TCVN **T I Ê U C H U Ẩ N Q U Ố C G I A**

TCVN DIN SPEC 92001-1:202x

TRÍ TUỆ NHÂN TẠO – QUY TRÌNH VÒNG ĐỜI VÀ YÊU CẦU CHẤT LƯỢNG – PHẦN 1: SIÊU MÔ HÌNH CHẤT LƯỢNG

Artificial Intelligence – Life Cycle Processes and Quality Requirements – Part 1: Quality Meta Model

(Tài liệu nghiệm thu Bộ)

HÀ NỘI - 2022

Lời nói đầu

TCVN DIN SPEC 92001-1:202x được xây dựng trên cơ sở tham khảo tài liệu DIN SPEC 92001-1 (2019) “Artificial Intelligence – Life Cycle Processes and Quality Requirements – Part 1: Quality Meta Model” của DIN SPEC (PAS).

TCVN DIN SPEC 92001-1:202x do Viện Khoa học Kỹ thuật Bưu điện - Học viện Công nghệ Bưu chính Viễn thông 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.

MỤC LỤC

Trí tuệ nhân tạo – Quy trình vòng đời và yêu cầu chất lượng – Phần 1: Siêu mô hình chất lượng

Artificial Intelligence - Life Cycle Processes and Quality Requirements - Part 1: Quality Meta Model

1 Phạm vi áp dụng

1.1 Mục đích

Mục đích của tiêu chuẩn này là thiết lập vòng đời đảm bảo chất lượng và minh bạch của các mô-đun trí tuệ nhân tạo (AI). Các tiêu chí chất lượng quan trọng được xác định và các vấn đề AI cụ thể sẽ được giải quyết. Để đạt được điều này, tiêu chuẩn này trình bày một tập hợp các yêu cầu chất lượng được cấu trúc trong siêu mô hình chất lượng cụ thể cho AI.

Điều quan trọng cần lưu ý rằng không phải tất cả các mô-đun AI đều áp đặt các yêu cầu chất lượng giống nhau. Do đó, tiêu chuẩn này đề xuất sự khác biệt giữa các mô-đun AI liên quan đến tính an toàn, bảo mật, quyền riêng tư và mức độ phù hợp về đạo đức của chúng. An toàn, bảo mật, quyền riêng tư hoặc đạo đức của một mô-đun AI yêu cầu phải xem xét và đáp ứng tất cả các yêu cầu chất lượng, trong khi các yêu cầu này ít nghiêm ngặt hơn, khi không đưa ra mức độ phù hợp này.

Tiêu chuẩn này phác thảo và xác định ba trụ cột chất lượng trung tâm là chức năng & hiệu năng, độ bền vững và tính dễ hiểu. Lưu ý rằng những trụ cột này không hoàn toàn tách rời, nhưng được đề xuất tạo điều kiện thuận lợi phân biệt một danh mục có cấu trúc của các yêu cầu chất lượng cụ thể. Các yêu cầu chất lượng này cũng được liên kết với các giai đoạn vòng đời và quy trình vòng đời khác nhau. Bằng cách này, sẽ trở nên rõ ràng khi và trong ngữ cảnh các yêu cầu nhất định phải được đáp ứng. Hơn nữa, sự khác biệt giữa các yêu cầu chất lượng liên quan tới các yếu tố ảnh hưởng đến mô hình, dữ liệu, nền tảng hoặc môi trường của mô-đun AI được tạo ra.

Tất cả các cân nhắc được tập hợp trong siêu mô hình chất lượng AI, được giới thiệu trong phần 4. Nó bao gồm đánh giá rủi ro, trụ cột chất lượng, các giai đoạn và quy trình vòng đời. Hơn nữa, cũng xác định sự khác biệt giữa mô hình, dữ liệu, nền tảng và môi trường. Trong DIN SPEC 92001-2, các yêu cầu AI cụ thể được đưa ra có liên quan đến các khía cạnh khác nhau của siêu mô hình chất lượng AI.

1.2 Lĩnh vực áp dụng

Tiêu chuẩn này áp dụng cho tất cả các giai đoạn vòng đời của mô-đun AI - khái niệm, phát triển, triển khai, vận hành và ngừng hoạt động - đồng thời đề cập đến các quy trình vòng đời khác nhau. Do thực tế rằng các công nghệ AI được sử dụng cho một phạm vi rộng lớn của các nhiệm vụ khác nhau, nên tiêu chuẩn này không chỉ nhắm đến một lĩnh vực cụ thể mà còn áp dụng cho các công ty và sản phẩm AI trên tất cả các lĩnh vực.

Tiêu chuẩn này áp dụng cho tất cả các loại mô-đun AI bao gồm ML và các hệ thống chuyên gia.

1.3 Giới hạn

Tiêu chuẩn này không định nghĩa hoặc liệt kê các thuật toán, phương pháp hoặc công nghệ là một phần của AI. Do đó, người dùng tiêu chuẩn này được yêu cầu đánh giá về siêu mô hình chất lượng AI đã định và việc áp dụng các yêu cầu chất lượng AI có liên quan.

Mặc dù yêu cầu đánh giá hành vi đạo đức trong quá trình phát triển mô-đun AI, nhưng tiêu chuẩn này sẽ không cung cấp bất kỳ yêu cầu cụ thể nào để xác định hành vi đạo đức.

Các cân nhắc vòng đời phần mềm trong tiêu chuẩn này tương thích với ISO/IEC/IEEE 12207:2017, Hệ thống và kỹ thuật phần mềm - Quy trình vòng đời phần mềm [3]. Tiêu chuẩn này giải thích các thuật ngữ, định nghĩa hoặc quy trình cụ thể của AI.

Tiêu chuẩn này đề xuất tách biệt giữa các mô-đun AI có mức độ rủi ro cao và thấp liên quan đến an toàn, bảo mật, quyền riêng tư và mức độ phù hợp về đạo đức. Nó cũng cung cấp các khía cạnh liên quan trong bối cảnh đánh giá rủi ro. Tiêu chuẩn này hoặc không thiết lập quy trình đánh giá rủi ro nghiêm ngặt cũng như không thiết lập khuôn khổ thiết kế đạo đức. Tuy nhiên, nó bị hạn chế bởi quy tắc ứng xử đạo đức của mỗi tổ chức. Tuân thủ các quy định được cho là đúng. Hơn nữa, các bên liên quan của tiêu chuẩn này được yêu cầu đánh giá hồ sơ rủi ro mô-đun AI của họ.

Các yêu cầu chất lượng được liệt kê trong DIN SPEC 92001-2 không phải là miền cụ thể. Có khả năng mở rộng các yêu cầu được đưa ra trong DIN SPEC 92001-2 cho các lĩnh vực ứng dụng cụ thể của AI trong bước tiêu chuẩn hóa tiếp theo.

1.4 Giới thiệu

Trí tuệ nhân tạo (AI) là một lĩnh vực phức tạp và đang phát triển nhanh chóng. Nó là một nhánh của khoa học máy tính bao gồm các nhiệm vụ liên quan đến trí thông minh của con người [1]. Đặc biệt, những tiến bộ gần đây trong lĩnh vực Máy học (ML), một mô hình cho phép các hệ thống tự động cải thiện hiệu suất của chúng bằng cách quan sát dữ liệu [2], đã dẫn đến sự gia tăng triển khai các công nghệ dựa vào AI bởi các công ty trên hầu hết các lĩnh vực.

Các thành phần AI phải đối mặt với các vấn đề về chất lượng phần mềm truyền thống và các vấn đề mới xảy ra ở cấp độ hệ thống. Tính mới quan trọng nhất là một số loại mô-đun AI chẳng hạn như Mạng nơ-ron nhân tạo - các quyết định của chúng và logic cơ bản - thường không thể hiểu đầy đủ bằng cách đánh giá mã. Ngoài ra, môi trường AI không cố định có thể dẫn đến những quyết định không thể đoán trước.

Vì những lý do này, việc đánh giá chất lượng của mô-đun AI vẫn là một thách thức lớn. Nó trở nên khó khăn hơn trong việc xác nhận, thẩm định và xác thực tính hợp lệ của một mô-đun AI trong quá trình hình thành, phát triển, triển khai, vận hành và ngừng hoạt động, vốn là những nhiệm vụ có phạm vi rộng lớn.

Tiêu chuẩn này nhằm mục đích cung cấp một cách tiếp cận thống nhất để đảm bảo chất lượng AI. Nó có thể áp dụng cho tất cả các thành phần AI và vượt xa các yêu cầu chất lượng thông thường trong các nhánh phát triển phần mềm khác. Các yêu cầu bổ sung giải quyết các thách thức cụ thể liên quan đến việc sử dụng các mô-đun AI, chẳng hạn như thành phần của tập dữ liệu, lựa chọn mô hình và môi trường không cố định. Các yêu cầu chất lượng phát sinh từ những thách thức mới này được sắp xếp theo ba trụ cột chất lượng là chức năng & hiệu suất, độ bền vững và tính dễ hiểu.

Tiêu chuẩn này giới thiệu một siêu mô hình chất lượng AI để phác thảo các khía cạnh chính của chất lượng AI bao gồm các trụ cột chất lượng AI đã đề cập ở trên. Để phân tích chất lượng AI, một phương pháp đánh giá rủi ro và vòng đời phần mềm phù hợp được cung cấp. Vòng đời AI nhất định phù hợp với tiêu chuẩn quốc tế về hệ thống và công nghệ phần mềm [3]. Phần thứ hai của thông số kỹ thuật này, DIN SPEC 92001-2, sẽ cung cấp các yêu cầu chất lượng AI cụ thể.

2 Tài liệu viện dẫn

Không có tài liệu tham khảo viện dẫn trong tiêu chuẩn này.

3 Thuật ngữ và định nghĩa

Đối với các mục đích của tiêu chuẩn này, các thuật ngữ và định nghĩa sau đây được áp dụng.

DIN và DKE duy trì cơ sở dữ liệu thuật ngữ để sử dụng trong tiêu chuẩn hóa tại các địa chỉ sau:

- DIN-TERMinologieportal: có tại https://www.din.de/go/din-term.

