## Học sâu là gì?

Học sâu là một phân ngành của học máy sử dụng mạng thần kinh nhiều lớp, gọi là mạng thần kinh sâu, để mô phỏng khả năng đưa ra quyết định phức tạp của bộ não người. Một số dạng học sâu đang thúc đẩy hầu hết các ứng dụng trí tuệ nhân tạo (AI) trong cuộc sống của chúng ta ngày nay.

Sự khác biệt chính giữa học sâu và học máy là cấu trúc của kiến trúc mạng thần kinh bên dưới. Các mô hình học máy truyền thống "không sâu" sử dụng mạng thần kinh đơn giản với một hoặc hai lớp tính toán. Mô hình học sâu sử dụng ba hoặc nhiều lớp - nhưng thường là hàng trăm hoặc hàng nghìn lớp - để huấn luyện các mô hình.

Trong khi các mô hình học có giám sát yêu cầu dữ liệu đầu vào có cấu trúc và được đánh nhãn để tạo ra đầu ra chính xác, các mô hình học sâu có thể sử dụng học không giám sát. Với học không giám sát, các mô hình học sâu có thể trích xuất các đặc điểm, tính năng và mối quan hệ cần thiết để tạo ra đầu ra chính xác từ dữ liệu thô và không có cấu trúc. Ngoài ra, các mô hình này thậm chí có thể đánh giá và tinh chỉnh đầu ra của chúng để tăng độ chính xác.

Học sâu là một khía cạnh của khoa học dữ liệu thúc đẩy nhiều ứng dụng và dịch vụ cải thiện tự động hóa, thực hiện các tác vụ phân tích và vật lý mà không cần sự can thiệp của con người. Điều này cho phép nhiều sản phẩm và dịch vụ hàng ngày - chẳng hạn như trợ lý kỹ thuật số, điều khiển TV bằng giọng nói, phát hiện gian lận thẻ tín dụng, xe tự lái và AI tạo ra.

\*\*Cách học sâu hoạt động\*\*

Mạng thần kinh, hoặc mạng thần kinh nhân tạo, cố gắng mô phỏng bộ não người thông qua sự kết hợp của dữ liệu đầu vào, trọng số và độ lệch - tất cả đều hoạt động như các neuron silicon. Các yếu tố này hoạt động cùng nhau để chính xác nhận dạng, phân loại và mô tả các đối tượng trong dữ liệu.

## Dịch thuật từ Tiếng Anh sang Tiếng Việt:

\*\*Mạng nơ-ron sâu\*\* bao gồm nhiều lớp các nút kết nối với nhau, mỗi lớp xây dựng dựa trên lớp trước để tinh chỉnh và tối ưu hóa dự đoán hoặc phân loại. Quá trình tính toán này được gọi là \*\*phân phối trực tiếp\*\*. Các lớp đầu vào và đầu ra của mạng nơ-ron sâu được gọi là \*\*các lớp khả kiến\*\*. Lớp đầu vào là nơi mô hình học sâu nhận dữ liệu để xử lý, và lớp đầu ra là nơi đưa ra dự đoán cuối cùng hoặc phân loại.

Một quá trình khác được gọi là \*\*phân phối ngược\*\* sử dụng các thuật toán, chẳng hạn như \*\*giảm gradient\*\*, để tính toán lỗi trong các dự đoán, sau đó điều chỉnh trọng số và độ lệch của hàm bằng cách di chuyển ngược lại qua các lớp để đào tạo mô hình. Cùng nhau, phân phối trực tiếp và phân phối ngược cho phép mạng nơ-ron đưa ra dự đoán và khắc phục bất kỳ lỗi nào. Theo thời gian, thuật toán trở nên ngày càng chính xác.

\*\*Học sâu\*\* đòi hỏi một lượng công suất tính toán khổng lồ. \*\*Card đồ họa (GPU) có hiệu suất cao\*\* là lý tưởng vì chúng có thể xử lý một lượng lớn các phép tính trong nhiều lõi với bộ nhớ dồi dào. \*\*Tính toán đám mây phân tán\*\* cũng có thể hỗ trợ. Mức công suất tính toán này là cần thiết để đào tạo các thuật toán sâu thông qua học sâu. Tuy nhiên, việc quản lý nhiều GPU trên địa phương có thể tạo ra một nhu cầu lớn về các nguồn lực nội bộ và trở nên vô cùng tốn kém để mở rộng. Đối với yêu cầu phần mềm, hầu hết các ứng dụng học sâu được mã hóa bằng một trong ba khung làm việc học tập này: \*\*JAX, PyTorch hoặc TensorFlow\*\*.

\*\*Decoding AI: Tóm tắt tin tức hàng tuần\*\*

Tham gia hội thảo của chúng tôi với các chuyên gia hàng đầu về kỹ thuật, nghiên cứu, lãnh đạo sản phẩm và nhiều hơn nữa khi họ phân tích thông tin AI để mang đến cho bạn những tin tức và洞察力 mới nhất về AI.

