

	<b>VIETTEL AI RACE</b>	TD080
	<b>Trí Tuệ Nhân Tạo Tạo Sinh (Generative AI) Trong Thiết Kế Thuốc Và Phát Hiện Dược Phẩm Mới</b>	Lần ban hành: 1

## 1. Giới thiệu

Phát triển thuốc truyền thống thường kéo dài 10–15 năm và tiêu tốn hàng tỷ đô la. Quy trình bao gồm nghiên cứu cơ bản, xác định mục tiêu (target), thiết kế phân tử, thử nghiệm lâm sàng và phê duyệt. Quá trình này tốn kém do phải sàng lọc hàng triệu hợp chất để tìm ra vài phân tử tiềm năng.

Trí tuệ nhân tạo tạo sinh (Generative AI) – bao gồm các mô hình như Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs) và Diffusion Models – đã mở ra khả năng tạo ra các phân tử mới đáp ứng tiêu chí được lý mong muốn, rút ngắn đáng kể thời gian và chi phí.

## 2. Trí tuệ nhân tạo tạo sinh là gì?

*Trí tuệ nhân tạo tạo sinh* (Generative AI) là một lĩnh vực của trí tuệ nhân tạo tập trung vào việc tạo ra dữ liệu mới, dựa trên dữ liệu đã được học trước đó. Dữ liệu này có thể bao gồm văn bản, hình ảnh, âm thanh, video, hoặc các loại nội dung khác. Các mô hình Generative AI không chỉ nhận biết và phân loại thông tin, mà còn có khả năng sáng tạo, mô phỏng và tái hiện thông tin. Một số ví dụ phổ biến nhất của trí tuệ nhân tạo tạo sinh là ChatGPT tạo văn bản, DeepFake mô phỏng khuôn mặt người trong video. Các mô hình này thường sở hữu một số đặc điểm cơ bản là:

- **Khả năng sáng tạo:** Tạo ra nội dung mới, độc đáo dựa trên các mẫu đã học.
- **Tính thích ứng cao:** Có thể áp dụng vào nhiều lĩnh vực, từ sáng tạo nội dung (viết báo, tạo hình ảnh) đến nghiên cứu khoa học (tạo mô hình phân tử).
- **Sử dụng dữ liệu huấn luyện:** Chất lượng nội dung tạo ra phụ thuộc nhiều vào quy mô và độ đa dạng của tập dữ liệu đã học.

	<b>VIETTEL AI RACE</b>	TD080
	<b>Trí Tuệ Nhân Tạo Tạo Sinh (Generative AI) Trong Thiết Kế Thuốc Và Phát Hiện Dược Phẩm Mới</b>	Lần ban hành: 1

- Dựa trên mô hình học sâu (Deep Learning): Thường sử dụng các kiến trúc mạng nơ-ron tiên tiến như GANs, VAEs và Transformers.
- Hạn chế sự giải thích: Mô hình thường được xem là hộp đen, khó giải thích rõ ràng lý do tại sao nó tạo ra kết quả cụ thể.

### 3. Trí tuệ nhân tạo tạo sinh hoạt động như thế nào?

Việc triển khai giải pháp trí tuệ nhân tạo tạo sinh cho mọi trường hợp sử dụng yêu cầu phải có nhiều nỗ lực từ các nhà khoa học và nhà phát triển. Đó là vì GenAI được thực hiện nhờ các bộ dữ liệu lớn và các thuật toán trí tuệ nhân tạo được đào tạo phức tạp. Công nghệ này được xây dựng và triển khai cùng với trí tuệ nhân tạo và xử lý ngôn ngữ tự nhiên (NLP), cho phép trí tuệ nhân tạo xử lý và hiểu ngôn ngữ của con người. GenAI và NLP cùng nhau có thể hiểu được lời nhắc người dùng để tạo ra phản hồi thích hợp, cho dù là văn bản, video, hình ảnh hay âm thanh.

Các giải pháp trí tuệ nhân tạo tạo sinh sử dụng các hệ thống trí tuệ nhân tạo được gọi là các mô hình ngôn ngữ lớn (LLM), sử dụng mạng thần kinh sâu để xử lý và tạo văn bản. Chúng được đào tạo trên lượng dữ liệu khổng lồ, tìm ra sự tương đồng giữa các loại dữ liệu và thông tin để tạo ra và cung cấp các kết quả đầu ra mới, mạch lạc.

LLM dựa vào cấu trúc transformer để xử lý các chuỗi đầu vào một cách song song, giúp cải thiện hiệu năng và tốc độ so với mạng nơ-ron cổ điển. Quy trình đào tạo mô hình cũng tiếp nhận thông tin từ các nhà khoa học dữ liệu và các chuyên gia về chủ đề giúp hướng dẫn quá trình học tập của thuật toán và dẫn dắt nó đến kết quả chính xác hơn.