- DKE-IEV: có tại http://www.dke.de/DKE-IEV.

3.1

Trí tuệ nhân tạo (AI) (artificial intelligence - AI)

Lĩnh vực gồm nhiều ngành học thuật, thường được coi là một nhánh của khoa học máy tính, xử lý các mô hình và hệ thống để thực hiện các chức năng chung liên quan đến trí thông minh của con người, chẳng hạn như lý luận và kiến thức [1].

3.2

Mô-đun trí tuệ nhân tạo (AI) (artificial intelligence (AI) module)

Mô-đun phần mềm bao gồm các thuật toán AI.

3.3

Độ khả dụng (availability)

<khả năng truy cập/khả năng sử dụng> thuộc tính có thể truy cập và có thể sử dụng được theo yêu cầu của một thực thể được ủy quyền

[NGUỒN: ISO/TS 21089:2018]

3.4

Mô-đun tiện ích (comfort module)

Mô-đun phần mềm không có liên quan đến an toàn hoặc bảo mật.

3.5

Thành phần (component)

Phần tử mức phi hệ thống có thể tách rời về mặt logic và kỹ thuật và gồm nhiều hơn một phần cứng hoặc của một hoặc nhiều đơn vị phần mềm.

CHÚ THÍCH 1: Một thành phần là một phần của một hệ thống.

[NGUỒN: ISO 26262-1:2011]

3.6

Tính bí mật (confidentiality)

<không được tiết lộ> thuộc tính mà thông tin không được cung cấp hoặc tiết lộ trái phép cho các cá nhân, thực thể hoặc quy trình.

CHÚ THÍCH 1: Tính bí mật dữ liệu là một từ đồng nghĩa.

[NGUỒN: ISO/TS 21089:2018, được sửa đổi – Chú thích 1 đưa vào đã được bổ sung thêm.]

3.7

Dữ liệu (data)

Biểu thị thông tin có thể diễn giải lại theo một cách chính thức có thể phù hợp cho truyền thông, giải thích hoặc xử lý.

CHÚ THÍCH 1: Điều này có thể bao gồm kiến thức chuyên môn.

[NGUỒN: ISO/IEC 25000:2014, được sửa đổi – Chú thích 1 đưa vào đã được bổ sung thêm.]

3.8

Toàn vẹn dữ liệu (data integrity)

Thuộc tính mà dữ liệu không bị thay đổi hoặc phá hủy một cách trái phép.

CHÚ THÍCH 1: Thường được sử dụng đồng nghĩa với tính toàn vẹn.

[NGUỒN: ISO 24534-5:2011, được sửa đổi – Chú thích 1 đưa vào đã được bổ sung thêm.]

3.9

Phần tử (element)

Các đơn vị phần mềm và các bộ phận phần cứng.

3.10

Môi trường (environment)

<hệ thống> bối cảnh xác định cài đặt và tình huống của tất cả các ảnh hưởng trên một hệ thống.

[NGUỒN: ISO/IEC/IEEE 12207:2017]

3.11

Lỗi (error)

Sự khác biệt giữa một giá trị hoặc điều kiện được tính toán, được quan sát hoặc được đo lường, và giá trị hoặc điều kiện thực sự, được chỉ định hoặc đúng về lý thuyết.

CHÚ THÍCH 1: Một lỗi có thể phát sinh do các điều kiện vận hành không lường trước được hoặc do lỗi bên trong hệ thống, hệ thống con hoặc thành phần đang được xem xét.

CHÚ THÍCH 2: Một lỗi có thể tự biểu hiện như một lỗi bên trong phần tử được xem xét và lỗi này cuối cùng có thể gây ra hư hỏng.

[NGUỒN: ISO 26262-1:2011]

3.12

Hỏng (failure)

Chấm dứt khả năng của một phần tử, để thực hiện một chức năng theo yêu cầu.

[NGUỒN: ISO 26262-1:2011, được sửa đổi – Chú thích 1 đưa vào đã bị xóa bỏ.]

3.13

Sai sót (fault)

Tình trạng bất thường có thể khiến một phần tử hoặc một mục bị hỏng.

[NGUỒN: ISO 26262-1:2011, được sửa đổi – Chú thích 1 đưa vào và Chú thích 2 đưa vào đã bị xóa bỏ.]

3.14

Phần cứng (hardware part)

Phần cứng không thể được chia nhỏ.

[NGUỒN: ISO 26262-1:2011]

3.15

Gây hại (harm)

Thương tích thân thể hoặc thiệt hại cho sức khỏe của con người.

[NGUỒN: ISO 26262-1:2011]

3.16

Siêu tham số (hyperparameters)

Các biến bên ngoài kiểm soát mô hình.

VÍ DỤ Tốc độ học của phương pháp giảm dần độ dốc là một siêu tham số.

CHÚ THÍCH 1: Siêu tham số không thể ước tính từ dữ liệu. Việc xác định siêu tham số ở đó phù hợp tốt với mô hình của một vấn đề cụ thể thường khó khăn trong thực hành. Do đó, các phương pháp kinh nghiệm được sử dụng.

[NGUỒN: [4]]

3.17

Mô hình suy luận (inference model)

Một phần tử cụ thể của mô hình không gian giải quyết một nhiệm vụ cụ thể ở một mức độ nhất định.

CHÚ THÍCH 1: Mô hình suy luận bao gồm siêu tham số được cố định. Vì vậy, nó đại diện cho một kiến trúc mô hình riêng biệt.

3.18

Sự toàn vẹn (integrity)

Được thiết kế sao cho không thể sửa đổi bất kỳ thông tin được lưu trữ điện tử nào, nếu không có sự cho phép thích hợp.

CHÚ THÍCH 1: Thường được sử dụng đồng nghĩa với tính toàn vẹn của dữ liệu.

[NGUỒN: ISO 17364:2013, được sửa đổi – Chú thích 1 đưa vào đã được bổ sung thêm.]

3.19

Vòng đời (life cycle)

Sự tiến triển của một hệ thống, sản phẩm, dịch vụ, đề án hoặc thực thể do con người tạo ra khác từ khi nhận thức cho đến khi dừng hoạt động.

[NGUỒN: ISO/IEC/IEEE 12207:2017]

3.20

Mô hình vòng đời (life cycle model)

Khuôn khổ các quy trình và hoạt động liên quan đến vòng đời, có thể tổ chức thành các giai đoạn, hoạt động như một tham chiếu chung cho giao tiếp và hiểu biết.

[NGUỒN: ISO/IEC/IEEE 12207:2017]

3.21

Học máy (machine learning)

Quá trình sử dụng các thuật toán thay vì thủ tục mã hóa cho phép học từ dữ liệu hiện có để dự đoán kết quả trong tương lai

3.22

Siêu mô hình (metamodel)

Mô hình xác định các khái niệm và mối quan hệ của chúng với một số ký hiệu mô hình hóa.

[NGUỒN: ISO/IEC 15909-2:2011, được sửa đổi – Chú thích 1 đưa vào đã bị xóa bỏ.]

3.23

Mô hình (model)

Sự trừu tượng của một số khía cạnh trong thực tế.

CHÚ THÍCH 1: Một mô hình có thể là một biểu diễn vật lý, toán học hoặc logic của một hệ thống, thực thể, hiện tượng hoặc quá trình.

[NGUỒN: ISO 19103:2015, được sửa đổi – Chú thích 1 đưa vào tương thích với ISO/IEC 18023-1:2006]

3.24

Mô hình không gian (model space)

Tập hợp các cách tiếp cận tiềm năng để giải quyết một nhiệm vụ ở một mức độ nhất định.

CHÚ THÍCH 1: Trong mô hình không gian, siêu tham số chưa được cố định.

3.25

(Mô hình) các tham số ((model) parameters)

Các biến cấu hình mô hình bên trong.

VÍ DỤ Trọng số trong mạng thần kinh là các tham số mô hình.

CHÚ THÍCH 1: Trong ML, mục tiêu là tìm các tham số mô hình. Chúng có thể được ước tính từ dữ liệu.

CHÚ THÍCH 2: Các tham số mô hình được cố định trong mô hình suy luận.

[NGUỒN: [4]]

3.26

Học ngoại tuyến (offline learning)

Học từ một lượng dữ liệu nhất định và tĩnh.

3.27

Học trực tuyến (online learning)

Cập nhật liên tục mô hình dựa trên các lô dữ liệu mới.

3.28

Nền tảng (platform)

Phần cứng, hệ điều hành và môi trường thời gian chạy.

3.29

Quy trình (process)

Tập hợp các hoạt động có liên quan hoặc tương tác với nhau để đạt được một tập hợp các mục tiêu.

3.30

Sản phẩm (product)

Kết quả của một quá trình.

CHÚ THÍCH 1: Có bốn loại sản phẩm chung đã được thống nhất: phần cứng (ví dụ: bộ phận cơ khí của động cơ); phần mềm (ví dụ: các thủ tục chương trình máy tính và có thể là tài liệu và dữ liệu liên quan); các dịch vụ (ví dụ: vận tải); và các vật liệu chế biến (ví dụ: chất bôi trơn). Phần cứng và vật liệu chế biến nói chung là sản phẩm hữu hình, trong khi phần mềm hoặc dịch vụ nói chung là vô hình.

[NGUỒN: ISO/IEC/IEEE 12207:2017]

3.31

Đề án (project)

Nỗ lực với các tiêu chí bắt đầu và kết thúc đã xác định được thực hiện để tạo ra một sản phẩm hoặc dịch vụ phù hợp với các nguồn lực và yêu cầu cụ thể.

CHÚ THÍCH 1: Một đề án đôi khi được xem như một quá trình duy nhất bao gồm các hoạt động được phối hợp và kiểm soát và bao gồm các hoạt động từ các quá trình Quản lý Kỹ thuật và các quy trình Kỹ thuật được định nghĩa trong tiêu chuẩn này.