\*\*Xem các tập podcast mới nhất\*\*

\*\*Các loại mô hình học sâu>\*\*

## Dịch thuật:

\*\*Các thuật toán học sâu vô cùng phức tạp, và có nhiều loại mạng thần kinh để giải quyết các vấn đề hoặc bộ dữ liệu cụ thể. Dưới đây là sáu loại mạng thần kinh đó. Mỗi loại có những ưu điểm riêng và được trình bày ở đây theo thứ tự phát triển của chúng, với mỗi mô hình tiếp theo được điều chỉnh để khắc phục nhược điểm của mô hình trước.\*\*

\*\*Một điểm yếu tiềm ẩn chung cho tất cả các mô hình này là các mô hình học sâu thường là "hộp đen", làm cho việc hiểu hoạt động bên trong của chúng trở nên khó khăn và gây ra thách thức về khả năng giải thích. Tuy nhiên, điều này có thể được cân bằng với những lợi ích chung về độ chính xác cao và khả năng mở rộng.\*\*

\*\*CNNs\*\*

\*\*Mạng thần kinh畳み込み (CNNs hoặc ConvNets) chủ yếu được sử dụng trong ứng dụng tầm nhìn máy tính và phân loại hình ảnh. Chúng có thể phát hiện các đặc điểm và mẫu trong hình ảnh và video, cho phép thực hiện các nhiệm vụ như phát hiện đối tượng, nhận dạng hình ảnh, nhận dạng mẫu và nhận dạng khuôn mặt. Các mạng này sử dụng các nguyên tắc từ đại số tuyến tính, đặc biệt là phép nhân ma trận, để xác định các mẫu trong hình ảnh.\*\*

\*\*CNNs là một loại mạng thần kinh cụ thể, được cấu thành bởi các lớp nút, bao gồm một lớp đầu vào, một hoặc nhiều lớp ẩn và một lớp đầu ra. Mỗi nút kết nối với nhau và có một trọng số và ngưỡng được liên kết. Nếu đầu ra của bất kỳ nút nào vượt quá giá trị ngưỡng được chỉ định, nút đó sẽ được kích hoạt, gửi dữ liệu đến lớp tiếp theo của mạng. Ngược lại, không có dữ liệu nào được truyền sang lớp tiếp theo của mạng.\*\*

## Dịch từ Tiếng Anh sang Tiếng Việt

\*\*Mạng nơ-ron Convolutional (CNN)\*\*

Mạng nơ-ron Convolutional (CNN) được cấu tạo bởi ít nhất ba loại lớp chính: lớp convolutional, lớp pooling và lớp hoàn toàn kết nối (FC). Cho các ứng dụng phức tạp, một CNN có thể chứa hàng ngàn lớp, mỗi lớp được xây dựng dựa trên các lớp trước đó.

Qua quá trình "convolution" - việc xử lý và tái xử lý đầu vào ban đầu - các mô hình chi tiết có thể được phát hiện. Với mỗi lớp, CNN tăng độ phức tạp của nó, xác định các phần lớn hơn của hình ảnh. Các lớp đầu tiên tập trung vào các tính năng đơn giản, chẳng hạn như màu sắc và cạnh. Khi dữ liệu hình ảnh di chuyển qua các lớp của CNN, nó bắt đầu nhận ra các yếu tố hoặc hình dạng lớn hơn của đối tượng cho đến khi cuối cùng xác định được đối tượng được muốn.

CNN được phân biệt với các mạng nơ-ron khác bởi hiệu suất vượt trội của chúng với đầu vào hình ảnh, giọng nói hoặc tín hiệu âm thanh. Trước khi CNN, các phương pháp trích xuất tính năng thủ công và tốn thời gian đã được sử dụng để xác định đối tượng trong hình ảnh. Tuy nhiên, CNN hiện cung cấp một cách tiếp cận khả năng mở rộng hơn cho các nhiệm vụ phân loại hình ảnh và nhận dạng đối tượng, và xử lý dữ liệu đa chiều. CNN cũng có thể trao đổi dữ liệu giữa các lớp để cung cấp việc xử lý dữ liệu hiệu quả hơn. Mặc dù có thể mất thông tin trong lớp pooling, điều này có thể bị bù đắp bởi lợi ích của CNN, giúp giảm độ phức tạp, cải thiện hiệu suất và hạn chế rủi ro quá điều chỉnh.

CNN cũng có một số nhược điểm, bao gồm việc yêu cầu tính toán cao - tốn thời gian và ngân sách, cần nhiều bộ xử lý đồ họa (GPU). Chúng cũng đòi hỏi các chuyên gia được đào tạo kỹ lưỡng với kiến thức đa lĩnh vực và việc kiểm tra cẩn thận các cấu hình, siêu tham số và cấu hình.

## Dịch thuật từ Tiếng Anh sang Tiếng Việt:

\*\*Mạng nơ-ron lặp (RNNs)\*\* thường được sử dụng trong các ứng dụng nhận dạng ngôn ngữ tự nhiên và giọng nói vì chúng sử dụng dữ liệu theo thứ tự hoặc dữ liệu thời gian. RNNs có thể được xác định bằng các vòng lặp phản hồi của chúng. Các thuật toán học này chủ yếu được sử dụng khi sử dụng dữ liệu thời gian để dự đoán kết quả trong tương lai. Các trường hợp sử dụng bao gồm dự đoán thị trường chứng khoán hoặc dự báo doanh số, hoặc các vấn đề thứ tự hoặc thời gian, chẳng hạn như dịch thuật ngôn ngữ, xử lý ngôn ngữ tự nhiên (NLP), nhận dạng giọng nói và thêm chú thích cho hình ảnh. Các chức năng này thường được tích hợp vào các ứng dụng phổ biến như Siri, tìm kiếm bằng giọng nói và Google Translate.

RNNs sử dụng "bộ nhớ" của chúng khi chúng lấy thông tin từ các đầu vào trước để ảnh hưởng đến đầu vào và đầu ra hiện tại. Trong khi mạng nơ-ron sâu truyền thống giả định rằng đầu vào và đầu ra độc lập với nhau, thì đầu ra của RNNs phụ thuộc vào các phần tử trước trong chuỗi. Mặc dù các sự kiện trong tương lai cũng sẽ hữu ích trong việc xác định đầu ra của một chuỗi cụ thể, nhưng mạng nơ-ron lặp đơn chiều không thể tính đến các sự kiện này trong các dự đoán của chúng.

