đây là các ví dụ về ứng dụng AI trong phát hiện gian lận:

- Xác định các khoản chi trả qua thẻ tín dụng gian lận;
- Xác định các ứng dụng cho vay hoặc tín dụng gian lận;
- Xác định các yêu cầu bảo hiểm gian lận;
- Xác định truy cập tài khoản gian lận.

10.3 Xe tự động

Các phương tiện không người lái được dự báo có thể trở nên phổ biến. Ngày nay nhiều công nghệ hỗ trợ bởi AI được áp dụng trên ô tô, chẳng hạn như các tính năng hỗ trợ người lái. Dưới đây là các ví dụ về ứng dụng AI trong phương tiện di chuyển:

- Định tuyến tối ưu (ví dụ: tìm tuyến đường nhanh nhất với điều kiện giao thông hiện tại);
- Chuyển làn đường tự động;
- Tránh vật thể (ví dụ: thao tác tự động của phanh, ga và đánh lái dựa trên việc phiên dịch các tín hiệu từ camera, cảm biến phát hiện ánh sáng và khoảng cách);
- Tự động hóa đi lại từ điểm A đến điểm B.

Các phương tiện di chuyển tự động dựa trên các công nghệ AI như thị giác máy tính và lập kế hoạch.

10.4 Bảo trì theo dự đoán

Không giống như bảo trì mang tính phòng ngừa, trong đó bảo trì dựa trên tuổi thọ mong muốn của các bộ phận (ví dụ: thời gian trung bình giữa các lần hỏng hóc), bảo trì theo dự đoán cung cấp hoạt động bảo dưỡng hoặc thay thế các bộ phận dựa trên các quan sát hành vi hoặc hiệu năng hiện tại cũng như tuổi thọ dự kiến của chúng. Dưới đây là các ví dụ về ứng dụng AI trong bảo trì dự đoán:

- Phát hiện các khoảng trống bên dưới đường ray (có thể dẫn đến tàu bị trật bánh);
- Phát hiện nhựa đường bị nứt hoặc hư hỏng;
- Phát hiện hỏng ổ trục trong động cơ điện;
- Phát hiện những dao động bất thường về công suất trong hệ thống cung cấp điện.

Phụ lục A

(Tham khảo)

Ánh xạ vòng đời của hệ thống AI với định nghĩa của OECD về vòng đời của hệ thống AI

Công cụ pháp lý của OECD đã xuất bản "Khuyến nghị của Hội đồng trí tuệ nhân tạo" [26].

Được đưa vào trong tiêu chuẩn này là những nội dung sau:

"Theo đề xuất của Ủy ban về Chính sách kinh tế số:

I. ĐỒNG Ý với mục đích của Khuyến nghị này, các thuật ngữ sau đây nên được hiểu như sau:

– Vòng đời hệ thống AI: Các giai đoạn vòng đời của hệ thống AI bao gồm:

i) 'thiết kế, dữ liệu và mô hình'; là một trình tự phụ thuộc vào ngữ cảnh bao gồm lập kế hoạch và thiết kế, thu thập và xử lý dữ liệu, cũng như xây dựng mô hình;

ii) 'xác minh và thẩm định';

iii) 'triển khai';

iv) 'vận hành và theo dõi'.

Các giai đoạn này thường diễn ra theo phương thức lặp đi lặp lại và không nhất thiết theo tuần tự. Quyết định cho ngừng sử dụng một hệ thống AI có thể xảy ra bất kỳ lúc nào trong giai đoạn vận hành và giám sát".

