

PlantGPT:

“TRỢ LÝ ẢO” VỀ DI TRUYỀN THỰC VẬT

“

Sự kết hợp giữa trí tuệ nhân tạo (AI) và khoa học thực vật đang mở ra trang mới cho nghiên cứu nông nghiệp, đặc biệt trong lĩnh vực phân tích chức năng gene. Một trong những bước tiến nổi bật của xu hướng này là sự ra đời của PlantGPT - hệ thống hỏi đáp thông minh dựa trên loài *Arabidopsis*¹, do các nhà khoa học thuộc Viện Hàn lâm Khoa học Trung Quốc, Đại học Nông nghiệp Hoa Nam và Đại học Thanh Hoa nghiên cứu phát triển. Công trình này đã được công bố trên *Tạp chí Advanced Science* vào tháng 5/2025².

”

Phương thức hoạt động của PlantGPT

PlantGPT khai thác sức mạnh của các mô hình ngôn ngữ lớn (LLMs) để cung cấp phân tích chuyên sâu và chính xác cho nghiên cứu chức năng gen ở thực vật. Khác với các cơ sở dữ liệu sinh học truyền thống vốn yêu cầu truy vấn kỹ thuật cụ thể, PlantGPT mang đến giao diện thân thiện bằng ngôn ngữ tự nhiên, đồng thời tích hợp dữ liệu omics đa tầng nhằm làm sáng tỏ các cơ chế điều hòa gen phức tạp.

PlantGPT được phát triển trên nền tảng mô hình Llama3-8B và được tinh chỉnh bằng bộ dữ liệu chuyên sâu, bao gồm hơn 60.000 bài báo nghiên cứu về thực vật, 13.993 kiểu hình của loài *Arabidopsis* và 23.323 chức năng gen. Hệ thống ứng dụng phương pháp tạo văn bản tăng cường truy xuất (retrieval-augmented generation - RAG), kết hợp với cơ sở dữ liệu Chroma, nhằm giảm thiểu các sai sót thường gặp ở các mô hình ngôn ngữ lớn thông thường. Cách tiếp cận này giúp đảm bảo chất lượng đầu ra cao, dựa trên kiến thức chuyên sâu, đặc biệt phù hợp với các nghiên cứu liên quan đến *Arabidopsis* - loài thực vật mô hình quan trọng trong nghiên cứu chức năng gen ở các cây trồng một hoặc hai lá mầm.

Nhờ tích hợp các kỹ thuật xử lý ngôn ngữ tự nhiên (NLP) hiện đại, PlantGPT có khả năng phân tích dữ liệu genomics với độ chính xác cao, nắm bắt được tính phức tạp, sự mơ hồ và các mối liên hệ dài hạn giữa các thông tin di truyền. Khác với các mô hình cũ như Bộ nhớ dài - ngắn hạn (LSTM) vốn bị giới hạn bởi bộ nhớ ngắn và cách xử lý tuần tự, PlantGPT sử dụng nền tảng mô hình ngôn ngữ lớn (LLM) để làm rõ mối quan hệ giữa gen và kiểu hình, từ đó hỗ trợ quá trình giải mã hệ gen hiệu quả hơn. Đáng chú ý, PlantGPT là nền tảng mã nguồn mở, có khả năng cập nhật liên tục và hoàn toàn miễn phí tại địa chỉ www.plantgpt.icu. Điều này giúp các nhà nghiên cứu trên toàn cầu dễ dàng tiếp cận và ứng dụng trong nhiều lĩnh vực, từ lai tạo giống cây trồng cho đến nghiên cứu sinh học tổng hợp. PlantGPT được xây dựng với ba mục tiêu cốt lõi:

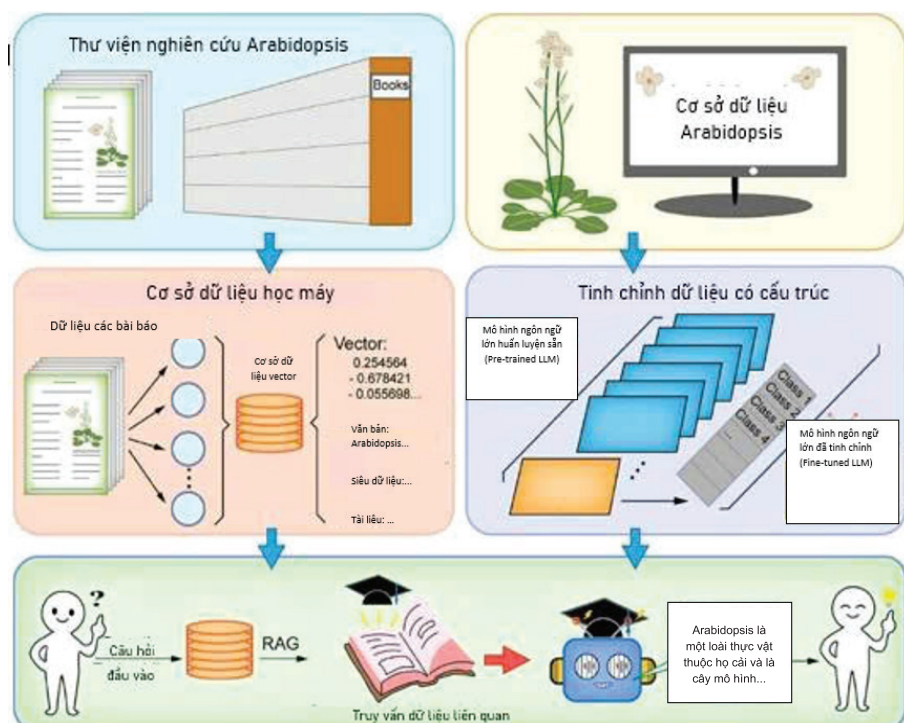
Phổ biến kiến thức: PlantGPT giúp kiến thức nông nghiệp và genomics dễ tiếp cận với công chúng, thu hẹp khoảng cách giữa khoa học phức tạp và người không chuyên.