RNNs chia sẻ các tham số trên mỗi lớp của mạng và chia sẻ cùng một tham số trọng số trong mỗi lớp của mạng, với trọng số được điều chỉnh thông qua các quá trình backpropagation và gradient descent để tạo điều kiện cho việc học tăng cường.

RNNs sử dụng thuật toán backpropagation qua thời gian (BPTT) để xác định các gradient, điều này hơi khác với backpropagation truyền thống vì nó đặc biệt cho dữ liệu theo chuỗi. Các nguyên tắc của BPTT giống như backpropagation truyền thống, nơi mô hình tự huấn luyện bằng cách tính toán lỗi từ lớp đầu ra đến lớp đầu vào. BPTT khác với cách tiếp cận truyền thống ở chỗ BPTT tổng hợp lỗi tại mỗi bước thời gian, trong khi mạng lưới truyền thông tuyến tính không cần phải tổng hợp lỗi vì chúng không chia sẻ các tham số trên mỗi lớp.

## Dịch Tiếng Anh sang Tiếng Việt

\*\*Lợi thế của mạng nơ-ron lặp (RNN) so với các loại mạng nơ-ron khác là chúng sử dụng cả xử lý dữ liệu nhị phân và bộ nhớ. RNN có thể lên kế hoạch cho nhiều đầu vào và đầu ra, vì vậy thay vì chỉ đưa ra một kết quả cho một đầu vào duy nhất, RNN có thể tạo ra đầu ra một-đến-nhiều, nhiều-đến-một hoặc nhiều-đến-nhiều.\*\*

\*\*Ngoài ra còn có các tùy chọn bên trong RNN. Ví dụ, mạng bộ nhớ ngắn hạn dài (LSTM) vượt trội hơn so với RNN đơn giản bằng cách học và hành động trên các sự phụ thuộc dài hạn.\*\*

\*\*Tuy nhiên, RNN thường gặp hai vấn đề cơ bản, được gọi là gradient nổ và gradient biến mất. Những vấn đề này được xác định bởi kích thước của gradient, đó là độ dốc của hàm mất mát dọc theo đường cong lỗi.\*\*

\*\*Khi gradient biến mất và quá nhỏ, nó tiếp tục trở nên nhỏ hơn, cập nhật các tham số trọng số cho đến khi chúng trở nên không đáng kể - tức là bằng không (0). Khi đó, thuật toán không còn học được nữa.\*\*

\*\*Gradient nổ xảy ra khi gradient quá lớn, tạo ra một mô hình không ổn định. Trong trường hợp này, trọng số của mô hình tăng quá lớn và cuối cùng sẽ được biểu diễn là NaN (không phải là một số). Một giải pháp cho những vấn đề này là giảm số lớp ẩn bên trong mạng nơ-ron, loại bỏ một số phức tạp trong các mô hình RNN.\*\*

\*\*Một số nhược điểm cuối cùng: RNN cũng có thể yêu cầu thời gian huấn luyện dài và khó sử dụng trên các tập dữ liệu lớn. Ốp-tim RNN thêm phức tạp khi chúng có nhiều lớp và tham số.\*\*

\*\*Autoencoders và variational autoencoders\*\*

\*\*Deep learning đã cho phép vượt qua việc phân tích dữ liệu số, bằng cách thêm việc phân tích hình ảnh, giọng nói và các loại dữ liệu phức tạp khác. Trong số những lớp mô hình đầu tiên đạt được điều này là variational autoencoders (VAEs). Chúng là những mô hình học sâu đầu tiên được sử dụng rộng rãi để tạo ra hình ảnh và giọng nói có thật, điều này đã tăng cường việc tạo mô hình tạo ra sâu bằng cách làm cho các mô hình dễ dàng hơn để mở rộng - đó là nền tảng của những gì chúng ta nghĩ về như AI tạo ra.\*\*

## Dịch thuật từ Tiếng Anh sang Tiếng Việt:

\*\*Autoencoders hoạt động bằng cách mã hóa dữ liệu chưa có nhãn vào một dạng biểu diễn nén, sau đó giải mã dữ liệu trở lại dạng ban đầu của nó.\*\* Autoencoders đơn giản được sử dụng cho một loạt các mục đích, bao gồm cả việc phục hồi lại hình ảnh bị lỗi hoặc mờ. \*\*Autoencoders biến thể (VAEs) đã bổ sung khả năng quan trọng không chỉ để phục hồi dữ liệu mà còn để đưa ra các biến thể của dữ liệu gốc.\*\*

\*\*Khả năng tạo ra dữ liệu mới đã kích hoạt một chuỗi liên tục các công nghệ mới, từ mạng đối kháng tạo ra (GANs) đến các mô hình lan truyền, có khả năng tạo ra các hình ảnh ngày càng chân thực hơn - nhưng là giả mạo.\*\* Theo cách này, VAEs đã đặt nền móng cho AI tạo ra ngày nay.

\*\*Autoencoders được xây dựng từ các khối mã hóa và giải mã, một kiến trúc cũng là nền tảng cho các mô hình ngôn ngữ lớn hiện nay.\*\* Các mã hóa nén một tập dữ liệu thành một dạng biểu diễn dày đặc, sắp xếp các điểm dữ liệu tương tự gần nhau hơn trong một không gian trừu tượng. Các giải mã lấy mẫu từ không gian này để tạo ra một cái gì đó mới trong khi vẫn giữ lại các đặc điểm quan trọng nhất của tập dữ liệu.

