

## Trí tuệ nhân tạo quá trình phát triển, thời cơ và thách thức

Nguyễn Thị Thu Nga<sup>1\*</sup>, Hoàng Đức Thọ<sup>2</sup>, Trần Mạnh Đông<sup>3</sup>, Phạm Duy Nam<sup>4</sup>

<sup>1</sup>Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam, 18 Hoàng Quốc Việt, Cầu Giấy, Hà Nội, Việt Nam;

<sup>2</sup>Học viện Kỹ thuật Mật mã, Ban Cơ yếu Chính phủ, 144 Chiến Thắng, Thanh Trì, Hà Nội, Việt Nam;

<sup>3</sup>Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam, 18 Hoàng Quốc Việt, Cầu Giấy, Hà Nội, Việt Nam;

<sup>4</sup>Trường Đại học Nguyễn Huệ, Tam Phước, Biên Hòa, Đồng Nai, Việt Nam.

\*Tác giả liên hệ: thungaln@gmail.com

Nhận bài: 30/9/2024; Hoàn thiện: 06/11/2024; Chấp nhận đăng: 12/12/2024; Xuất bản: 25/12/2024.

DOI: <https://doi.org/10.54939/1859-1043.j.mst.100.2024.146-150>

### TÓM TẮT

Bài báo trình bày về quá trình phát triển, thời cơ và thách thức của trí tuệ nhân tạo (AI), một lĩnh vực có nhiều ứng dụng trong thực tiễn: y học, kinh tế quản lý, giáo dục, quân sự,... Phạm vi của trí tuệ nhân tạo gồm: thuật toán, heuristic, thuật toán di truyền, hệ chuyên gia, mạng nơ-ron nhân tạo và logic mờ,... Bài báo cũng phân tích việc tạo ra những cỗ máy thông minh có thể suy nghĩ và đưa ra quyết định một cách độc lập và sử dụng công nghệ này một cách tùy tiện, nguy hiểm có thể dẫn đến thảm họa (như trường hợp sử dụng năng lượng hạt nhân). Tuy nhiên, những lo ngại về tầm quan trọng của con người bị giảm sút là không có cơ sở vì con người không chỉ tạo ra cỗ máy thông minh, mà còn tạo ra cỗ máy có ý thức và cá tính riêng.

**Từ khóa:** Trí tuệ nhân tạo; Hệ chuyên gia; Mạng nơ-ron nhân tạo; Logic mờ; Máy thông minh.

### 1. MỞ ĐẦU

Khái niệm "trí tuệ nhân tạo" được sử dụng phổ biến, nhưng không dễ định nghĩa, do thiếu một định nghĩa rõ ràng và chính xác về trí tuệ. Trí tuệ là khả năng xử lý thông tin ở cấp độ trừu tượng, sáng tạo, không máy móc. Nhận thức hiện đại về trí tuệ được hiểu là khả năng tương tác được tạo ra trong các lĩnh vực cảm xúc, tâm lý con người. Thuật ngữ "trí tuệ nhân tạo" lần đầu tiên được định nghĩa vào năm 1955 bởi John McCarthy: "Trí tuệ nhân tạo là một lĩnh vực nghiên cứu cố gắng bắt chước trí thông minh của con người trong một cỗ máy. Lĩnh vực AI bao gồm các hệ thống có cơ sở tri thức, hệ thống chuyên gia, nhận dạng hình ảnh, học tự động, hiểu ngôn ngữ tự nhiên, người máy và các hệ thống khác".