[NGUỒN: ISO/IEC/IEEE 12207:2017]

3.32

Bảo đảm chất lượng (quality assurance)

Phần quản lý chất lượng tập trung vào việc cung cấp niềm tin rằng các yêu cầu chất lượng sẽ được đáp ứng.

[NGUỒN: ISO 9000:2015]

3.33

Trụ cột chất lượng (quality pillar)

Một trong ba đặc điểm chất lượng AI: chức năng & hiệu năng, độ bền vững và tính có thể bao hàm.

3.34

Yêu cầu chất lượng (quality requirement)

Yêu cầu cần thiết để bảo đảm chất lượng.

3.35

Độ tin cậy (reliability)

Khả năng của một thiết bị hoặc một hệ thống thực hiện chức năng dự kiến của nó trong các điều kiện sử dụng nhất định trong một khoảng thời gian hoặc số chu kỳ đã được xác định.

[NGUỒN: ISO/TS 17574:2017]

3.36

Yêu cầu (requirement)

Tuyên bố biên dịch hoặc thể hiện nhu cầu và các ràng buộc và điều kiện có liên quan của nó.

[NGUỒN: ISO/IEC/IEEE 12207:2017]

3.37

Rủi ro (risk)

Ảnh hưởng của sự không chắc chắn lên các đối tượng.

CHÚ THÍCH 1: Ảnh hưởng là sự sai lệch so với dự kiến - tích cực hoặc tiêu cực. Một ảnh hưởng tích cực còn được gọi là một cơ hội.

CHÚ THÍCH 2: Các mục tiêu có thể có các khía cạnh khác nhau (chẳng hạn như mục tiêu tài chính, sức khỏe và an toàn, và môi trường) và có thể áp dụng ở các cấp độ khác nhau (chẳng hạn như chiến lược, toàn tổ chức, đề án, sản phẩm và quá trình).

CHÚ THÍCH 3: Rủi ro thường được đặc trưng bởi tham chiếu đến các sự kiện và hậu quả tiềm ẩn, hoặc sự kết hợp của những điều này.

CHÚ THÍCH 4: Rủi ro thường được thể hiện dưới dạng kết hợp các hậu quả của một sự kiện (bao gồm cả những thay đổi về hoàn cảnh) và khả năng xảy ra liên quan.

CHÚ THÍCH 5: Độ không đảm bảo là trạng thái, thậm chí một phần, thiếu thông tin liên quan đến hiểu biết hoặc kiến thức về một sự kiện, hậu quả của nó hoặc khả năng xảy ra.

[NGUỒN: ISO/IEC/IEEE 12207:2017]

3.38

An toàn (safety)

Kỳ vọng rằng một hệ thống, trong các điều kiện xác định, không dẫn đến tình trạng mà cuộc sống, sức khỏe, tài sản hoặc môi trường của con người bị đe dọa.

CHÚ THÍCH 1: Cho đến nay, thuật ngữ này thường được sử dụng chung theo nghĩa độ tin cậy. Theo nghĩa hạn chế của nó, có nghĩa là khả năng của một cấu trúc chống lại tất cả các tác động, cũng như một số hiện tượng ngẫu nhiên cụ thể mà nó sẽ phải chịu đựng trong quá trình xây dựng và sử dụng theo dự kiến (các trạng thái giới hạn cuối cùng có liên quan).

CHÚ THÍCH 2: An toàn có nghĩa là không có rủi ro không thể chấp nhận được.

[NGUỒN: ISO/IEC/IEEE 12207:2017, được sửa đổi – Chú thích 1 đưa vào phù hợp với ISO 8930:1987, Chú thích 2 đưa vào phù hợp với ISO 5840-3:2013]

3.39

Bảo mật (security)

Chống lại các hoạt động cố ý được thiết kế để gây nên tổn hại hoặc thiệt hại cho hoặc bởi chuỗi cung ứng.

CHÚ THÍCH 1: Bảo mật là sự kết hợp của tính bí mật, tính toàn vẹn và tính sẵn sàng.

CHÚ THÍCH 2: Đánh giá bảo mật bao gồm các câu hỏi về rủi ro có thể xảy ra đối với một mô-đun AI gây ra bởi con người hoặc các thành phần phần mềm khác gây ra và các khía cạnh riêng tư.

[NGUỒN: ISO/TR 12773-2:2009, được sửa đổi – Chú thích 1 đưa vào phù hợp với ISO 28001:2007, Chú thích 2 đưa vào đã được bổ sung thêm]

3.40

Yếu tố phần mềm (software element)

Yếu tố hệ thống đó là phần mềm

[NGUỒN: ISO/IEC/IEEE 12207:2017]

3.41

Mô-đun phần mềm (software module)

Đơn vị chương trình đóng gói các thủ tục và cấu trúc dữ liệu cần thiết để cung cấp một chức năng mong muốn nhất định.

3.42

Giai đoạn (stage)

Các pha khác nhau trong vòng đời của một phần mềm hoặc mô-đun phần mềm.

3.43

Bên liên quan (stakeholder)

Cá nhân hoặc tổ chức có quyền, chia sẻ, đòi hỏi hoặc lợi ích trong một hệ thống hoặc trong việc sở hữu các đặc điểm đáp ứng nhu cầu và mong muốn của họ

VÍ DỤ: Người dùng cuối, tổ chức người dùng cuối, người ủng hộ, nhà phát triển, nhà sản xuất, nhà đào tạo, nhà bảo trì, người định đoạt, người mua, tổ chức cung cấp và cơ quan quản lý.

CHÚ THÍCH 1: Một số bên liên quan có thể có lợi ích đối lập nhau hoặc đối lập hệ thống.

[NGUỒN: ISO/IEC/IEEE 12207:2017]

3.44

Hệ thống (system)

Sự kết hợp của các yếu tố tương tác được tổ chức để đạt được một hoặc nhiều mục đích đã định.

CHÚ THÍCH 1: Một hệ thống đôi khi được xem như một sản phẩm hoặc dịch vụ mà nó cung cấp.

CHÚ THÍCH 2: Trong thực tế, việc giải thích ý nghĩa của nó thường được làm rõ bằng cách sử dụng danh từ kết hợp, ví dụ: hệ thống máy bay hoặc hệ thống quản lý cơ sở dữ liệu. Ngoài ra, từ hệ thống được thay thế đơn giản bằng một từ đồng nghĩa phụ thuộc vào nội dung, ví dụ: máy bay hoặc cơ sở dữ liệu, mặc dù điều này có khả năng che lấp quan điểm nguyên tắc hệ thống.

CHÚ THÍCH 3: Một hệ thống có thể bao gồm thiết bị liên quan, phương tiện, vật liệu, phần mềm, phần sụn, tài liệu kỹ thuật, dịch vụ và nhân sự được yêu cầu cho các hoạt động và hỗ trợ ở mức độ cần thiết để sử dụng trong môi trường dự kiến của nó.

CHÚ THÍCH 4: Xem để so sánh: hệ thống cho phép, hệ thống quan tâm, hệ thống của các hệ thống.

[NGUỒN: ISO/IEC/IEEE 12207:2017]

3.45

Phần tử hệ thống (system element)

Thành viên của một tập hợp các yếu tố cấu thành một hệ thống.

VÍ DỤ: Phần cứng, phần mềm, dữ liệu, con người, quy trình (ví dụ: quy trình cung cấp dịch vụ cho người dùng), thủ tục (ví dụ: hướng dẫn vận hành), phương tiện, vật liệu và các thực thể xuất hiện một cách tự nhiên hoặc bất kỳ sự kết hợp nào.

CHÚ THÍCH 1: Một phần tử hệ thống là một phần tách rời của một hệ thống có thể được triển khai để đáp ứng đầy đủ các yêu cầu quy định.

[NGUỒN: ISO/IEC/IEEE 12207:2017]

3.46

Nhiệm vụ (task)

Hành động được yêu cầu, khuyến nghị hoặc được phép, nhằm đóng góp vào việc đạt được một hoặc nhiều kết quả của một quá trình.

[NGUỒN: ISO/IEC/IEEE 12207:2017]

3.47

Kiểm thử (testing)

Quá trình lập kế hoạch, chuẩn bị, triển khai và thực hiện kiểm thử để phát hiện lỗi hoặc sự bất thường trong quá trình triển khai hệ thống có thể dẫn đến lỗi hệ thống.

CHÚ THÍCH 1: Tùy thuộc vào phạm vi kiểm thử, các mức kiểm thử khác nhau có thể được áp dụng: Bộ kiểm thử chứng minh tính đúng đắn của một bộ/mô-đun phần mềm đơn lẻ, kiểm thử tích hợp kiểm tra triển khai giao diện phù hợp, kiểm thử hệ thống cho thấy các lỗi của hệ thống được kiểm thử như vậy và kiểm thử chấp nhận thường được thực hiện bởi khách hàng trên nền tảng đích để người dùng tương tác chính xác với hệ thống.

3.48

Người dùng (user)

Cá nhân hoặc tổ chức sử dụng hệ thống hoặc phần mềm để đạt được các mục tiêu nhất định.

3.49

Xác nhận tính hợp lệ (validation)

Hoạt động để chứng minh rằng một mô-đun AI đáp ứng mục đích sử dụng của nó khi được đặt vào mục tiêu của nó.

CHÚ THÍCH 1: Các kỹ thuật mô phỏng hệ thống và kiểm thử hộp đen cung cấp các phương tiện quan trọng để xác nhận tính hợp lệ.

3.50

Xác minh (verification)

Hoạt động để đảm bảo rằng một mô-đun AI được xây dựng chính xác theo thông số kỹ thuật của nó.

CHÚ THÍCH 1: Trong các giai đoạn phát triển, triển khai và vận hành của vòng đời mô-đun AI, các yêu cầu sẽ được đặt vào thiết kế và sau đó được triển khai thành quy định. Việc xác minh đảm bảo rằng kết quả quy định là đúng đối với thông số kỹ thuật. Xem xét thiết kế, xem xét các kỹ thuật quy định, kiểm tra, kiểm tra mô hình và kiểm tra hộp trắng cung cấp các phương tiện quan trọng để xác minh hệ thống.