\*\*Lợi thế lớn nhất của autoencoders là khả năng xử lý các lô lớn dữ liệu và hiển thị dữ liệu đầu vào dưới dạng nén, khiến các khía cạnh quan trọng nhất nổi bật - cho phép các nhiệm vụ phát hiện và phân loại sự bất thường.\*\* Điều này cũng tăng tốc truyền tải và giảm yêu cầu lưu trữ. Autoencoders có thể được đào tạo trên dữ liệu chưa có nhãn nên chúng có thể được sử dụng trong trường hợp dữ liệu có nhãn không có sẵn. Khi sử dụng đào tạo không giám sát, có lợi thế về tiết kiệm thời gian: các thuật toán học sâu tự động học và đạt được độ chính xác mà không cần kỹ thuật xử lý đặc trưng thủ công. Ngoài ra, VAEs có thể tạo ra dữ liệu mẫu mới cho việc tạo văn bản hoặc hình ảnh.

## Nhược điểm của Autoencoders và Mô hình GAN

\*\*Autoencoders\*\*

Có một số nhược điểm của autoencoders. Việc đào tạo các cấu trúc sâu hoặc phức tạp có thể làm hao tốn tài nguyên tính toán. Và trong quá trình đào tạo không giám sát, mô hình có thể bỏ qua các thuộc tính cần thiết và thay vào đó chỉ đơn thuần sao chép dữ liệu đầu vào. Autoencoders cũng có thể bỏ qua các liên kết dữ liệu phức tạp trong dữ liệu cấu trúc nên không thể xác định chính xác các mối quan hệ phức tạp.

\*\*GANs (Mạng Nơ-ron đối kháng sinh sản)\*\*

Mạng đối kháng sinh sản (GANs) là các mạng nơ-ron được sử dụng cả trong và ngoài lĩnh vực trí tuệ nhân tạo (AI) để tạo ra dữ liệu mới giống với dữ liệu huấn luyện ban đầu. Những dữ liệu này có thể bao gồm các hình ảnh trông giống như khuôn mặt người - nhưng chúng được tạo ra, chứ không phải chụp ảnh từ những người thật. Phần "đối kháng" trong tên gọi xuất phát từ sự tương tác qua lại giữa hai phần của GAN: một bộ tạo và một bộ phân biệt.

\* \*\*Bộ tạo\*\* tạo ra một thứ gì đó: hình ảnh, video hoặc âm thanh và sau đó sản xuất một đầu ra với một sự thay đổi. Ví dụ, một con ngựa có thể được biến thành một con ngựa vằn với một mức độ chính xác nhất định. Kết quả phụ thuộc vào đầu vào và mức độ được huấn luyện của các lớp trong mô hình tạo ra cho trường hợp sử dụng này.

\* \*\*Bộ phân biệt\*\* là đối thủ, nơi kết quả tạo ra (hình ảnh giả) được so sánh với các hình ảnh thật trong bộ dữ liệu. Bộ phân biệt cố gắng phân biệt giữa các hình ảnh thật và giả, video hoặc âm thanh.

GANs đào tạo chính chúng. Bộ tạo tạo ra những giả mạo trong khi bộ phân biệt học cách phát hiện sự khác biệt giữa những giả mạo của bộ tạo và các ví dụ thật. Khi bộ phân biệt có thể đánh dấu giả mạo, bộ tạo sẽ bị phạt. Chu trình phản hồi tiếp tục cho đến khi bộ tạo thành công trong việc sản xuất đầu ra mà bộ phân biệt không thể phân biệt.

## Dịch thuật từ Tiếng Anh sang Tiếng Việt

\*\*Mô hình GAN\*\*

Lợi ích chính của mô hình GAN là tạo ra đầu ra có tính chân thực, khó phân biệt với bản gốc, từ đó có thể được sử dụng để đào tạo thêm các mô hình học máy. Việc thiết lập một mô hình GAN để học là đơn giản, vì chúng được đào tạo bằng cách sử dụng dữ liệu chưa được đánh nhãn hoặc với nhãn rất ít. Tuy nhiên, nhược điểm tiềm ẩn là bộ tạo và bộ phân biệt có thể luân phiên cạnh tranh trong một thời gian dài, dẫn đến việc tiêu tốn nhiều tài nguyên hệ thống. Một hạn chế trong việc đào tạo là có thể cần một lượng dữ liệu đầu vào rất lớn để có được đầu ra thỏa đáng. Một vấn đề tiềm ẩn khác là "mode collapse", khi bộ tạo chỉ sản xuất một tập hợp hạn chế các đầu ra thay vì một loạt đa dạng hơn.

\*\*Mô hình Diffusion\*\*

Mô hình Diffusion là các mô hình tạo ra được đào tạo bằng quá trình khuếch tán về phía trước và ngược lại của việc thêm tiếng ồn theo cấp số và làm giảm tiếng ồn. Mô hình Diffusion tạo ra dữ liệu - thường là hình ảnh - tương tự như dữ liệu mà chúng được đào tạo, nhưng sau đó ghi đè dữ liệu được sử dụng để đào tạo chúng. Chúng dần dần thêm tiếng ồn Gaussian vào dữ liệu huấn luyện cho đến khi nó không còn nhận biết được, sau đó học một quá trình "giảm tiếng ồn" ngược lại có thể tổng hợp đầu ra (thường là hình ảnh) từ đầu vào tiếng ồn ngẫu nhiên.

Một mô hình Diffusion học cách giảm thiểu sự khác biệt giữa các mẫu được tạo ra và mục tiêu mong muốn. Mọi sự sai lệch đều được định lượng và các tham số của mô hình được cập nhật để giảm thiểu tổn thất - đào tạo mô hình để tạo ra các mẫu tương tự như dữ liệu huấn luyện gốc.