Hỗ trợ nhà nghiên cứu trẻ: Bằng cách đơn giản hóa truy cập dữ liệu bộ gen thực vật, PlantGPT giúp các nhà khoa học trẻ khám phá và hiểu các cơ chế di truyền.

Hướng dẫn chiến lược cho chuyên gia: PlantGPT cung cấp gợi ý hữu ích cho các nhà nghiên cứu, hỗ trợ thiết kế thí nghiệm và chiến lược lai tạo.

¹Arabidopsis là chi thực vật có hoa trong họ cải và là loài thực vật có bộ gen đầu tiên được giải mã gen.

²R. Zhang, Y. Wang, W. Yang, et al. (2025), "PlantGPT: An Arabidopsis-based intelligent agent that answers questions about plant functional genomics", *Advanced Science*, DOI: 10.1002/advs.202503926.



Cơ chế hoạt động của PlantGPT. Nguồn: Advanced Science.

Không chỉ giới hạn trong nghiên cứu trên loài Arabidopsis, PlantGPT còn cho thấy tiềm năng ứng dụng rộng rãi trên nhiều cây trồng khác như lúa, lúa mì và ngô. Nhờ tính linh hoạt cao, nền tảng này được đánh giá là công cụ quan trọng trong nghiên cứu chức năng gen ở cây lương thực, góp phần giải quyết các thách thức cấp bách hiện nay như tăng năng suất, nâng cao khả năng chống chịu và thích ứng với biến đổi khí hậu.

Tiềm năng lớn để phát triển nông nghiệp bền vững

PlantGPT ra đời vào thời điểm then chốt, khi thế giới đang đối mặt với thách thức đảm bảo an ninh lương thực trong bối cảnh biến đổi khí hậu ngày càng gia tăng. Theo dự báo, dân số toàn cầu sẽ đạt gần 9,7 tỷ người vào năm 2050, kéo theo nhu cầu tăng 70% sản lượng lương thực, một mục tiêu khó đạt được nếu chỉ dựa vào các phương pháp lai tạo cây trồng truyền thống, vốn tốn nhiều thời gian và công sức.

Với khả năng đẩy nhanh quá trình khám phá gen và dự đoán chính xác đặc điểm kiểu hình, PlantGPT mang đến giải pháp công nghệ đột phá, góp phần phát triển các giống cây trồng có năng suất cao, khả năng chịu hạn tốt và thích nghi với điều kiện môi trường khắc nghiệt.

Không chỉ dừng lại ở nông nghiệp, tiềm năng ứng dụng PlantGPT trong lĩnh vực sinh học tổng hợp còn mở ra cơ hội thiết kế các mạch gen mới, giúp tối ưu hóa hiệu suất cây trồng. Trong bối cảnh biến đổi khí hậu ngày càng phức tạp, những công cụ công nghệ như PlantGPT được kỳ vọng sẽ đóng vai trò quan trọng trong việc xây dựng một nền nông nghiệp bền vững và thích ứng tốt với tương lai.

Dù có nhiều triển vọng, PlantGPT vẫn đối mặt với không ít thách thức. Các vấn đề đạo đức như thiên vị thuật toán và bảo vệ quyền riêng tư dữ liệu đang đặt ra yêu cầu về một khung pháp lý chặt chẽ, nhằm đảm bảo AI được sử dụng một cách minh bạch và có trách nhiệm. Bên cạnh đó, bộ gen thực vật vốn rất phức tạp với nhiều

cơ chế điều hòa đa dạng, đòi hỏi các mô hình AI phải liên tục được cải tiến. Để ứng phó, PlantGPT áp dụng cách tiếp cận mã nguồn mở và quy trình xác thực dữ liệu nghiêm ngặt, góp phần thiết lập tiêu chuẩn mới cho các công cụ AI trong lĩnh vực nông nghiệp.

*
* *
*

PlantGPT đánh dấu một bước ngoặt trong nghiên cứu chức năng gen thực vật, kết hợp sức mạnh của AI với kiến thức sinh học sâu rộng để giải quyết các thách thức nông nghiệp cấp bách. Thiết kế sáng tạo, khả năng tiếp cận mã nguồn mở và tính ứng dụng đa loài khiến nó trở thành nền tảng quan trọng cho nghiên cứu tương lai về cải thiện cây trồng và thích ứng khí hậu. Bằng cách tạo ra mã nguồn mở để truy cập kiến thức genomics và trao quyền cho các nhà nghiên cứu, PlantGPT không chỉ thúc đẩy khoa học mà còn góp phần đảm bảo an ninh lương thực toàn cầu. Khi AI tiếp tục phát triển, PlantGPT là hình mẫu cho việc tích hợp công nghệ với sinh học, mở đường cho một tương lai nông nghiệp bền vững.

Phạm Thịnh (theo *Advanced Science*)