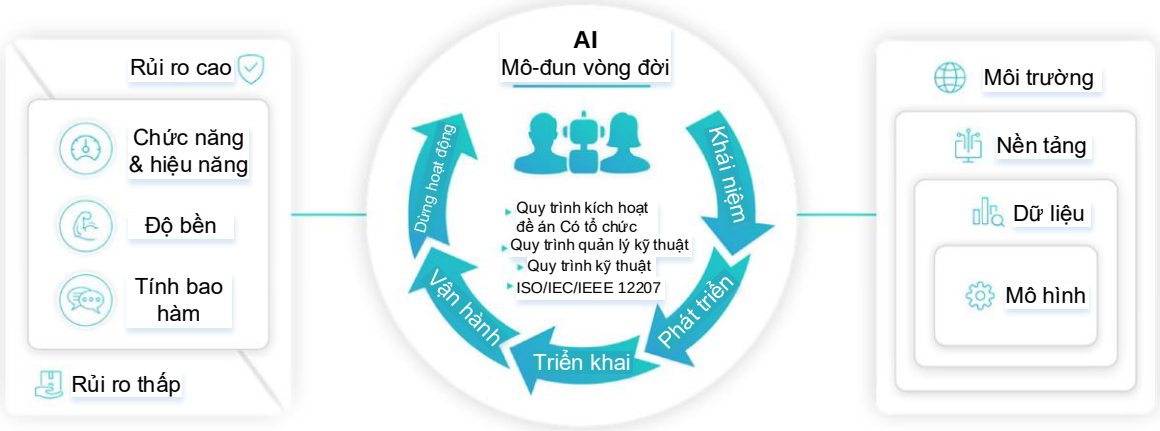
4 Ký tự và các từ viết tắt

Từ viết tắt	Tiếng Anh	Tiếng Việt
AI	Artificial Intelligence	Trí tuệ nhân tạo
ML	Machine Learning	Học máy

5 Siêu mô hình chất lượng

5.1 Giới thiệu

Đảm bảo chất lượng cao của các mô-đun AI nhất định là một nhiệm vụ khó khăn, đặc biệt là trong học máy (ML), do phản ứng không thể đoán trước của chúng đối với các đầu vào không biết trước và sự thiếu minh bạch của nó, điều này đặt ra những thách thức mới để bảo đảm chất lượng AI. Giải quyết những thách thức này theo cách có cấu trúc là cơ sở bắt buộc để phát triển và tích hợp thành công các mô-đun AI mạnh mẽ, an toàn và đáng tin cậy. Cung cấp cách sử dụng dễ dàng và dễ hiểu với cấu trúc lĩnh vực rộng lớn của những vấn đề và thách thức về chất lượng AI là đóng góp chính của tiêu chuẩn này. Để đạt được mục tiêu này, tiàêu chuẩn này mô tả một siêu mô hình chất lượng AI như trong Hình 1. Nó cung cấp cấu trúc cơ sở cho các yêu cầu chất lượng AI cụ thể được đưa ra trong DIN SPEC 92001-2.



Hình 1 - Siêu mô hình chất lượng AI

Siêu mô hình bao gồm các khía cạnh quan trọng nhất cần được tính đến để tạo điều kiện thuận lợi cho thiết kế các mô-đun AI chất lượng cao: Vòng đời của một mô-đun AI, được minh họa bằng vòng tròn ở giữa Hình 1, gồm khái niệm các giai đoạn vòng đời, phát triển, triển khai, vận hành, dừng hoạt động và các quy trình vòng đời. Siêu mô hình chất lượng này thừa nhận

ba nhóm chính của quy trình vòng đời, tức là quy trình kích hoạt đề án của tổ chức, quy trình quản lý kỹ thuật và quy trình kỹ thuật. Các quy trình được chỉ định cho ba nhóm này có thể được tìm thấy trong [3]. Yêu cầu chất lượng đối với các mô-đun AI cần được liên kết với vòng đời. Các yêu cầu nhất định, chẳng hạn như liên quan đến ngăn chặn sai lệch có hại, được yêu cầu cho phát triển cũng như trong triển khai và vận hành vòng đời của thành phần AI.

Các đặc điểm chất lượng chính, gọi là trụ cột chất lượng, cần được tính đến trong toàn bộ vòng đời của mô-đun AI, là chức năng & hiệu năng, độ bền vững và tính dễ hiểu. Chúng được mô tả ở phần bên trái của Hình 1. Những trụ cột chất lượng này kéo theo những thách thức chất lượng dành riêng cho AI cấp bách nhất. Chúng chỉ ra các vấn đề triển khai trung tâm của các mô-đun AI so với phần mềm cổ điển, chẳng hạn như xây dựng tập huấn luyện và độ bền vững đối với các ví dụ đối nghịch. Ba trụ cột chất lượng này không hoàn toàn tách rời nhau. Chẳng hạn, độ bền vững có thể được coi là một phần của chức năng & hiệu năng, vì khả năng thích ứng với các môi trường không biết có thể là một yêu cầu chức năng trong một ứng dụng nhất định. Tuy nhiên, sự tách biệt được đề xuất giúp nhấn mạnh các khía cạnh quan trọng khác nhau của việc đánh giá và đảm bảo chất lượng AI.

Tiêu chuẩn này phân biệt giữa các mô-đun AI có tính an toàn, bảo mật, quyền riêng tư, hoặc liên quan đến đạo đức và những mô-đun không có. Theo cách này, các mô-đun AI được phân chia thành hai loại rủi ro. Sau đây, các mô-đun AI có liên quan đến an toàn, bảo mật, quyền riêng tư, hoặc đạo đức được tóm tắt trong các thành phần có (có tiềm năng) rủi ro cao và sau đó trong các thành phần có rủi ro thấp. Đối với các mô-đun AI có rủi ro cao, sai lệch so với các yêu cầu chất lượng là không được phép hoặc phải được căn chỉnh, trong khi đối với các mô-đun AI có rủi ro thấp thì điều này ít nghiêm ngặt hơn. Phân loại rủi ro này được thể hiện trong Bảng 1.

Trong Hình 1, đánh giá rủi ro cơ bản này được quan sát dưới dạng một hình chữ nhật được đặt phía sau ba trụ cột chất lượng. Bước đầu tiên trong giai đoạn khái niệm của mô-đun AI là đánh giá xem mô-đun AI được hỏi có tiềm ẩn rủi ro cao hay thấp liên quan đến các khía cạnh an toàn, bảo mật, quyền riêng tư hoặc đạo đức hay không. Điều này không phải là một phần của tiêu chuẩn này. Nếu một quy trình đánh giá rủi ro đã được đặt trong một công ty, thì việc phân loại rủi ro cũng có thể được ánh xạ tương ứng với những gì liên quan tới rủi ro cao hoặc thấp, tương ứng, trong tiêu chuẩn này. Một nhóm các chuyên gia trong công ty nên được triệu tập cho nhiệm vụ này.

Bảng 1 - Phân loại các yêu cầu liên quan đến các mô-đun AI có rủi ro cao hoặc thấp theo bắt buộc, khuyến nghị cao và khuyến nghị

Lớp Mô-đun AI **Lớp yêu cầu**	**Rủi ro cao**	**Rủi ro thấp**
Bắt buộc	Không sai lệch so với yêu cầu cho phép	Không sai lệch so với yêu cầu cho phép
Khuyễn nghị mức cao	Sai lệch so với yêu cầu chỉ căn chỉnh	Sai lệch so với yêu cầu chỉ căn chỉnh
Khuyến nghị	Sai lệch so với yêu cầu chỉ căn chỉnh	Sai lệch so với yêu cầu chỉ căn chỉnh

Trong DIN SPEC 92001-2, mỗi yêu cầu chất lượng đã liệt kê được liên kết với một trong ba trụ cột chất lượng, với một hoặc nhiều giai đoạn vòng đời, và với một hoặc nhiều quy trình vòng đời. Ngoài ra, nó cũng nêu rõ yêu cầu chất lượng đề cập đến môi trường, nền tảng, dữ

liệu hoặc mô hình của mô-đun AI. Cách tiếp cận theo cấu trúc này được thúc đẩy bởi thực tế là trong nhiều trường hợp, chất lượng AI không thể được đánh giá bằng cách tiến hành đánh giá mã hóa đơn giản. Chất lượng AI phụ thuộc vào thiết kế của mô hình AI được sử dụng, bao gồm kiến trúc, siêu tham số và thuật toán huấn luyện trong trường hợp ML, nhưng cũng phụ thuộc vào chất lượng của dữ liệu được sử dụng. Tuy nhiên, trong nhiều trường hợp, chất lượng dữ liệu không thể đánh giá được nếu không đưa vào tính toán nền tảng và môi trường của mô-đun AI. Ví dụ, trong lĩnh vực lái xe tự động, cần lưu ý đến tiếng ồn của camera và dịch chuyển trong phân phối dữ liệu tự nhiên, chẳng hạn như các điều kiện thời tiết khác nhau cần phải được thể hiện trong tập dữ liệu. Những sự phụ thuộc này giữa mô hình, dữ liệu, nền tảng và môi trường được minh họa ở phía bên phải của siêu mô hình chất lượng AI trong Hình 1.

Tóm lại, trong tiêu chuẩn này, mỗi mô-đun AI được xem như hoặc là có rủi ro cao hoặc là có rủi ro thấp hoặc được giả định rằng ánh xạ các loại rủi ro nội bộ thành rủi ro cao và rủi ro thấp tương ứng được thực hiện. Để đảm bảo an toàn, bảo mật, quyền riêng tư hoặc đạo đức có liên quan đến các mô-đun AI, tiêu chuẩn này yêu cầu xem xét tất cả các yêu cầu chất lượng đã liệt kê. Những sai lệch tiềm ẩn của các mô-đun AI như vậy cần được căn chỉnh sâu. Mọi yêu cầu chất lượng, được đưa ra trong DIN SPEC 92001-2, được liên kết với một trong ba trụ cột chất lượng, một hoặc nhiều giai đoạn và quy trình vòng đời cũng như loại mô hình, dữ liệu, nền tảng và môi trường. Sau đây, các phần khác nhau của siêu mô hình chất lượng AI sẽ được thảo luận chi tiết hơn.