Ngoài chất lượng hình ảnh, mô hình Diffusion có lợi thế là không cần đào tạo đối kháng, điều này làm tăng tốc quá trình học và cũng cung cấp khả năng kiểm soát quá trình gần như chính xác. Việc đào tạo ổn định hơn so với GAN và mô hình Diffusion không dễ bị "mode collapse".

## Dịch Tiếng Anh sang Tiếng Việt

\*\*Mô hình phân tán\*\*

Tuy nhiên, so với GANs, mô hình phân tán có thể yêu cầu nhiều tài nguyên tính toán hơn để đào tạo, bao gồm cả việc fine-tuning. IBM Research® cũng đã phát hiện ra rằng hình thức AI tạo ra này có thể bị tấn công bằng các backdoor ẩn, trao quyền kiểm soát cho người tấn công đối với quá trình tạo hình ảnh, khiến các mô hình phân tán AI bị lừa tạo ra các hình ảnh bị thao túng.

\*\*Mô hình Transformer\*\*

Mô hình Transformer kết hợp kiến trúc encoder-decoder với cơ chế xử lý văn bản và đã cách mạng hóa cách đào tạo các mô hình ngôn ngữ. Một encoder chuyển đổi văn bản thô, chưa được đánh dấu thành các đại diện được gọi là embeddings; decoder lấy các embeddings này cùng với các đầu ra trước của mô hình và dự đoán từng từ trong một câu một cách liên tiếp.

Bằng cách đoán các ô trống, encoder học cách các từ và câu liên quan đến nhau, xây dựng một đại diện mạnh mẽ về ngôn ngữ mà không cần phải đánh dấu các bộ phận của tiếng nói và các tính năng ngữ pháp khác. Transformer, thực tế, có thể được pretraining từ đầu mà không cần một nhiệm vụ cụ thể. Sau khi các đại diện mạnh mẽ này được học, các mô hình sau đó có thể được chuyên biệt hóa - với rất ít dữ liệu - để thực hiện một nhiệm vụ được yêu cầu.

Một số đổi mới làm cho điều này trở nên khả thi. Transformer xử lý các từ trong một câu đồng thời, cho phép xử lý văn bản song song, tăng tốc quá trình đào tạo. Các kỹ thuật trước đây bao gồm mạng nơ-ron lặp (RNNs) đã xử lý các từ một cách từng bước. Transformer cũng đã học được vị trí của các từ và mối quan hệ của chúng - ngữ cảnh này cho phép chúng suy ra nghĩa và loại bỏ bối cảnh của các từ như "nó" trong các câu dài.

## Dịch Tiếng Anh sang Tiếng Việt:

\*\*Bằng cách loại bỏ nhu cầu định nghĩa một nhiệm vụ trước, các mô hình transformer đã làm cho việc pre-training các mô hình ngôn ngữ trên lượng lớn văn bản thô trở nên thực tế, cho phép chúng phát triển mạnh về kích thước. Trước đây, dữ liệu được đánh dấu được thu thập để đào tạo một mô hình trên một nhiệm vụ cụ thể. Với các transformer, một mô hình được đào tạo trên một lượng dữ liệu khổng lồ có thể được điều chỉnh cho nhiều nhiệm vụ bằng cách fine-tuning nó trên một lượng nhỏ dữ liệu đánh dấu cụ thể cho nhiệm vụ.\*\*

\*\*Các transformer ngôn ngữ ngày nay được sử dụng cho các nhiệm vụ không tạo ra văn bản như phân loại và trích xuất thực thể cũng như các nhiệm vụ tạo ra văn bản bao gồm dịch máy, tóm tắt và trả lời câu hỏi. Transformer đã khiến nhiều người ngạc nhiên với khả năng tạo ra các cuộc hội thoại thuyết phục, bài luận và nội dung khác.\*\*

\*\*Các transformer xử lý ngôn ngữ tự nhiên (NLP) cung cấp sức mạnh đáng kinh ngạc vì chúng có thể chạy song song, xử lý nhiều phần của một chuỗi đồng thời, điều này sau đó làm tăng đáng kể tốc độ đào tạo. Transformer cũng theo dõi các sự phụ thuộc dài hạn trong văn bản, cho phép chúng hiểu ngữ cảnh tổng thể một cách rõ ràng hơn và tạo ra đầu ra vượt trội. Ngoài ra, transformer còn có khả năng mở rộng và linh hoạt hơn để được tùy chỉnh cho từng nhiệm vụ.\*\*

\*\*Về mặt hạn chế, do sự phức tạp của chúng, các transformer đòi hỏi các nguồn tài nguyên tính toán khổng lồ và thời gian đào tạo dài. Ngoài ra, dữ liệu đào tạo phải chính xác, không thiên vị và phong phú để sản xuất kết quả chính xác.\*\*

\*\*Các trường hợp sử dụng học sâu\*\*

\*\*Số lượng ứng dụng của học sâu ngày càng tăng. Đây chỉ là một số cách mà nó hiện đang giúp các doanh nghiệp trở nên hiệu quả hơn và phục vụ khách hàng tốt hơn.\*\*

\*\*Đa dạng ứng dụng>\*\*

## Dịch từ Tiếng Anh sang Tiếng Việt:

\*\*Trí tuệ nhân tạo thế hệ mới (Generative AI) có thể nâng cao năng lực của các nhà phát triển và thu hẹp khoảng cách kỹ năng ngày càng mở rộng trong các lĩnh vực hiện đại hóa ứng dụng và tự động hóa IT.\*\*

Trí tuệ nhân tạo thế hệ mới cho lập trình trở nên khả thi nhờ những đột phá gần đây trong công nghệ mô hình ngôn ngữ lớn (LLM) và xử lý ngôn ngữ tự nhiên (NLP). Nó sử dụng các thuật toán học sâu và mạng nơ-ron lớn được đào tạo trên các bộ dữ liệu khổng lồ của mã nguồn hiện có. Mã đào tạo thường đến từ mã nguồn mở được công bố rộng rãi từ các dự án mã nguồn mở.