Trí tuệ nhân tạo (Artificial Intelligence, AI) là một nhánh của công nghệ thông tin liên quan đến việc thiết kế xây dựng các máy móc và thuật toán, mà hoạt động của chúng có dấu hiệu trí tuệ, có khả năng thích ứng một cách tự nhiên với các điều kiện thay đổi, đưa ra các quyết định phức tạp, học hỏi, suy luận trừu tượng [1],... Chủ đề của trí tuệ nhân tạo là nghiên cứu và xác định các quy tắc chi phối hành vi trí tuệ của con người và sử dụng chúng trong các thuật toán và chương trình máy tính biết sử dụng các quy tắc này. Có hai cách tiếp cận về trí tuệ nhân tạo: Trí tuệ nhân tạo mạnh (Strong AI) và Trí tuệ nhân tạo yếu (Weak AI). Những người ủng hộ AI yếu cho rằng "máy tính có thể xây dựng và kiểm tra các giả thuyết về bộ não. Bộ não thực hiện rất nhiều phép tính theo cách các ấn tượng giác quan được xử lý trước khi được hình thành trong tâm trí chúng ta ở một mức độ nào đó và có thể tạo ra các mô hình toán học toàn diện cho các vấn đề được phân tích và triển khai chúng dưới dạng các chương trình máy tính để đạt được các mục tiêu cụ thể. Những tiếp cận táo bạo hơn liên quan đến AI mạnh, khi đó một máy tính được lập trình đúng cách sẽ tương đương với bộ não và sẽ có các yếu tố trí tuệ của con người. Có thể tạo ra các cấu trúc và chương trình "tự học", như mô hình mạng nơ-ron, phát triển các quy trình giải quyết bài toán bằng cách "dạy" các chương trình đó và nhận được câu trả lời thích ứng.

Sự đầu tư nghiên cứu về AI trên thế giới đã mang lại nhiều kết quả to lớn và nhiều ứng dụng thiết thực trong nhiều lĩnh vực: y học, kinh tế, quản lý, quân sự,... Đó là những giải pháp liên quan đến:

- Các công nghệ dựa trên logic mờ.
- Hệ thống chuyên gia, tức là cơ sở dữ liệu mở rộng được cây "trí tuệ nhân tạo".
- Máy dịch văn bản - hệ thống SYSTRANS [3].
- Mạng nơ ron - được sử dụng trong xấp xỉ và nội suy, nhận dạng và phân loại, nén, dự đoán...
- Khai thác dữ liệu - các lĩnh vực liên quan đến thông tin, thu thập kiến thức, kỹ thuật phân tích,...
- Nhận dạng quang học - chương trình nhận dạng người dựa trên ảnh khuôn mặt hoặc nhận dạng tự động các đối tượng được chỉ định trong ảnh vệ tinh,...
- Nhận dạng giọng nói - nhận dạng nội dung của câu nói và nhận dạng người nói,...
- Nhận dạng chữ viết tay - đã được sử dụng trên quy mô lớn,...
- Sáng tác nhân tạo: chương trình tự động sáng tác các thể thơ, diễn giải các tác phẩm âm nhạc,...
- Kinh tế học - các hệ thống đánh giá tự động được sử dụng phổ biến.

## 2. PHẠM VI TRÍ TUỆ NHÂN TẠO

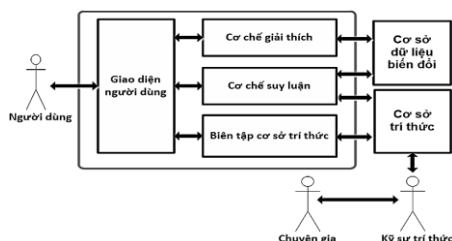
Trí tuệ nhân tạo bao gồm các thuật toán tiến hóa, heuristic, thuật toán di truyền, hệ chuyên gia, mạng nơ ron nhân tạo và logic mờ.

### Thuật toán tiến hóa

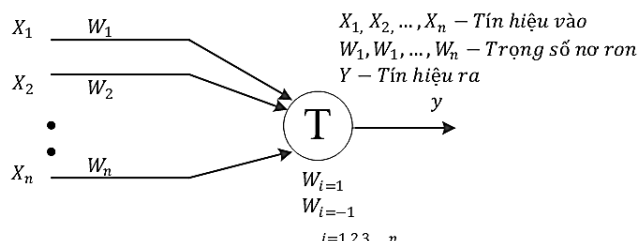
Thuật toán tiến hóa (Evolutionary Algorithms - EA) là một kỹ thuật tìm kiếm và tối ưu hóa dựa trên các nguyên tắc bắt nguồn từ lý thuyết tiến hóa. Thường được sử dụng trong khoa học quản lý để giải quyết các bài toán về tối ưu hóa tổ hợp, và phân bổ tài nguyên.