5.2 Mối quan hệ mô-đun AI và hệ thống phần mềm

Các hệ thống phần mềm bao gồm các phần tử hệ thống tương tác với nhau, trong đó mỗi phần tử có mục đích và yêu cầu riêng tương ứng. Mô-đun AI là một trong những phần tử bao gồm các phương pháp và thuật toán AI tương ứng. Như một phần tử của hệ thống phần mềm, nên nó liên quan và tương tác với các phần tử khác chẳng hạn như phần cứng, phần mềm hoặc dữ liệu và với môi trường xung quanh như con người. Do đó, tiêu chuẩn này tập trung vào đảm bảo chất lượng của các đồ tạo tác AI trong hệ thống phần mềm. Những đồ tạo tác này có thể là hệ thống lai ghép. Cần lưu ý rằng các tiêu chuẩn, yêu cầu và quy định khác có thể áp dụng cho toàn bộ hệ thống phần mềm và do đó áp dụng cho mô-đun AI. Để cung cấp một khuôn khổ cho DevOps của các mô-đun AI đáng tin cậy, một siêu mô hình chất lượng đã được đề xuất và mô tả trong tiêu chuẩn này.

5.3 Đánh giá rủi ro

Để kiểm soát các rủi ro như vi phạm các mục tiêu an toàn, bảo mật, quyền riêng tư hoặc đạo đức, bắt buộc phải nhận dạng và đánh giá rủi ro. Điều này nên được xem xét từ rất sớm trong vòng đời, vì những thay đổi sau này thường dẫn đến nỗ lực cao. Nên thiết lập quản lý rủi ro trong toàn bộ vòng đời.

Đánh giá rủi ro liên quan đến các khía cạnh an toàn, bảo mật, quyền riêng tư, đạo đức và pháp lý của một mô-đun AI. An toàn đề cập đến kỳ vọng rằng một hệ thống, trong những trường hợp đã định, không dẫn đến tình trạng nguy hiểm đến đời sống, sức khỏe, tài sản hoặc môi trường của con người. Bảo mật đề cập đến tác động tiêu cực mà con người hoặc máy móc khác có thể có trên mô-đun AI. Tính bí mật, tính toàn vẹn và tính khả dụng là các chủ đề bảo mật chính. Quyền riêng tư đề cập đến việc thu thập và xử lý dữ liệu cá nhân phù hợp với các quy định có liên quan. Chẳng hạn, chủ thể dữ liệu ở Châu Âu có quyền xóa dữ liệu riêng tư của họ [5].

Một khía cạnh bổ sung của đánh giá rủi ro là mức độ phù hợp về đạo đức của miền mà mô-đun AI được triển khai. Do đó, cần thiết lập một khuôn khổ đạo đức cho vòng đời của mô-đun AI, trong đó hành vi đạo đức phải được xác định trong một quy tắc ứng xử minh bạch. Cảnh giác liên tục và nhận thức liên quan đến các câu hỏi đạo đức sẽ chiếm ưu thế [6] thông qua

toàn bộ sự phát triển của mô-đun AI. Các khía cạnh đạo đức bao gồm xem xét sự thiên vị có chủ ý và không chủ ý của mô-đun AI. Ví dụ, để đảm bảo tính công bằng, có thể cần đưa sự thiên vị có chủ ý vào mô-đun AI. Do đó, để đảm bảo hành vi đạo đức của mô-đun AI, không chỉ cần xem xét các khía cạnh phân tích công nghệ AI mà còn phải thu thập kiến thức về môi trường xã hội mà mô-đun AI hoạt động.

Trong tiêu chuẩn này, chúng tôi phân chia các mô-đun AI thành hai loại, đó là các mô-đun AI rủi ro cao và rủi ro thấp. Các lớp rủi ro nội bộ được giả định là được ánh xạ tới hoặc là rủi ro cao hoặc là rủi ro thấp. Các mô-đun AI rủi ro cao (được gọi là các mô-đun AI "quan trọng") có liên quan đến tính an toàn, bảo mật, quyền riêng tư hoặc đạo đức. Các miền có mức độ liên quan như vậy có thể là lái xe tự động, chẩn đoán y tế và xếp hạng tín dụng. Các mô-đun AI rủi ro thấp không đáp ứng bất kỳ mục tiêu nào về an toàn, bảo mật, quyền riêng tư hoặc đạo đức, mặc dù đã được đánh giá rủi ro theo công nghệ tiên tiến nhất. Chúng được gọi là các mô-đun AI "thoải mái". Tuy nhiên, hệ thống phần mềm chứa mô-đun AI có thể có sự liên quan về an toàn, bảo mật, quyền riêng tư hoặc đạo đức. Hầu hết các hệ thống khuyến nghị hoặc quảng cáo, chẳng hạn như lựa chọn phim tự động, là ví dụ cho các mô-đun AI rủi ro thấp.

Trong DIN SPEC 92001-2, các yêu cầu được phân loại dựa trên mức độ liên quan của chúng chuyển sang bắt buộc, khuyến nghị cao và khuyến nghị, xem Bảng 1.

Đối với các mô-đun AI rủi ro thấp, cho phép sai lệch so với các yêu cầu khuyến nghị mà không cần căn chỉnh thêm. Chỉ cho phép sai lệch so với các yêu cầu được khuyến nghị cao đối với các mô-đun AI rủi ro thấp trong các trường hợp ngoại lệ và có căn chỉnh tương xứng, trong khi đó, sai lệch so với các yêu cầu bắt buộc như thiết lập quy trình xác định và đánh giá rủi ro sẽ không được chấp nhận. Chỉ cho phép sai lệch so với các yêu cầu được khuyến nghị và khuyến nghị cao trong những trường hợp ngoại lệ và có căn chỉnh tương xứng, trong khi không được phép sai lệch so với các yêu cầu bắt buộc.

Do sự hiểu biết sâu về miền cụ thể của hệ thống là cần thiết để nhận dạng các mục tiêu liên quan đến các khía cạnh an toàn, bảo mật, quyền riêng tư và đạo đức, nên tiêu chuẩn này không chứa bất kỳ thủ tục tương ứng nào để đánh giá mức độ nghiêm trọng. Phân loại này phải đảm bảo phù hợp với tính tiên tiến của các tiêu chuẩn trong lĩnh vực cụ thể. Các chủ đề mẫu mực, có thể được xem xét ở đây, sẽ đánh giá sự kiện xảy ra, tính nghiêm ngặt và khả năng kiểm soát hậu quả của sự cố mô-đun AI có thể xảy ra. Đánh giá rủi ro của các mô-đun AI cũng bị ảnh hưởng bởi mức độ tác động của nó đối với môi trường. Có thể đánh giá hoặc mô-đun AI có tác động trực tiếp lên môi trường hoặc nó chỉ đưa ra các khuyến nghị, chẳng hạn như khuyến nghị chẩn đoán y tế với sự giải thích cụ thể. Khả năng kiểm soát chỉ có thể dễ dàng đảm bảo trong trường hợp sau đây. Điều này dẫn đến phải tính đến yếu tố khác: sự hiện diện hay vắng mặt của giám sát bên ngoài và thể loại của nó. Đó là, sự tồn tại của hệ thống tác động khác như phải xem xét đối với người giám sát. Điều này được liên kết chặt chẽ hơn nữa với mức độ tự chủ của mô-đun AI. Mức độ tự chủ cao có thể làm tăng phạm vi tác động có thể xảy ra do hành vi ngoài ý muốn. Một chủ đề quan trọng khác nữa liên quan đến loại hình học: Các thuật toán học ngoại tuyến không còn thay đổi dài hơn trong hoạt động, sẽ tạo điều kiện thuận lợi cho việc đánh giá. Ngược lại, các thuật toán học trực tuyến tiếp tục thay đổi trong hoạt động, điều này có thể khiến việc đánh giá trở nên khó khăn hơn nhiều hoặc thậm chí là không thể, tùy thuộc vào các phương pháp tiếp cận thuật toán [7].

5.4 Môi trường, nền tảng, dữ liệu, mô hình

Các ảnh hưởng đối với mô-đun AI, chẳng hạn như thiết lập phần cứng chung hoặc mô hình toán học đã chọn, cần được xem xét trong quá trình đánh giá chất lượng. Để giải quyết các yếu tố ảnh hưởng của mô-đun AI, tiêu chuẩn này đề xuất một phương pháp tiếp cận phân cấp. Như được minh họa ở phía bên phải của Hình 1, nó phân biệt giữa môi trường, nền tảng,

dữ liệu và mô hình. Những khía cạnh này tham chiếu đến các phân lớp. Điều này cho phép thiết kế các quy trình đảm bảo chất lượng có cấu trúc bằng cách ánh xạ các yêu cầu chất lượng tới các lớp khác nhau.

Vì môi trường, nền tảng, dữ liệu và mô hình có thể tác động đến tất cả các giai đoạn vòng đời, cho nên chúng phải được xem xét trong toàn bộ vòng đời của một mô-đun AI. Đoạn văn bản sau đây trình bày các định nghĩa của bốn lớp.

Môi trường là tình huống mà mô-đun AI được cài đặt tổng thể và nó có thể tương tác với mô-đun đó. Liên quan đến cài đặt mô-đun phần mềm trong hệ thống phần mềm [3], môi trường bao gồm các khía cạnh bên ngoài hệ thống phần mềm có ảnh hưởng đến mô-đun. Ví dụ về môi trường AI rất đa dạng. Môi trường của các nhiệm vụ phân tích hình ảnh có thể là nguồn từ những hình ảnh được cảm nhận bởi hệ thống phần mềm và mô-đun AI, chẳng hạn như máy ảnh. Trong các trò chơi như cờ vua, cờ vây hoặc bài xì phé, luật chơi và biểu diễn trạng thái trò chơi được coi như môi trường. Ô tô tự lái được đặt trong môi trường thế giới thực có chứa các chướng ngại vật có thể xảy ra trên đường và các ô tô khác. Lưu ý rằng nhìn chung, đánh giá môi trường không phải là một quá trình cố định. Đó là theo quyết định của một nhóm chuyên gia. Xem Phụ lục A đối với ví dụ chi tiết về các khía cạnh môi trường khác nhau.