\*\*Các lập trình viên có thể nhập các lệnh văn bản đơn giản mô tả những gì họ muốn mã thực hiện. Các công cụ trí tuệ nhân tạo thế hệ mới sẽ gợi ý các đoạn mã hoặc các hàm đầy đủ, làm cho quá trình lập trình hiệu quả hơn bằng cách xử lý các tác vụ lặp đi lặp lại và giảm thiểu việc lập trình thủ công.\*\* Trí tuệ nhân tạo thế hệ mới cũng có thể dịch mã từ một ngôn ngữ sang ngôn ngữ khác, giúp đơn giản hóa các dự án chuyển đổi hoặc hiện đại hóa mã, chẳng hạn như cập nhật các ứng dụng cũ bằng cách dịch COBOL sang Java.

\*\*Tầm nhìn máy tính (Computer Vision)\*\*

Tầm nhìn máy tính là một lĩnh vực của trí tuệ nhân tạo (AI) bao gồm phân loại hình ảnh, phát hiện đối tượng và phân đoạn ngữ nghĩa. Nó sử dụng học máy và mạng nơ-ron để dạy máy tính và các hệ thống học tập để trích xuất thông tin có ý nghĩa từ các hình ảnh kỹ thuật số, video và các đầu vào trực quan khác - và đưa ra các đề xuất hoặc thực hiện các hành động khi hệ thống phát hiện lỗi hoặc vấn đề. Nếu AI cho phép máy tính suy nghĩ, tầm nhìn máy tính cho phép chúng nhìn, quan sát và hiểu.

\*\*Bởi vì một hệ thống tầm nhìn máy tính thường được đào tạo để kiểm tra sản phẩm hoặc theo dõi tài sản sản xuất, nó thường có thể phân tích hàng ngàn sản phẩm hoặc quy trình mỗi phút, nhận ra các lỗi hoặc vấn đề không thể nhìn thấy bằng mắt thường.\*\* Tầm nhìn máy tính được sử dụng trong các ngành công nghiệp từ năng lượng và công utilities đến sản xuất và ô tô.

## Nhận diện hình ảnh: Giải mã thế giới thị giác của máy tính

Nhận diện hình ảnh (Computer vision) đòi hỏi một lượng lớn dữ liệu và sau đó thực hiện phân tích liên tục trên dữ liệu đó cho đến khi nó phân biệt và cuối cùng nhận ra hình ảnh. Ví dụ, để huấn luyện máy tính nhận ra lốp xe ô tô, nó cần được cung cấp một số lượng lớn hình ảnh lốp xe và các vật phẩm liên quan đến lốp xe để học cách phân biệt và nhận ra một lốp xe, đặc biệt là một lốp xe không có lỗi.

Nhận diện hình ảnh sử dụng các mô hình thuật toán để cho phép máy tính tự học về ngữ cảnh của dữ liệu hình ảnh. Nếu đủ dữ liệu được đưa vào mô hình, máy tính sẽ "nhìn" vào dữ liệu và tự học cách phân biệt một hình ảnh với hình ảnh khác. Thuật toán cho phép máy học một cách tự chủ, thay vì phải được lập trình để nhận ra một hình ảnh.

Nhận diện hình ảnh cho phép các hệ thống trích xuất thông tin có ý nghĩa từ các hình ảnh kỹ thuật số, video và các đầu vào thị giác khác, và dựa trên các đầu vào đó, thực hiện các hành động. Khả năng này cung cấp lời khuyên phân biệt nó với các nhiệm vụ nhận dạng hình ảnh đơn giản. Một số ứng dụng phổ biến của nhận diện hình ảnh ngày nay có thể được nhìn thấy trong:

\* \*\*Ô tô:\*\* Mặc dù thời đại xe tự lái chưa hoàn toàn đến, nhưng công nghệ nền tảng đã bắt đầu được tích hợp vào ô tô, cải thiện an toàn cho người lái và hành khách thông qua các tính năng như phát hiện đường phân làn.

\* \*\*Y tế:\*\* Nhận diện hình ảnh đã được tích hợp vào công nghệ chụp X-quang, cho phép bác sĩ dễ dàng xác định các khối u ung thư trong cơ thể khỏe mạnh hơn.

\* \*\*Marketing:\*\* Các nền tảng truyền thông xã hội cung cấp các đề xuất về người có thể xuất hiện trong một bức ảnh đã được đăng trên hồ sơ, giúp việc đánh dấu bạn bè trong album ảnh dễ dàng hơn.

\* \*\*Bán lẻ:\*\* Tìm kiếm bằng hình ảnh đã được tích hợp vào một số nền tảng thương mại điện tử, cho phép các thương hiệu đề xuất các mặt hàng sẽ bổ sung cho tủ quần áo hiện có.

\* \*\*Dịch vụ chăm sóc khách hàng:\*\*

## AI đang giúp doanh nghiệp hiểu và đáp ứng tốt hơn nhu cầu ngày càng tăng của người tiêu dùng.

Với sự gia tăng của việc mua sắm trực tuyến cá nhân hóa cao, mô hình bán hàng trực tiếp cho người tiêu dùng và dịch vụ giao hàng, AI thế hệ mới có thể giúp khai thác thêm nhiều lợi ích có thể cải thiện chăm sóc khách hàng, chuyển đổi tài năng và hiệu suất của các ứng dụng.