### Hệ chuyên gia

Hệ chuyên gia là các chương trình máy tính được thiết kế để đưa ra quyết định và cung cấp cho người dùng thông tin muốn có. Hệ thống này mô hình hóa kiến thức của một chuyên gia trong một lĩnh vực khoa học nhất định. Ý tưởng của một hệ chuyên gia là thu thập kiến thức mà một chuyên gia có và chuyển sang một chương trình máy tính dựa trên cơ sở kiến thức và các quy tắc suy luận cụ thể,...



**Hình 1.** Sơ đồ hệ chuyên gia [1].



**Hình 2.** Mô hình nơ-ron nhân tạo McCulloch-Pitts [1].

### Mạng nơ ron nhân tạo

Mạng nơ-ron nhân tạo ANN (Artificial neural networks) là một mô hình toán học được xây dựng dựa trên các mạng nơ-ron sinh học, gồm có một nhóm các nơ-ron nhân tạo nối với nhau và xử lý thông tin bằng cách truyền theo các kết nối và tính giá trị mới tại các nút. Mạng nơ-ron nhân tạo là một hệ thống thích ứng (adaptive system) tự thay đổi cấu trúc của mình dựa trên các thông tin bên ngoài hay bên trong chảy qua mạng trong quá trình học. Các công trình nghiên cứu phổ biến nhất về mạng nơ-ron nhân tạo như: nén ảnh, tái tạo ảnh và nhận dạng ảnh, nhận dạng khuôn mặt,... trong việc giám sát, bảo mật, lưu trữ thông tin,... Hệ thống nhận diện khuôn mặt được sử dụng rộng rãi trên thiết bị điện thoại thông minh; kiểm soát tình trạng ra vào tại các tòa nhà chung cư, văn phòng, trung tâm thương mại; giám sát an ninh trật tự tại khu dân cư,...

Ngày 8/10/2024 tại Thủ đô Stockholm, Viện Hàn lâm Khoa học Hoàng gia Thụy Điển đã quyết định trao Giải thưởng Nobel Vật lý năm 2024 cho 2 nhà khoa học John J. Hopfield và Geoffrey E. Hinton "vì những khám phá và phát minh nền tảng cho phép học máy sử dụng mạng nơ-ron nhân

tạo". John Hopfield đã tạo ra một cấu trúc có thể lưu trữ và tái tạo thông tin. Còn Geoffrey Hinton đã phát minh ra một phương pháp có thể tự khám phá các thuộc tính trong dữ liệu; phương pháp này quan trọng đối với các mạng nơ-ron nhân tạo lớn đang được sử dụng.

Mạng nơ-ron được ứng dụng để xác minh chữ ký, nhận dạng các ký tự viết tay hoặc phân tích, nhận dạng tín hiệu khác như: nhận dạng tiếng nói và tín hiệu âm thanh và phân tích văn bản viết bằng ngôn ngữ tự nhiên; các ứng dụng trong chẩn đoán y tế, trong phân tích tín hiệu radar,... Mạng nơ-ron nhân tạo liên hệ chặt chẽ với lĩnh vực trí tuệ nhân tạo, đặc biệt là các hệ chuyên gia, ANN trong quân sự được sử dụng định vị đối tượng và sử dụng vũ khí, theo dõi mục tiêu,...

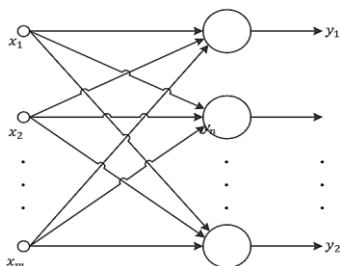
McCulloch và Pitts đã đưa ra định nghĩa mô hình nơ-ron nhân tạo. Cấu trúc và nguyên tắc hoạt động của mô hình này dựa trên đối tác sinh học. Giả định ở đây là thay thế nơ-ron bằng một đơn vị nhị phân. Mô hình đề xuất như sau:

Tín hiệu vào tại các điểm  $x_1, x_2, \dots, x_n$  có giá trị nhị phân là 0 hoặc 1. Nếu tại thời điểm  $k$  tại một điểm xuất hiện xung (impuls) thì điểm đó nhận giá trị 1, nếu không có xung thì nhận giá trị 0. Tín hiệu ra nhận giá trị  $y$ . Quy tắc hoạt động nơ-ron có dạng:

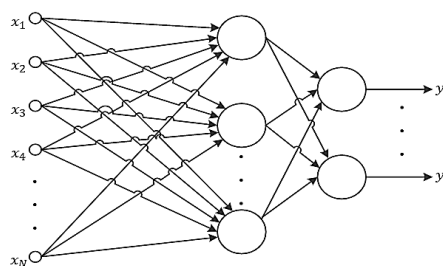
$$y^{k+1} = \begin{cases} 1, & \text{gdy } \sum_{i=1}^n w_i x_i^k \geq T, \\ 0, & \text{gdy } \sum_{i=1}^n w_i x_i^k < T, \end{cases} \quad (1)$$

Trong đó:  $k = 0, 1, 2, \dots$  là các thời điểm liên tiếp. Trong khoảng thời gian  $k$  và  $k+1$ , có một đơn vị thời gian trôi qua,  $w_i$  là trọng số nhân được gán cho đầu vào  $i$  với màng tế bào thần kinh. Đối với synap kích thích  $w_i = +1$ , đối với synap ức chế  $w_i = -1$ . Dưới ngưỡng  $T$ , nơ-ron không hoạt động.

*Các loại mạng nơ-ron nhân tạo phổ biến*



**Hình 3.** Mạng đơn hướng một lớp [9]. **Hình 4.** Mạng đơn hướng nhiều lớp [9].



*Mạng đơn hướng một lớp* - Các nơ-ron được sắp xếp trong một lớp, được cung cấp năng lượng từ các nút đầu vào. Luồng tín hiệu trong loại mạng này luôn chạy theo một hướng được xác định: từ lớp đầu vào đến lớp đầu ra. Việc lựa chọn trọng số diễn ra ở đây trong quá trình dạy mạng, tức là khớp các tín hiệu đầu ra  $y_i$  với giá trị mong đợi  $d_i$

*Mạng đơn hướng nhiều lớp* - Đặc điểm phân biệt các mạng này với các mạng khác là ngoài các lớp đầu vào và đầu ra, còn có ít nhất một lớp tế bào nơ-ron ẩn. Các tín hiệu đầu vào được đưa đến lớp tế bào nơ-ron ẩn đầu tiên, lớp tế bào nơ-ron này lại là tín hiệu nguồn cho lớp tiếp theo. Giá trị của tín hiệu đầu vào bị ảnh hưởng bởi trọng số của cả hai lớp.

*Mạng đệ quy*: Có thể phân biệt *mạng đệ quy một lớp* với một lớp nơ-ron đầu ra và *mạng đệ quy nhiều lớp* với một lớp ẩn bổ sung. Chúng khác với các mạng đơn hướng ở chỗ có phản hồi giữa các lớp đầu ra và đầu vào.

**Lô gic mờ**

Hệ thống Fuzzy Logic (*Lô gic mờ*) là một hệ thống điều khiển hoặc quyết định dựa trên nguyên tắc "mờ" thay vì logic truyền thống giữa đúng và sai, chấp nhận sự mập mờ và không rõ ràng trong

thông tin đầu vào và xử lý dữ liệu theo cách mà logic nhị phân không thể làm được. Trong logic mờ, giữa trạng thái mô tả không đúng và trạng thái đúng có một số giá trị trung gian [1],...

### ***Những công cụ được chọn để xây dựng mạng lưới nơ ron nhân tạo***

**Matlab** - là một môi trường tương tác trong đó các tính toán kỹ thuật được thực hiện, được trang bị ngôn ngữ lập trình riêng. Về mặt đồ họa, có thể vẽ biểu đồ hai, ba chiều; cũng như trực quan hóa kết quả tính toán dưới dạng bản vẽ tĩnh và hiển thị hình ảnh động. Các hàm chuyên biệt (hộp công cụ) mà Matlab trang bị để: tính toán ma trận, mạng nơ-ron, xử lý ảnh, xử lý tín hiệu, logic mờ,...

**Thư viện Hộp công cụ Mạng nơ ron** có hơn 200 m-file, được chia thành hai nhóm: **NNET** – gồm các tệp chức năng cho phép thiết kế, đào tạo và mô phỏng mạng nơ ron nhân tạo và **NNDEMOS** – gồm các tập lệnh demo m-file với nhiều ví dụ về ứng dụng thực tế của mạng nơ ron nhân tạo.