Để hiện thực hóa các giải pháp trí tuệ nhân tạo tạo sinh, doanh nghiệp có thể đào tạo các mô hình từ đầu hoặc chọn mô hình đã được đào tạo sẵn có thể điều chỉnh theo nhu cầu cụ thể. Ví dụ: có thể đào tạo thuật toán chatbot GenAI theo các thuộc tính cụ thể của cơ sở khách hàng và mô hình kinh doanh của tổ chức. Hoặc

	<b>VIETTEL AI RACE</b>	TD080
	<b>Trí Tuệ Nhân Tạo Tạo Sinh (Generative AI) Trong Thiết Kế Thuốc Và Phát Hiện Dược Phẩm Mới</b>	Lần ban hành: 1

một mô hình có mục đích tạo ra văn bản để tiếp thị nội dung có thể được chuyên môn hóa hoặc tinh chỉnh để tập trung vào một ngành và đối tượng cụ thể. Nhiều mô hình dành riêng cho từng lĩnh vực cũng đang nổi lên với tốc độ nhanh chóng. Các mô hình này được đào tạo trên các bộ dữ liệu nhỏ hơn, nhằm mục tiêu cụ thể hơn so với các mô hình lớn hơn. Kết quả mới đây cho thấy các mô hình nhỏ hơn này có thể tái tạo độ chính xác của các mô hình lớn hơn nếu được đào tạo dựa trên dữ liệu được chọn lọc cẩn thận.

Các nhà phát triển cũng có thể sử dụng chức năng [tao tăng cường truy xuất \(RAG\)](#) để bổ sung cho các mô hình được đào tạo trước với dữ liệu cập nhật, độc quyền và bí mật từ cơ sở dữ liệu vector trong quá trình suy luận. Điều này đơn giản hóa quy trình tùy chỉnh và cập nhật, cũng như cho phép gán nguồn thông tin được tạo ra.

Dưới đây là tổng quan tóm lược (mặc dù nâng cao về kỹ thuật) về ba loại mô hình trí tuệ nhân tạo tạo sinh phổ biến:

- Mạng đối nghịch tạo sinh (GAN):** Các mạng này liên quan đến hai mạng thần kinh, một bộ tạo và một bộ phân biệt, được đào tạo đồng thời. Bộ tạo tạo ra các phiên bản dữ liệu mới, trong khi bộ phân biệt đánh giá chúng so với một bộ dữ liệu thực. Mục tiêu của bộ tạo là tạo ra dữ liệu không thể phân biệt được với dữ liệu thực, trong khi mục tiêu của bộ phân biệt là phân biệt chính xác giữa hai dữ liệu này. Theo thời gian, bộ tạo được cải thiện, tạo ra dữ liệu ngày càng thực tế.
- Bộ tự mã hóa biến thể (VAE):** VAE kết hợp các nguyên tắc từ mạng nơ-ron và mô hình xác suất để tạo ra các phiên bản dữ liệu mới thông qua quy trình mã hóa và giải mã. Mô hình VAE bắt đầu bằng cách nén dữ liệu đầu vào thành một đại diện đơn giản hóa các đặc điểm của nó. Sau đó, nó giải mã phiên bản đơn giản hóa đó, cố gắng tái cấu trúc dữ liệu đầu vào về dạng ban đầu. Thông qua quy trình này, VAE học được các tính năng và

	<b>VIETTEL AI RACE</b>	TD080
	<b>Trí Tuệ Nhân Tạo Tạo Sinh (Generative AI) Trong Thiết Kế Thuốc Và Phát Hiện Dược Phẩm Mới</b>	Lần ban hành: 1

thông số cơ bản của dữ liệu, giúp nó tạo ra các kết quả đầu ra thực tế và có thể tùy chỉnh nhiều hơn.

- **Mô hình transformer:** Các mô hình như Generative Pre-trained Transformer (GPT) có thể tạo ra văn bản mạch lạc và liên quan đến bối cảnh. Các mô hình này, ban đầu được thiết kế dành cho các công việc xử lý ngôn ngữ tự nhiên, được đào tạo trên các bộ dữ liệu lớn và có thể hoàn thành các lời nhắc văn bản, dịch ngôn ngữ, trả lời các câu hỏi và thậm chí có thể tạo ra các văn bản sáng tạo.