\*\*AI cho phép doanh nghiệp áp dụng cách tiếp cận khách hàng tập trung bằng cách khai thác những洞察 giá trị từ phản hồi và thói quen mua hàng của khách hàng.\*\* Cách tiếp cận dựa trên dữ liệu này có thể giúp cải thiện thiết kế và bao bì sản phẩm và có thể giúp thúc đẩy sự hài lòng của khách hàng cao và tăng doanh thu.

\*\*AI thế hệ mới cũng có thể hoạt động như trợ lý nhận thức cho chăm sóc khách hàng, cung cấp hướng dẫn ngữ cảnh dựa trên lịch sử cuộc trò chuyện, phân tích cảm xúc và bản ghi cuộc gọi trung tâm.\*\* Ngoài ra, AI thế hệ mới có thể cho phép trải nghiệm mua sắm cá nhân hóa, thúc đẩy sự trung thành của khách hàng và mang lại lợi thế cạnh tranh.

\*\*Lao động kỹ thuật số\*\*

Các tổ chức có thể tăng cường đội ngũ lao động của mình bằng cách xây dựng và triển khai tự động hóa quy trình robot (RPA) và lao động kỹ thuật số để hợp tác với con người nhằm tăng năng suất hoặc hỗ trợ khi cần sự hỗ trợ. Ví dụ: điều này có thể giúp các nhà phát triển tăng tốc việc cập nhật phần mềm cũ.

\*\*Lao động kỹ thuật số sử dụng các mô hình cơ sở để tự động hóa và cải thiện năng suất của người lao động có kiến thức bằng cách cho phép tự động hóa tự phục vụ một cách nhanh chóng và đáng tin cậy - mà không gặp phải các rào cản kỹ thuật.\*\* Để tự động hóa việc thực hiện tác vụ hoặc gọi API, một mô hình lấp đầy khe cắm dựa trên LLM cấp doanh nghiệp có thể xác định thông tin trong cuộc trò chuyện và thu thập tất cả thông tin cần thiết để hoàn thành một hành động hoặc gọi API mà không cần nhiều công sức thủ công.

## Dịch Tiếng Anh sang Tiếng Việt

\*\*Thay vì để các chuyên gia kỹ thuật ghi lại và mã hóa các luồng hành động lặp đi lặp lại cho người lao động có chuyên môn, các tự động hóa lao động kỹ thuật số được xây dựng trên nền tảng các hướng dẫn và minh họa hội thoại được điều khiển bởi mô hình có thể được người lao động có chuyên môn sử dụng cho tự động hóa tự phục vụ. Ví dụ, để tăng tốc việc tạo ứng dụng, các trợ lý kỹ thuật số không cần mã có thể giúp người dùng cuối, những người không có chuyên môn lập trình, bằng cách hiệu quả dạy, giám sát và xác minh mã.\*\*

\*\*Trí tuệ nhân tạo tạo ra (Generative AI)\*\*

Trí tuệ nhân tạo tạo ra (còn được gọi là gen AI) là một loại trí tuệ nhân tạo tự động tạo ra văn bản, hình ảnh, video, dữ liệu hoặc nội dung khác dựa trên yêu cầu hoặc yêu cầu của người dùng.

Trí tuệ nhân tạo tạo ra dựa trên các mô hình học sâu có thể học từ các mẫu trong nội dung hiện có và tạo ra nội dung mới, tương tự dựa trên việc đào tạo đó. Nó có ứng dụng trong nhiều lĩnh vực - bao gồm dịch vụ khách hàng, tiếp thị, phát triển phần mềm và nghiên cứu - và mang lại tiềm năng khổng lồ để đơn giản hóa các luồng công việc doanh nghiệp thông qua việc tạo và tăng cường nội dung tự động nhanh chóng.

Trí tuệ nhân tạo tạo ra xuất sắc trong việc xử lý các nguồn dữ liệu đa dạng như email, hình ảnh, video, tệp âm thanh và nội dung mạng xã hội. Dữ liệu không cấu trúc này tạo nên nền tảng để tạo mô hình và đào tạo liên tục của trí tuệ nhân tạo tạo ra, vì vậy nó có thể duy trì hiệu quả theo thời gian. Việc sử dụng dữ liệu không cấu trúc này có thể nâng cao dịch vụ khách hàng thông qua các trợ lý ảo và tạo điều kiện cho việc phân luồng email hiệu quả hơn. Trong thực tế, điều này có thể có nghĩa là hướng dẫn người dùng đến các nguồn lực phù hợp, cho dù đó là kết nối họ với đại diện phù hợp hay hướng dẫn họ đến các hướng dẫn người dùng và câu hỏi thường gặp.

\*\*Mặc dù những hạn chế và rủi ro được thảo luận nhiều, nhiều doanh nghiệp đang tiến lên phía trước, thận trọng khám phá cách tổ chức của họ có thể tận dụng trí tuệ nhân tạo tạo ra để cải thiện các luồng công việc nội bộ và nâng cao sản phẩm và dịch vụ của họ. Đây là ranh giới mới: Làm thế nào để làm cho nơi làm việc hiệu quả hơn mà không tạo ra các vấn đề pháp lý hoặc đạo đức.\*\*

\*\*Xử lý ngôn ngữ tự nhiên và nhận dạng giọng nói>\*\*

## Dịch thuật từ Tiếng Anh sang Tiếng Việt:

\*\*NLP (Ngôn Ngữ Tự Do)\*\* kết hợp ngôn ngữ tính toán - mô hình dựa trên quy tắc của ngôn ngữ con người - với các mô hình thống kê và học máy để cho phép máy tính và thiết bị kỹ thuật số nhận ra, hiểu và tạo ra văn bản và giọng nói. NLP cung cấp sức mạnh cho các ứng dụng và thiết bị có thể dịch văn bản từ một ngôn ngữ sang ngôn ngữ khác, phản hồi các lệnh nhập văn bản hoặc giọng nói, nhận dạng hoặc xác thực người dùng dựa trên giọng nói. Nó giúp tóm tắt các khối lượng văn bản lớn, đánh giá ý định hoặc cảm xúc của văn bản hoặc giọng nói và tạo ra văn bản, đồ họa hoặc nội dung khác theo yêu cầu.