**Mạng Neural Statistica SNN** (Statistica Neural Networks) - công cụ có sẵn để thiết kế và mô phỏng mạng nơ ron. Tính năng quan trọng nhất của SNN gồm: Quy trình xử lý dữ liệu đầu vào, đầu ra và xử lý kết quả, mã hóa các biến danh nghĩa,... xây dựng các mạng kết nối với kiến trúc đa dạng và kích thước không giới hạn.

### ***Một vài dự án AI tiêu biểu***

**Dự án SingularityNET** do Tiến sĩ Ben Goertzel và Tiến sĩ David Hanson thành lập vào năm 2017 nhằm mục đích cung cấp công cụ thương mại cho các ứng dụng phi tập trung dựa trên công nghệ AI thông qua việc cải thiện khả năng tương tác. Các sản phẩm đáng chú ý của SingularityNET gồm: AI Marketplace và AI Publisher giúp cho các nhà phát triển triển khai các dịch vụ AI và AGIX Token Tools.

**ChatGPT** (*Chat Generative Pre-training Transformer*) là một chatbot do công ty OpenAI của Mỹ phát triển và ra mắt vào tháng 11 năm 2022. ChatGPT được ra mắt dưới dạng nguyên mẫu vào tháng 11 năm 2022 và nhanh chóng thu hút sự chú ý. Sau khi phát hành ChatGPT, OpenAI được định giá 29 tỷ USD[8]. Đến ngày 4 tháng 12 năm 2022, OpenAI ước tính ChatGPT đã có hơn một triệu người dùng và đến 31 tháng 1 năm 2023, ứng dụng đã đạt 100 triệu người dùng chỉ sau 2 tháng ra mắt [2].

**Fetch.AI (FET)** là một giao thức mở sử dụng cơ chế đồng thuận Useful Proof-Of-Work, tận dụng trí tuệ nhân tạo (AI) để tạo ra một hệ thống máy học (machine learning) phi tập trung có thể thực thi các tác vụ trên blockchain. Là một giao thức liên chuỗi (interchain protocol), Fetch.AI cung cấp một môi trường mở một cách tự động. Dự án được công bố vào tháng 2/2019 trên Binance Launchpad.

## **3. MỐI ĐE DỌA VÀ THÁCH THỨC CỦA VIỆC SỬ DỤNG TRÍ TUỆ NHÂN TẠO**

Trước khi nói một cách nghiêm túc về những mối đe dọa thực sự phát sinh từ việc sử dụng trí tuệ nhân tạo, cần trả lời câu hỏi: khi nào những cỗ máy thực sự thông minh sẽ được tạo ra? Có rất nhiều ý tưởng về robot trong tương lai và có nhiều dự án được thực hiện để chế tạo robot hình người, nhưng trong tương lai máy móc thông minh sẽ không phải là robot biết đi và biết nói. Nguyên tắc tương tự cũng được sử dụng khi xây dựng Cỗ máy thông minh (IM). Máy móc sử dụng một bộ giác quan, nhưng khác với giác quan mà con người sử dụng. Với sự phát triển của khoa học, trí thông minh của máy móc được thể hiện trong việc tạo ra mô hình thế giới riêng của chúng và tìm hiểu về thế giới đó thông qua quan sát và hướng dẫn của nhà thiết kế. Máy móc thông minh vật lý có thể được tích hợp vào máy bay, ô tô hoặc bất kỳ nơi nào phụ thuộc vào mục đích của chính máy.

Nếu AI phát triển hoàn thiện, nó có khả năng thay thế con người trong các công việc trí tuệ như chăm sóc sức khỏe, phục vụ, sản xuất theo dây chuyền tự động, công việc văn phòng,... Theo Bill Joy, người đồng sáng lập và Giám đốc khoa học của Sun Microsystems: "*Khi AI trở nên hoàn thiện và thông minh hơn, chúng ta sẽ cho phép mình nghe theo những quyết định của máy móc, vì đơn giản là các cỗ máy luôn đưa ra quyết định chính xác hơn con người*" [6].