Nhiều chiến lược khác nhau có thể được sử dụng trong quá trình tạo ra để cân bằng giữa tính sáng tạo và sự gắn kết trong kết quả đầu ra. Hiện có nhiều nghiên cứu nhằm mục đích làm cho các mô hình này minh bạch hơn, giảm thiên kiến và cải thiện độ chính xác thực tế. Hiện cũng đã có chuyển động hướng đến các mô hình có thể đồng thời làm việc với nhiều loại dữ liệu, chẳng hạn như văn bản, hình ảnh và âm thanh.

#### 4. Cách Generative AI hỗ trợ thiết kế thuốc

Generative AI có thể học từ cơ sở dữ liệu hóa học khổng lồ để hiểu mối quan hệ giữa cấu trúc phân tử và tính chất dược học.

- Mô hình tạo ra các phân tử mới có đặc điểm tối ưu: khả năng gắn kết với protein mục tiêu, tính ổn định, và ít độc tính.
- Sử dụng reinforcement learning để tối ưu dần các phân tử dựa trên phản hồi từ các mô phỏng sinh học.
- Kết hợp mô phỏng động lực học phân tử (molecular dynamics) và AI tạo sinh giúp dự đoán tương tác phân tử nhanh hơn.

#### 5. Ứng dụng thực tế

	<b>VIETTEL AI RACE</b>	TD080
	<b>Trí Tuệ Nhân Tạo Tạo Sinh (Generative AI) Trong Thiết Kế Thuốc Và Phát Hiện Dược Phẩm Mới</b>	Lần ban hành: 1

- Tìm hợp chất kháng virus mới: Generative AI đã được sử dụng để đề xuất phân tử ức chế protein của SARS-CoV-2 trong giai đoạn đầu dịch COVID-19.
- Kháng sinh thế hệ mới: Tìm ra hợp chất Halicin bằng mô hình học sâu, hiệu quả chống vi khuẩn kháng thuốc.
- Thuốc điều trị ung thư: AI gợi ý cấu trúc phân tử tương tác đặc hiệu với thụ thể tế bào ung thư, hỗ trợ thiết kế thuốc đích (targeted therapy).
- Tối ưu hóa thuốc sẵn có: Đề xuất biến thể cấu trúc để cải thiện dược động học hoặc giảm tác dụng phụ.

## 6. Quy trình tích hợp AI trong phát triển thuốc

- Xác định mục tiêu (Target Identification): AI phân tích dữ liệu gen và proteomics để tìm đích tác động.
- Sinh phân tử mới (Molecule Generation): Mô hình tạo sinh để xuất hàng nghìn cấu trúc phân tử.
- Đánh giá in silico: Mô phỏng tính chất hóa lý, độ hòa tan, tính độc.
- Tổng hợp và thử nghiệm: Các phân tử triển vọng được tổng hợp trong phòng thí nghiệm và thử nghiệm in vitro/in vivo.

## 7. Ưu điểm nổi bật

- Tốc độ vượt trội: Rút ngắn thời gian sàng lọc từ nhiều năm xuống vài tháng.
- Tiết kiệm chi phí: Giảm đáng kể chi phí nghiên cứu ban đầu.
- Khả năng khám phá không gian hóa học rộng: Khám phá hàng tỷ cấu trúc mà phương pháp truyền thống khó tiếp cận.
- Cá nhân hóa y học: Đề xuất thuốc phù hợp với đặc điểm gen của từng nhóm bệnh nhân.

	<b>VIETTEL AI RACE</b>	TD080
	<b>Trí Tuệ Nhân Tạo Tạo Sinh (Generative AI) Trong Thiết Kế Thuốc Và Phát Hiện Dược Phẩm Mới</b>	Lần ban hành: 1

## 8. Thách thức

- Độ tin cậy của mô hình: Phân tử do AI đề xuất cần xác thực thực nghiệm.
- Dữ liệu đào tạo: Cần cơ sở dữ liệu hóa học lớn, chuẩn hóa, chất lượng cao.
- Vấn đề pháp lý và đạo đức: Quy định về bằng sáng chế, quyền sở hữu trí tuệ và trách nhiệm khi thuốc thất bại.
- Khả năng giải thích: Khó hiểu tại sao mô hình chọn một cấu trúc cụ thể.

## 9. Xu hướng tương lai

- Kết hợp Generative AI và Quantum Computing: Tăng độ chính xác của mô phỏng lượng tử trong tương tác phân tử.
- Học liên kết (Federated Learning): Cho phép nhiều viện nghiên cứu cùng huấn luyện mô hình mà không chia sẻ dữ liệu bí mật.
- Tự động hóa phòng thí nghiệm (Lab Automation): Tích hợp robot tổng hợp hóa học với AI để khép kín quy trình thiết kế–thử nghiệm.
- AI đa mô thức (Multimodal AI): Kết hợp dữ liệu gen, hình ảnh tế bào, và văn bản y khoa để thiết kế thuốc đa mục tiêu.