và

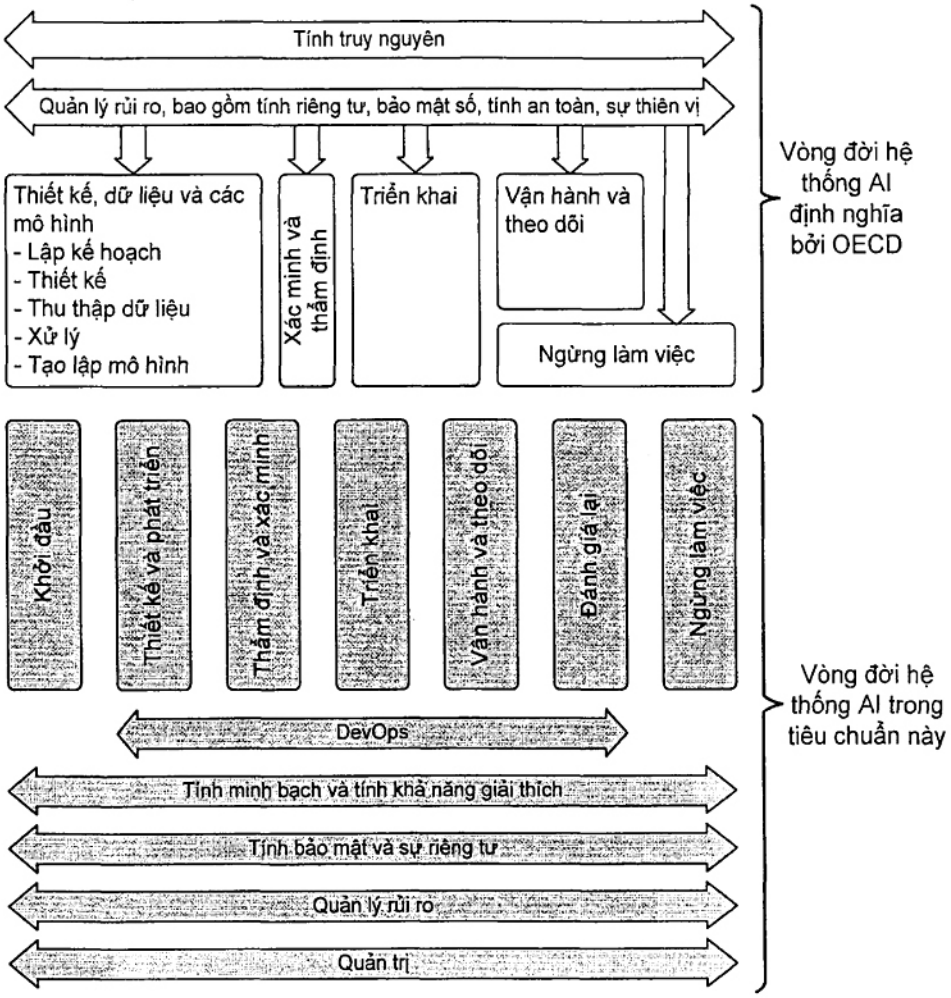
"Độ bền vững, bảo mật và an toàn

a) Hệ thống AI phải mạnh mẽ, bảo mật và an toàn trong toàn bộ vòng đời của chúng trong điều kiện sử dụng bình thường, lường trước việc sử dụng hoặc lạm dụng, hoặc các điều kiện bất lợi khác, chúng hoạt động một cách phù hợp và không gây ra rủi ro an toàn không đáng có.

b) Để đạt được mục tiêu này, các tác nhân AI phải đảm bảo khả năng truy xuất, bao gồm những vấn đề liên quan đến bộ dữ liệu, quá trình và các quyết định được đưa ra trong vòng đời của hệ thống AI, cho phép phân tích các kết quả và phản hồi của hệ thống AI đối với yêu cầu truy vấn, phù hợp với bối cảnh và tân tiến nhất.

c) Các tác nhân AI cần dựa trên vai trò, bối cảnh và khả năng hành động của chúng, áp dụng phương pháp quản lý rủi ro có hệ thống cho từng giai đoạn của vòng đời hệ thống AI một cách liên tục để giải quyết các rủi ro liên quan đến hệ thống AI, bao gồm quyền riêng tư, bảo mật số, tính an toàn và sự thiên vị".

Hình A.1 cho thấy định nghĩa này về vòng đời của hệ thống AI có thể được ánh xạ như thế nào với vòng đời của hệ thống AI được mô tả trong Điều 6.



Hình A.1 – Ánh xạ đến vòng đời hệ thống AI của OECD

Thư mục tài liệu tham khảo

[1] TCVN 10539:2014 ISO/IEC 12207:2008, Kỹ thuật hệ thống và phần mềm – Các quá trình vòng đời phần mềm (Systems and software engineering — Software life cycle processes).