Viễn cảnh về sự xuất hiện của những IM có thể suy nghĩ và đưa ra quyết định một cách độc lập khiến mọi người lo lắng. Tuy vậy, những lo ngại này là vô căn cứ vì con người có thể tạo ra không chỉ một cỗ máy thông minh mà còn cả một cỗ máy có ý thức và tính cách riêng. Tất nhiên, không thể loại trừ khả năng công nghệ này có thể được sử dụng một cách nguy hiểm và có thể dẫn đến thảm họa, như trường hợp năng lượng hạt nhân. Có rất nhiều cuộc thảo luận về các mối đe dọa phát sinh từ việc sử dụng AI như: vấn đề đảm bảo quyền riêng tư, gián điệp công nghiệp và phá hoại máy tính,...

Tóm lại: Công nghệ AI đã mang lại thành công lớn trong nhiều ngành công nghiệp khác nhau. Năm 2016, thị trường toàn cầu của AI đạt trị giá 4 tỷ USD nhưng dự đoán sẽ lên tới 169 tỷ USD vào năm 2025 và 15.700 tỷ USD vào năm 2035. Với xu thế phát triển công nghệ và ứng dụng trong đời sống xã hội đang thay đổi không ngừng, công nghệ AI đang là điểm đến của các nhà khoa học trong tương lai.

#### 4. KẾT LUẬN

Bài báo trình bày về quá trình phát triển, thời cơ và thách thức của trí tuệ nhân tạo, một trong những hướng phát triển thú vị nhất của công nghệ thông tin, thu hút rất nhiều tâm huyết của con người. Sự đầu tư nghiên cứu trí tuệ nhân tạo (AI) của nhiều cơ quan nghiên cứu trên thế giới mang lại nhiều thành tựu và có những ứng dụng thiết thực trong nhiều lĩnh vực kinh tế - xã hội và an ninh - quốc phòng. Viễn cảnh tạo ra những cỗ máy thông minh có thể suy nghĩ và đưa ra quyết định một cách độc lập khiến con người lo lắng. Tuy nhiên, những lo ngại về tầm quan trọng của con người bị giảm sút là không có cơ sở vì con người không chỉ tạo ra cỗ máy thông minh, mà còn tạo ra cỗ máy có ý thức và cá tính riêng.

#### TÀI LIỆU THAM KHẢO

- [1]. <http://pl.wikipedia.org>.
- [2]. [https://vi.wikipedia.org/wiki/Tri\\_tue\\_nhan\\_tao](https://vi.wikipedia.org/wiki/Tri_tue_nhan_tao).
- [3]. <http://www.systransoft.com/>
- [4]. L. Hou *et al.*, “Statistical neural network (SNN) for predicting signal-to-noise ratio (SNR) from static parameters and its validation in 16-bit, 125-MSPS analog-to-digital converters (ADCs)”. *Rev. Sci. Instrum.* 93 (8): 084701, (2022).
- [5]. R. Kurzweil, *The Age of Spiritual Machines*, Books Penguin, (2009).
- [6]. James, “ChatGPT creator OpenAI is in talks to sell shares in a tender offer that would double the startup's valuation to \$29 billion Insider”, Lakshmi Varanasi, (2023).
- [7]. OpenAI, “ChatGPT: Optimizing Language Models for Dialogue”, (2022).
- [8]. Clark, Jack “Why 2015 Was a Breakthrough Year in Artificial Intelligence”. *Bloomberg News*, (2015).
- [9]. S. Osowski, “Sieci neuronowe w ujęciu algorytmicznym”, Warszawa, (2016).

#### ABSTRACT

##### Artificial Intelligence development, opportunities and challenges

*The article presents the development process, opportunities and challenges of artificial intelligence (AI), a field with many practical applications: Medicine, Management Economics, Education, Military, etc. The scope of artificial intelligence includes algorithms, heuristics, genetic algorithms, expert systems, artificial neural networks, fuzzy logic, etc. The article also analyzes the creation of intelligent machines that can think and make decisions independently and the arbitrary and dangerous use of this technology that can lead to disaster (as in the case of nuclear energy). However, concerns about the reduced importance of humans are unfounded because humans not only create intelligent machines but also create machines with their own consciousness and personality.*

**Keywords:** Artificial intelligence; Expert systems; Artificial neural networks; Fuzzy logic; Intelligence machine.