Nền tảng liên quan đến các thuộc tính và ràng buộc của phần cứng và hệ điều hành. Điều này bao gồm, ví dụ như giao diện (giao tiếp), sức mạnh xử lý, sự phụ thuộc vào các yếu tố phần mềm khác, hệ thống thời gian chạy, mối liên quan độ tin cậy và độ chính xác của các thành phần phần cứng. Ví dụ như độ chính xác và độ tin cậy của máy ảnh hoặc cảm biến.

Dữ liệu được sử dụng bởi một mô-đun AI đóng một vai trò quan trọng đối với hiệu năng và chất lượng của nó. Do số lượng và việc sử dụng các nguồn dữ liệu lớn ngày càng gia tăng, điều đặc biệt quan trọng là phải giải quyết các vấn đề về tính đại diện, tính sẵn có và chất lượng của dữ liệu. Trong ngữ cảnh của ML, những khía cạnh này có thể có ảnh hưởng đáng kể để đạt được các kết quả dự kiến. Do đó, dữ liệu là một trong những chủ đề chính khi lập luận về các yêu cầu của các mô-đun AI. Trong trường hợp các mô-đun AI có mức độ học dựa trên dữ liệu thấp, dữ liệu có thể được coi là trừu tượng hơn và ít cấu trúc hơn so với ML. Tuy nhiên, các mô-đun AI này vẫn sử dụng dữ liệu để thực hiện các thuật toán của chúng.

Thuật ngữ mô hình trong ngữ cảnh của tiêu chuẩn này đề cập đến hai thực thể riêng biệt: mô hình không gian và mô hình suy luận. Mô hình không gian bao gồm tất cả tập hợp các cách tiếp cận tiềm năng để giải quyết nhiệm vụ vấn đề hiện tại. Các thuật toán, mô hình toán học, kiến trúc và cấu hình tham số có thể dẫn đến các giải pháp phù hợp cho nhiệm vụ quy định được bao gồm trong tập hợp này. Mô hình suy luận là một phần tử cụ thể của mô hình không gian. Do đó, nó bao gồm một kiến trúc mô hình cụ thể có cấu hình tham số cố định. Cấu hình này được cấp phát từ mô hình không gian thông qua một phương pháp lựa chọn, chẳng hạn như thuật toán huấn luyện trên một số tập dữ liệu. Mô hình suy luận có thể được sử dụng để giải quyết nhiệm vụ dự định ở một mức độ nhất định. Trong bối cảnh phân loại hình ảnh với mạng thần kinh tích chập, mô hình không gian cấu thành tất cả các kiến trúc đã chọn trước đó có thể khác nhau về độ sâu, cấu trúc và bao gồm tất cả các tham số có thể, chẳng hạn như trọng số và độ lệch. Ngoài ra, các thuật toán dẫn đến một mô hình suy luận sau này là một phần của mô hình không gian. Chẳng hạn, các kỹ thuật tối ưu hóa hoặc phương pháp tiến hóa bao gồm siêu tham số, chẳng hạn như sử dụng tỷ lệ học hoặc chiến lược chính quy hóa. Mô hình không gian có thể được tăng cường khi xem xét các cách tiếp cận mới đối với nhiệm vụ, ví dụ như bằng các cây quyết định thích ứng.

Sự sắp xếp đồ họa của các lớp trong Hình 1 làm nổi bật tác động mà một lớp có thể có trên các lớp tiếp theo. Một thay đổi đáng kể trong môi trường có thể hàm ý một thiết lập phần cứng khác và thay đổi dữ liệu có sẵn. Hơn nữa, tăng hoặc giảm dữ liệu có sẵn có thể dẫn đến các

lựa chọn mô hình khác nhau. Các tác động có thể bao gồm từ những thích ứng nhỏ của các tham số đến xem xét lại mô hình toán học đã lựa chọn. Do đó, chúng có tác động đáng kể đến đảm bảo chất lượng. Những thay đổi về môi trường, nền tảng, dữ liệu và mô hình có thể diễn ra bằng cách lặp lại các giai đoạn vòng đời. Do đó, chuyển dịch giữa các giai đoạn cần được thực hiện với sự thận trọng bổ sung liên quan đến các yêu cầu dựa trên các lớp. Với những ảnh hưởng liên quan đến nhau đã định, đòi hỏi phải liên tục đánh giá lại các yêu cầu liên quan đến chất lượng của mô-đun AI [7].

Các yêu cầu cụ thể về chức năng & hiệu năng, độ bền vững và tính dễ hiểu, tác động trên các lớp và đưa ra chi tiết hơn về cách đạt được các tiêu chuẩn chất lượng dự kiến. Chúng được đưa ra trong DIN SPEC 92001-2.

5.5 Vòng đời

5.5.1 Khái quát

Mọi hệ thống phần mềm đều có vòng đời xác định sự tiến hóa của nó từ khi hình thành cho đến khi ngừng hoạt động và bao gồm các giai đoạn và quy trình của vòng đời [3]. Các giai đoạn vòng đời thường bao gồm các khía cạnh khái niệm hóa, hiện thực hóa, sử dụng, tiến hóa và loại bỏ. Hệ thống phần mềm tiến triển qua các giai đoạn bằng cách thực hiện các quy trình vòng đời bao gồm lập kế hoạch, thực hiện và đánh giá. Chúng được liên kết với một hoặc nhiều giai đoạn, tùy thuộc vào quy trình cụ thể. Là một phần của hệ thống phần mềm, mô-đun AI cho phép tạo ra một vòng đời tương tự với các giai đoạn khái niệm, phát triển, triển khai, vận hành và ngừng hoạt động, cũng như các nhóm quy trình, quy trình hỗ trợ đề án tổ chức, quy trình quản lý kỹ thuật và quy trình kỹ thuật. Tiêu chuẩn này dựa trên [3]. Nó không xác định hoặc ra lệnh cho một mô hình vòng đời mới. Nếu tồn tại một vòng đời cho toàn bộ hệ thống phần mềm, thì có thể chấp nhận sử dụng vòng đời cụ thể này cho mô-đun AI. Có thể ánh xạ các giai đoạn vòng đời của mô-đun AI với các giai đoạn vòng đời của vòng đời phần mềm hiện có. Tuy nhiên, khuyến nghị sử dụng một số loại vòng đời để phát triển một mô-đun AI. Nó cần được ánh xạ tới vòng đời được trình bày trong tiêu chuẩn này.

Các giai đoạn vòng đời được đề cập ở trên là một hướng dẫn. Mỗi bên liên quan của tiêu chuẩn này được yêu cầu đánh giá xem các giai đoạn vòng đời đó có thể áp dụng được hay không, và điều chỉnh chúng nếu cần. Điều này cho phép liên kết các yêu cầu với một hoặc nhiều quy trình vòng đời và một hoặc nhiều giai đoạn vòng đời. Cần thiết phải thực hiện các hoạt động đảm bảo chất lượng phù hợp trước khi dịch chuyển diễn ra từ một giai đoạn sang giai đoạn tiếp theo. Các hoạt động này nên bao gồm xác minh và xác nhận tính hợp lệ sử dụng các phương pháp thử nghiệm thích hợp. Đặc biệt đối với các mô-đun AI có rủi ro cao, các yêu cầu chất lượng về chức năng & hiệu năng, độ bền vững và tính dễ hiểu phải được liên kết với vòng đời để ngăn chặn hành vi bất ngờ và có khả năng gây hại.

Mô tả tiến trình phát triển của một mô-đun AI từ khái niệm đến khi ngừng hoạt động, các giai đoạn trong vòng đời là sự hướng dẫn cho sự phát triển của nó. Mỗi giai đoạn có mục đích và đóng góp riêng cho vòng đời. Sử dụng các giai đoạn trong vòng đời cung cấp một khuôn khổ cho khả năng quan sát và kiểm soát cao đối với đề án và các quy trình [3]. Trong mỗi giai đoạn vòng đời, các phương pháp thử nghiệm thích hợp sẽ được xem xét và các tiêu chí thử nghiệm sẽ được thu thập cho giai đoạn thực tế và vòng đời tiếp theo. Tiêu chuẩn này đề xuất năm giai đoạn vòng đời sau đây cho các mô-đun AI được mô tả dưới dạng mũi tên màu xanh lam trong Hình 1.

a) Giai đoạn đầu tiên của vòng đời là giai đoạn khái niệm gồm tất cả các hành động diễn ra trước khi phát triển mô-đun AI, tức là quá trình xác định vấn đề, phân tích và tìm kiếm mô hình không gian phù hợp. Dựa trên vấn đề cụ thể, các mô hình phù hợp nên được xác định và phân tích liên quan đến các thuộc tính như hội tụ và các giả định đầu vào. Trong giai đoạn

này, không có siêu tham số mô hình nào được chọn và không thực hiện đánh giá mô hình cuối cùng. Đây là một phần của giai đoạn phát triển. Ngoài ra, các tiêu chí có khả năng chấp nhận nên được xác định cho các bước đảm bảo chất lượng hơn nữa. Ví dụ, khuyến nghị vận hành hóa vấn đề sao cho công thức của nó chứa các hành động khả thi đối với giải pháp.