Một tập con của NLP là NLP thống kê, kết hợp các thuật toán máy tính với các mô hình học máy và học sâu. Phương pháp này giúp tự động trích xuất, phân loại và dán nhãn các yếu tố của dữ liệu văn bản và giọng nói, sau đó gán khả năng thống kê cho mỗi ý nghĩa có thể của các yếu tố đó. Ngày nay, các mô hình học sâu và kỹ thuật học dựa trên RNN cho phép các hệ thống NLP "học" khi làm việc và trích xuất ý nghĩa chính xác hơn từ những khối lượng lớn dữ liệu văn bản thô, không được cấu trúc và chưa được dán nhãn.

\*\*Nhận dạng giọng nói\*\* - cũng được gọi là nhận dạng giọng nói tự động (ASR), nhận dạng giọng nói máy tính hoặc giọng nói sang văn bản - là một khả năng cho phép một chương trình xử lý giọng nói của con người thành định dạng văn bản.

Mặc dù nhận dạng giọng nói thường bị nhầm lẫn với nhận dạng giọng nói, nhưng nhận dạng giọng nói tập trung vào việc dịch chuyển giọng nói từ định dạng lời nói sang định dạng văn bản, trong khi nhận dạng giọng nói chỉ nhằm mục đích xác định giọng nói của một người dùng cụ thể.

\*\*Ứng dụng trong ngành công nghiệp\*\*

Các ứng dụng học sâu thực tế đang ở khắp mọi nơi, và được tích hợp rất tốt vào các sản phẩm và dịch vụ đến nỗi người dùng không nhận ra quy trình xử lý dữ liệu phức tạp đang diễn ra ở phía sau. Một số ví dụ về điều này bao gồm:

\* \*\*Dịch vụ khách hàng học sâu:\*\*

## Dịch thuật từ Tiếng Anh sang Tiếng Việt

Nhiều tổ chức tích hợp công nghệ học sâu vào các quy trình dịch vụ khách hàng của họ. Chatbot thường được sử dụng trong nhiều ứng dụng, dịch vụ và cổng thông tin chăm sóc khách hàng. Chatbot truyền thống sử dụng ngôn ngữ tự nhiên và thậm chí là nhận dạng hình ảnh, thường được tìm thấy trong các menu tương tự trung tâm gọi điện. Tuy nhiên, các giải pháp chatbot tinh vi hơn cố gắng xác định, thông qua việc học, liệu có nhiều câu trả lời cho các câu hỏi mơ hồ trong thời gian thực hay không. Dựa trên các câu trả lời nó nhận được, chatbot sẽ cố gắng trả lời trực tiếp những câu hỏi này hoặc chuyển hướng cuộc trò chuyện đến người dùng con người.

Các trợ lý ảo như Siri của Apple, Amazon Alexa hoặc Google Assistant mở rộng ý tưởng của chatbot bằng cách cho phép chức năng nhận dạng giọng nói. Điều này tạo ra một phương pháp mới để tương tác với người dùng theo cách cá nhân hóa.

\*\*Phân tích dữ liệu dịch vụ tài chính\*\*

Các tổ chức tài chính thường xuyên sử dụng phân tích dự đoán để thúc đẩy giao dịch thuật toán của cổ phiếu, đánh giá rủi ro kinh doanh cho việc phê duyệt khoản vay, phát hiện gian lận và giúp quản lý danh mục đầu tư và tín dụng cho khách hàng.

\*\*Quản lý hồ sơ y tế\*\*

Ngành y tế đã được hưởng lợi rất lớn từ khả năng học sâu kể từ khi hóa số hóa hồ sơ và hình ảnh bệnh viện. Các ứng dụng nhận dạng hình ảnh có thể hỗ trợ các chuyên gia chụp ảnh y tế và bác sĩ radio, giúp họ phân tích và đánh giá nhiều hình ảnh hơn trong thời gian ngắn hơn.

\*\*Lực lượng thực thi pháp luật sử dụng học sâu\*\*

Các thuật toán học sâu có thể phân tích và học từ dữ liệu giao dịch để xác định các mẫu nguy hiểm cho thấy hoạt động gian lận hoặc phạm tội có thể xảy ra. Nhận dạng giọng nói, tầm nhìn máy tính và các ứng dụng học sâu khác có thể cải thiện hiệu quả và hiệu quả của phân tích điều tra bằng cách trích xuất các mẫu và bằng chứng từ các bản ghi âm âm thanh và video, hình ảnh và tài liệu. Khả năng này giúp lực lượng thực thi pháp luật phân tích các khối lượng dữ liệu lớn một cách nhanh chóng và chính xác hơn.

## Dịch từ Tiếng Anh sang Tiếng Việt:

\*\*Nguồn:\*\* https://www.ibm.com/think/topics/deep-learning#:~:text=Deep%20learning%20is%20a%20subset,applications%20in%20our%20lives%20today.

\*\*Deep learning là một phân ngành của học máy, sử dụng các mạng thần kinh nhân tạo có nhiều lớp để phân tích dữ liệu phức tạp. Nó cho phép máy móc học cách nhận dạng các mẫu phức tạp, tương tự như cách con người học. Deep learning đang được sử dụng trong nhiều ứng dụng trong cuộc sống của chúng ta ngày nay, chẳng hạn như nhận dạng giọng nói, dịch thuật tự động và y học.\*\*