[2] ISO/IEC 15288:2015, Systems and software engineering — System life cycle processes.

[3] ISO/IEC/IEEE 15289:2019, Systems and software engineering — Content of life-cycle informationitems (documentation).

[4] ISO/IEC 17788:2014, Information technology — Cloud computing — Overview and vocabulary.

[5] ISO/IEC 17789:2014, Information technology — Cloud computing — Reference architecture.

[6] TCVN 13238:2020 ISO/IEC 20546:2019, Công nghệ thông tin – Dữ liệu lớn – Tổng quan và từ vựng (Information technology — Big data — Overview and vocabulary).

[7] ISO/IEC 20547-3:2020, Information technology — Big data reference architecture — Part 3: Reference architecture.

[8] ISO/IEC 20889:2018, Privacy enhancing data de-identification terminology and classification oftechniques.

[9] ISO/IEC 20924:2021, Information technology — Internet of Things (IoT) — Vocabulary.

[10] ISO/IEC 23053, Information technology — Artificial Intelligence (AI) — Framework for ArtificialIntelligence (AI) Systems Using Machine Learning (ML).

[11] ISO/IEC/TR 23188:2020, Information technology — Cloud computing — Edge computing landscape.

[12] ISO/IEC 23894, Information technology — Artificial intelligence — Risk management.

[13] ISO/IEC/TR 24027:2021, Information technology — Artificial intelligence (AI) — Bias in AI systemsand AI aided decision making.

[14] ISO/IEC/TR 24028:2020, Information technology — Artificial intelligence — Overview oftrustworthiness in artificial intelligence.

[15] ISO/IEC/TR 24029-1:2021, Artificial Intelligence (AI) — Assessment of the robustness of neuralnetworks — Part 1: Overview.

[16] ISO/IEC 27040:2015, Information technology — Security techniques — Storage security.

[17] TCVN 13117:2020 ISO/IEC 30141:2018, Internet vạn vật – Kiến trúc tham chiếu (Internet of Things (IoT) — Reference Architecture).

[18] Graves A., Abdel-rahman Mohamed, Geoffrey E. Hinton, Speech recognition with deep

recurrentneural networks, IEEE International Conference on Acoustics, Speech and Signal Processing,2013, DOI:10.1109/ICASSP.2013.6638947.

[19] Jena A., Reasoners and rule engines: Jena inference support, https://jena.apache.org/documentation/inference/index.html.

[20] Artificial Intelligence Methodologies and Their Application to Diabetes,https://pubmed.ncbi.nlm.nih.gov/28539087/.

[21] Elman Jeffrey L., Finding structure in time, Cognitive science 14.2 (1990): 179-211.

[22] Hochreiter Sepp, Schmidhuber Juergen, Long short-term memory, Neural computation 9.8(1997): 1735-1780.

[23] Japanese Society of Artificial Intelligence, AI Map Beta, https://www.ai-gakkai.or.jp/pdf/aimap/AIMap_EN_20190606.pdf.©ISO/IEC 2022 – All rights reserved 59ISO/IEC FDIS 22989:2022(E).

[24] Zadeh L.A., Soft computing and fuzzy logic, IEEE Software, 1994, vol. 11, issue 6.

[25] Rigla M., Gema García-Sáez B., Pons, M., Artificial Intelligence Methodologies and TheirApplication to Diabetes Hernando, Journal of diabetes science and technology, 2018, DOI:10.1177/1932296817710475.

[26] Recommendation of the councile on artificial intelligence,https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

[27] Rozenblit J.W., Cognitive computing: Principles, architectures, and applications, In: Proc. 19thEuropean Conf. on Modelling and Simulation (ECMS) (2005).

[28] S. C., The CRISP-DM model: the new blueprint for data mining, J Data Warehousing (2000); 5:13 — 22.

[29] Stuart Russell and Peter Norvig, Artificial Intelligence: A Modern Approach, (3rd Edition) (Essex, England: Pearson, 2009).

[30] Taxonomy and Definitions for Terms Related to Driving Automation Systems for OnRoad Motor Vehicles, SAE — On-Road Automated Driving (ORAD) committee, https://Saemobilus.Sae.org/content/J3016_201806/.