b) Giai đoạn phát triển gồm một số hoạt động, bao gồm thiết kế và đặc tả hệ thống, tạo nguyên mẫu và triển khai, tích hợp, theo dõi lỗi và sửa lỗi, xác minh và xác nhận tính hợp lệ bao gồm thử nghiệm ở các cấp độ khác nhau (chức năng, tích hợp, thử nghiệm, hiệu năng & độ bền vững), đóng gói, tài liệu hóa, phiên bản, v.v… Trong ngữ cảnh của AI, các phương pháp tiếp cận phát triển theo hướng dữ liệu được sử dụng để xây dựng một mô hình giao thoa liên quan đến các phương pháp tiếp cận kỹ thuật phần mềm cổ điển: Các hoạt động như vậy hàm chứa thu thập dữ liệu, phân tích dữ liệu và lập trình thực tế hoặc các nỗ lực huấn luyện. Chúng được đặt trong môi trường thực nghiệm chứ chưa phải trong môi trường mà mô-đun AI sẽ được sử dụng sau này. Trong trường hợp các mô hình ML, tập dữ liệu phải được phân tích, thấu hiểu và các biến có liên quan đến mục tiêu hoặc vấn đề phải được xác định. Đối với các hệ chuyên gia, tri thức chuyên gia nên được chính thức hóa dưới dạng các quy tắc và hệ quy tắc phức hợp. Trong giai đoạn này, siêu tham số mô hình được so sánh liên quan đến chất lượng của mô hình cụ thể. Các biện pháp và số liệu khác nhau để đánh giá chất lượng mô hình có thể được xem xét. Mục đích là để tìm một mô hình với các siêu tham số cụ thể giải quyết thỏa đáng vấn đề. Ví dụ, một mạng thần kinh tích chập có thể thực hiện đầy đủ việc phát hiện đối tượng và phân đoạn ngữ nghĩa. Nó có thể được huấn luyện trên một tập dữ liệu gồm các hình ảnh được dán nhãn trước. Các kiến trúc và tham số khác nhau có thể được xem xét và kiểm tra chất lượng tổng thể của chúng. Biểu diễn của tập dữ liệu có thể được điều chỉnh cho phù hợp với mô hình đã chọn vì một số mô hình ML cần một hình dạng đầu vào cụ thể. Chẳng hạn, trong các nhiệm vụ phát hiện đối tượng, kích thước đầu vào được điều chỉnh phù hợp với kiến trúc mạng cụ thể. Vào cuối giai đoạn này, mô hình suy luận được sử dụng để giải quyết vấn đề được xác định và xác nhận theo các tiêu chí có thể chấp nhận đã xác định. Ngoài ra, tập dữ liệu được khởi lập. Có thể cần quay lại giai đoạn khái niệm nếu mô hình không gian phải được điều chỉnh cho phù hợp với tập dữ liệu.

c) Giai đoạn triển khai thể hiện sự chuyển dịch từ phát triển sang vận hành. Đối với các mô-đun AI có mức độ học dựa trên dữ liệu cao, việc triển khai bao gồm huấn luyện mô hình trên hệ thống máy chủ và xuất sang hệ thống đích. Đối với các mô-đun AI có mức độ học dựa trên dữ liệu thấp, dịch chuyển từ máy chủ sang hệ thống mục tiêu cũng có liên quan. Chẳng hạn, khả năng chấp nhận mô-đun AI bởi bên liên quan là một phần của hệ thống mục tiêu và phải đạt được. Lưu ý rằng triển khai bắt đầu giai đoạn hoạt động. Vì vậy, không thể phân định rạch ròi giữa triển khai và vận hành. Triển khai và phát triển cả hai đều là một phần của sự phát triển. Đối với nhiều vấn đề phát triển liên quan đến AI, sẽ rất hữu ích khi nhấn mạnh sự khác biệt giữa cài đặt huấn luyện là một phần của phát triển và môi trường thực tế nơi mô-đun AI đang được sử dụng trong khi triển khai và vận hành. Đặc biệt, việc xác minh độ bền vững của các mô-đun AI có rủi ro cao đặt ra một thách thức lớn trong giai đoạn này và giai đoạn tiếp theo. Điều này là do việc chuyển dịch từ môi trường nhân tạo sang môi trường thực tế, nơi các khía cạnh an toàn, bảo mật, quyền riêng tư, hoặc đạo đức đều quan trọng. Do đó, trong giai đoạn này, việc lùi lại giai đoạn phát triển là phổ biến do khó chuyển giao mô hình được huấn luyện sang thế giới thực.

d) Giai đoạn vận hành đề cập đến các khía cạnh bảo trì và đánh giá trong môi trường nơi sử dụng mô-đun AI. Trong giai đoạn này, sự khác biệt lớn với phát triển phần mềm cổ điển được quan sát, vì các thuật toán ML có thể tiếp tục học từ dữ liệu thông qua học trực tuyến và do đó tiếp tục thay đổi sau khi huấn luyện trong môi trường thử nghiệm.

e) Giai đoạn ngừng hoạt động quy định việc giải thể và ngừng cung cấp mô-đun AI cũng như

chuyển dịch sang một mô-đun AI mới. Trong giai đoạn này, một mô-đun AI có thể bị xóa khỏi hệ thống phần mềm hoặc thay đổi đáng kể để một mô-đun AI mới được tạo ra. Điều này bắt đầu một vòng đời mới. Do đó, điều này cũng có thể được hiểu là mô-đun AI ban đầu đã ngừng hoạt động.

Năm giai đoạn vòng đời này minh họa các giai đoạn mà một mô-đun AI trải qua trong suốt thời gian sống của nó. Chúng không tạo nên một quá trình tuyến tính từ ý tưởng đến ngừng hoạt động. Ngược lại, có thể cần quay lại giai đoạn trước của vòng đời nhiều lần để hoàn thành đầy đủ các yêu cầu ở đó liên kết đến giai đoạn cụ thể này. Chẳng hạn, nếu một lỗi được tìm thấy ở đâu đó trong vòng đời, thì nó có thể cần phải quay lại giai đoạn nơi nó bắt nguồn để sửa lỗi và sau đó chuyển qua các giai đoạn vòng đời một lần nữa. Như đã đề cập ở trên, các giai đoạn này có thể được ánh xạ tới vòng đời hiện có của hệ thống phần mềm mà mô-đun AI là một phần trong đó. Cần đảm bảo chất lượng thông qua đảm bảo quy trình. Do đó, trước khi tiến hành giai đoạn tiếp theo, cần phải kiểm tra xem tất cả các giai đoạn yêu cầu đã được đáp ứng để thỏa mãn hoàn toàn hay chưa.

5.5.2 Quy trình vòng đời

Để liên kết các trụ cột chất lượng với vòng đời, các yêu cầu được chính thức hóa cho từng quy trình vòng đời, dựa trên môi trường, nền tảng, dữ liệu và mô hình. Các quy trình vòng đời đại diện cho tập hợp các hoạt động có liên quan hoặc tương tác lẫn nhau [3] được gán cho một hoặc nhiều giai đoạn vòng đời. Các nhóm quy trình trong tiêu chuẩn này đề cập đến những nhóm được xác định trong [3] và được minh họa trong Hình 1. Các quy trình được xác định theo tiêu đề, mục đích và kết quả. Các quy trình được tổ chức thành ba nhóm:

a) Các quá trình cho phép đề án của tổ chức “liên quan đến việc cung cấp các nguồn lực để cho phép dự án đáp ứng nhu cầu và mong đợi của các bên liên quan của tổ chức” [3]. Hầu hết các quy trình trong nhóm này chỉ bị ảnh hưởng một chút bởi những thách thức mới được đưa ra bởi AI. Tuy nhiên, người dùng tiêu chuẩn này cần đánh giá xem có cần thay đổi các quy trình hiện có hay không. Chẳng hạn, những cách mà các quy trình này cần được tinh chỉnh bao gồm thiết lập các tiêu chí đánh giá chất lượng áp dụng cho chức năng & hiệu năng, độ bền vững và tính dễ hiểu của các mô-đun AI.

b) Các quy trình quản lý kỹ thuật “liên quan đến quản lý các nguồn lực và tài sản được phân bổ bởi tổ chức quản lý và có áp dụng chúng để thực hiện các thỏa thuận mà tổ chức hoặc các tổ chức tham gia [...]. Đặc biệt, chúng liên quan đến lập kế hoạch về mặt chi phí, thang thời gian và đạt được, đến kiểm tra các hành động để giúp đảm bảo rằng chúng tuân thủ các kế hoạch và tiêu chí hiệu năng, cũng như việc xác định và lựa chọn các hành động khắc phục [...]” [3]. Ngoài ra, các biện pháp cụ thể với các tiêu chí chất lượng tương ứng cần được xác định để cho phép đánh giá nếu mô-đun AI đáp ứng các tiêu chí chức năng & hiệu năng, độ bền vững và tính dễ hiểu.

c) Các quy trình kỹ thuật “biến đổi nhu cầu của các bên liên quan thành sản phẩm hoặc dịch vụ bằng các hành động kỹ thuật trong suốt vòng đời” [3]. Chúng đảm bảo rằng hiệu năng bền vững và chất lượng tổng thể đạt được khi áp dụng mô-đun AI. Đây là nhóm các quy trình bị ảnh hưởng nhiều nhất bởi các thách thức dành riêng cho AI. Chẳng hạn, một khía cạnh quan trọng cần được xem xét trong quá trình phân tích hệ thống là để đảm bảo mức độ diễn giải cần thiết của mô-đun AI. Đối với các mô-đun ML, cần đặc biệt chú ý đến việc phân tích và lập mô hình dữ liệu phù hợp.

Các quy trình vòng đời của ba nhóm quy trình này được liệt kê trong [3]. Chúng cung cấp một khuôn khổ và cấu trúc chung cho vòng đời của một mô-đun AI. Do việc phát triển và sử dụng các mô-đun AI kéo theo những thách thức cụ thể, nên các quy trình đó được trang bị các yêu cầu dành riêng cho AI để đảm bảo rằng những thách thức đó được giải quyết thỏa đáng.

Chúng sẽ được giải quyết rõ ràng với các yêu cầu chất lượng tương ứng trong DIN SPEC 92001-2.

Các quy trình thỏa thuận là một nhóm quy trình khác trong [3] không phải là một phần của tiêu chuẩn này. Các quy trình thỏa thuận "là các quy trình tổ chức áp dụng ngoài phạm vi vòng đời của dự án, cũng như vòng đời của đề án. Các thỏa thuận cho phép […] nhận thức rõ giá trị và hỗ trợ các chiến lược kinh doanh cho các tổ chức [...]." [3]. Trong khi các quy trình thỏa thuận áp dụng cho toàn bộ hệ thống phần mềm, chúng không liên quan đến một thành phần phần mềm và các thách thức dành riêng cho AI. Do đó, DIN SPEC này không bao gồm các quy trình thỏa thuận.

5.6 Trụ cột chất lượng AI

Trong toàn bộ vòng đời, các đặc điểm chất lượng AI dưới dạng các yêu cầu cần được xem xét. Tiêu chuẩn này giới thiệu một cách tiếp cận để bao quát phổ đủ rộng của các khía cạnh chất lượng phần mềm liên quan đến AI và để nhấn mạnh tầm quan trọng của các yêu cầu AI cụ thể. Nó cho phép phát triển và thực thi các mô-đun AI hiệu quả, mạnh mẽ, an toàn và đáng tin cậy. Đối với điều này, ba đặc điểm chất lượng chính được thể hiện là chức năng & hiệunăng, độ bền vững và tính dễ hiểu.

a) Chức năng & hiệu năng thể hiện mức độ mà một mô-đun AI có khả năng hoàn thành nhiệm vụ dự kiến của nó trong các điều kiện đã nêu. Mục tiêu chất lượng giải quyết các vấn đề liên quan đến chính thức hóa vấn đề, phân tích nhiệm vụ, thu thập, phân tích và xử lý dữ liệu. Đánh giá hiệu năng và lựa chọn mô hình là các chủ đề xa hơn được đề cập trong trụ cột chất lượng này. Cần phải xác định chính xác vấn đề hoặc mục tiêu trước khi phát triển và phân tích nó theo các ràng buộc và giả định liên quan đến môi trường, nền tảng, dữ liệu và mô hình. Sau khi phân tích vấn đề, các giải pháp tiềm năng cần được chính thức hóa và đánh giá. Để tìm ra các giải pháp phù hợp, các phép đo và chỉ số hiệu năng thích hợp sẽ được lựa chọn cho nhiệm vụ và dữ liệu đã cho. Các yêu cầu liên quan đến dữ liệu được nêu để cải thiện khái quát hóa và phát triển các phương pháp giải pháp phù hợp. Ngoài ra, các yêu cầu về tính dễ hiểu của các phương pháp và biện pháp đánh giá hiệu năng được cung cấp.

b) Độ bền vững biểu thị khả năng của mô-đun AI đối phó với dữ liệu đầu vào sai, nhiễu, không nhận biết và đối nghịch. Do tính phức tạp của môi trường mô-đun AI có thể xuất phát từ tính không cố định và tính đa chiều cao, độ bền vững là một vấn đề chất lượng AI quan trọng. Do đó, độ bền vững của mô-đun AI cần phải được định lượng đầy đủ và đáp ứng các yêu cầu được xác định trong phân tích rủi ro. Sự phụ thuộc của mô hình vào môi trường, nền tảng và dữ liệu phải được xem xét. Kết quả của sự cố có thể hoặc là gây ra bởi phân loại sai do dịch chuyển phân phối hoặc là một cuộc tấn công đối nghịch cần phải được phân bổ. Sự dịch chuyển phân phối xảy ra khi mô-đun AI tiếp xúc với các điểm dữ liệu bên ngoài tập dữ liệu huấn luyện hoặc kiểm tra. Khả năng xảy ra một cuộc tấn công đối nghịch phải được giải quyết cụ thể, vì điều này gây ra rủi ro lớn đối với hoạt động của các mô-đun AI trong các cài đặt liên quan đến an toàn và bảo mật. Đối với điều này, kiến thức của đối thủ về mô-đun AI và phạm vi nhiễu loạn tương ứng sẽ được đánh giá và các chiến lược phòng thủ được yêu cầu để được lựa chọn phù hợp và được giám sát liên tục trong quá trình phát triển và triển khai.

c) Tính dễ hiểu thể hiện mức độ mà một bên liên quan có nhu cầu đã định có thể hiểu biết nguyên nhân của đầu ra của mô-đun AI. Các nguyên nhân bao gồm lý do cho một đầu ra cụ thể, tức là đầu vào dẫn đến nó, và toàn bộ quá trình ra quyết định. Điều này có nghĩa rằng thành phần AI minh bạch và có thể giải thích được. Hơn nữa, sự hiểu biết định tính giữa các biến đầu vào và phản hồi được cung cấp liên quan đến mức độ chuyên môn và nhu cầu hiểu biết của các bên liên quan. Chẳng hạn, nhà phát triển mô-đun AI cần hiểu biết không chỉ dữ liệu và mô hình suy luận mà còn cả mô hình không gian và khung toán học. Trong khi đó, chủ

thể dữ liệu chỉ cần hoặc muốn hiểu biết cách dữ liệu được lưu trữ và sử dụng thêm. Ngoài ra, các ràng buộc được coi là có thể đặt ra các nhu cầu về tính dễ hiểu khác nhau ở mỗi cấp độ. Pháp luật là một trong những ràng buộc bên ngoài có thể xảy ra. Trụ cột chất lượng này tập trung vào tính minh bạch và khả năng diễn giải của mô hình đã chọn. Cái gọi là mô hình hộp trắng hoàn toàn dễ hiểu và tiết lộ quá trình ra quyết định sẽ được ưu tiên sử dụng. Các mô hình hộp đen hoặc hộp xám có thể không rõ ràng, tức là ánh xạ đầu vào đến đầu ra phần lớn không thể hiểu được đối với các bên liên quan. Nếu không có sẵn các mô hình khác, thì tính dễ hiểu cần phải tăng cường sử dụng các giải thích sau đại học.

Các trụ cột chất lượng này phù hợp tốt với các yêu cầu chất lượng AI cụ thể cần được đáp ứng để giải quyết các thách thức chính đặt ra với AI. Mô tả chi tiết về cả ba trụ cột và các nhiệm vụ liên quan được đưa ra trong các phần khác của DIN SPEC 92001.

Tóm lại, đảm bảo chất lượng được phân chia thành ba phần: Vòng đời, các yếu tố ảnh hưởng và ba trụ cột chất lượng. Việc xem xét các yếu tố ảnh hưởng môi trường, nền tảng, dữ liệu và mô hình là cần thiết. Nó nâng cao nhận thức về các vấn đề chất lượng có thể phát sinh trong các giai đoạn và quy trình vòng đời khác nhau của mô-đun AI. Toàn bộ vòng đời được hướng dẫn bởi ba trụ cột chất lượng cần được giải quyết. Tất cả các yêu cầu đảm bảo chất lượng được tập hợp trong các đặc tính chất lượng này. Do đó, siêu mô hình chất lượng AI bao gồm tất cả các khía cạnh đảm bảo chất lượng AI.

THƯ MỤC TÀI LIỆU THAM KHẢO

[1] ISO/IEC 2382:2015. Information technology — Vocabulary. Tech. rep. ISO and IEC, 2015.

[2] P. Stone et al. Artificial Intelligence and Life in 2030. Tech. rep. Stanford University, 2016.

[3] ISO/IEC/IEEE 12207:2017. Systems and software engineering — Software life cycle processes. Tech. rep. ISO, IEC and IEEE, 2017.

[4] C. M. Bishop. Pattern Recognition and Machine Learning. Springer Science+Business Media, 2006.

[5] Council of European Union. Council regulation (EU) 2016/679 — General Data Protection Regulation (GDPR). Tech. rep. Council of European Union, 2018.

[6] A. Kung, O. Maridat, and G. Pellischeck. Cyber Security Framework. WP 200. Tech. rep. Automat, 2015.

[7] P. V. Wesel and A. E. Goodloe. Challenges in the Verification of Reinforcement Learning Algorithms. Tech. rep. NASA/TM-2017-219628, 2017.

DIN SPEC 92001-2, Artificial Intelligence — Life Cycle Processes and Quality Requirements — Part 2: Technical and Organizational Requirements

ISO 5840-3:2013, Cardiovascular implants — Cardiac valve prostheses — Part 3: Heart valve substitutes implanted by transcatheter techniques

ISO 8930:1987, General principles on reliability for structures — List of equivalent terms

ISO 9000:2015, Quality management systems — Fundamentals and vocabulary

ISO 17364:2013, Supply chain applications of RFID — Returnable transport items (RTIs) and returnable packaging items (RPIs)

ISO 19103:2015, Geographic information — Conceptual schema language

ISO 24534-5:2011, Intelligent transport systems — Automatic vehicle and equipment identification — Electronic Registration Identification (ERI) for vehicles — Part 5: Secure communications using symmetrical techniques

ISO 26262-1:2011, Road vehicles — Functional safety — Part 1: Vocabulary

ISO 28001:2007, Security management systems for the supply chain — Best practices for implementing supply chain security, assessments and plans — Requirements and guidance

ISO/IEC 18023-1:2006, Information technology — SEDRIS — Part 1: Functional specification

ISO/IEC 25000, System und Software-Engineering — Qualitätskriterien und Bewertung von System- und Softwareprodukten (SQuaRE) — Leitfaden für SQuaRE

ISO/IEC 38505-1:2017, Information technology — Governance of IT — Governance of data — Part 1: Application of ISO/IEC 38500 to the governance of data

ISO/TR 12773-2:2009, Business requirements for health summary records — Part 2: Environmental scan

ISO/TS 17574:2017, Electronic fee collection — Guidelines for security protection profiles

ISO/TS 21089:2018, Health informatics — Trusted end-to-end information flows

S. J. Russell and P. Norvig. Artificial intelligence: a modern approach. Prentice Hall series in artificial intelligence. Prentice Hall, 1995.

W. Bourne, R. Gallimard and J. Tunnicliffe. Multi-Agent Systems. url: https://www.doc.ic.ac.uk/project/examples/2005/163/g0516302/environments/environments.html (visited on 11/30/2018).