

LLMs & NLP dành cho Tổ chức & Doanh nghiệp

# Giải pháp huấn luyện dữ liệu AI dùng riêng

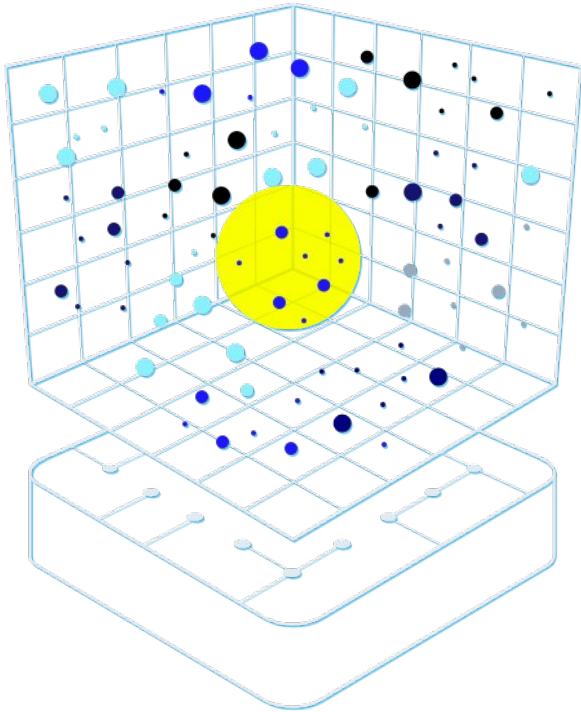


Trình bày: Tạ Quang Thái  
Tech advisor - myGPT

# Giải thích một số thuật ngữ

- **LLMs:** Mô hình ngôn ngữ lớn (*trong phạm vi này nó dùng mạng neuron - nơ ron để biểu diễn*)
- **Foundation Model:** Mô hình ngôn ngữ cơ bản (gốc)
- **OpenAI, Meta, Anthropic, Nvidia...:** Một số tổ chức giới thiệu và huấn luyện các mô hình ngôn ngữ gốc
- **NLP:** Xử lý ngôn ngữ tự nhiên
- **GPT:** Thuật toán sản sinh chữ dựa trên Mô hình ngôn ngữ lớn (LLM) được huấn luyện sẵn
- **GPT3, GPT4, Llama 2, Claude 2, LaMDA, PaLM, Mistral7B, PhoGPT...:** Là một số mô hình ngôn ngữ được huấn luyện sẵn với nhiều đóng gói với số lượng tham số khác nhau từ 3tỷ, 13 tỷ, 65 tỷ ... đến 175 tỷ tham số, 1000 tỷ, 2xxx tỷ. (tham số: là các nốt nơ ron)
- **ChatGPT:** Là một ứng dụng chat của OpenAI để chat với Mô hình GPT3, GPT4...
- **myGPT:** Là giải pháp của huấn luyện dữ liệu riêng cho các tổ chức, doanh nghiệp (*bao gồm cả ứng dụng, lưu trữ và huấn luyện dữ liệu riêng dựa trên kỹ thuật ánh xạ với GPT3-4-5, sau này GG, FB, hay Open Source có gì hay hơn được mình chứng thì mình có thể ứng dụng*)

# Công nghệ sử dụng của myGPT



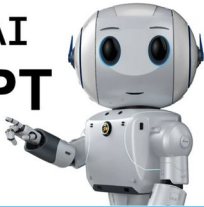
- **Data processing & back-end application:** R, Python, Node.js, Next.js...
- **NLP Framework:** NTKK, Bert, OpenNLP, TensorFlow, Keras, Theano, spaCy...
- **LLMs framework:** Langchain, Awaresome-LLM, Lamini, Llama\_index, CodiumAI, Auto-GPT...
- **Vectors Data store (notable):** Pinecone, Supabase, Elasticsearch, Keras, Faiss, Milvus...
- **Front application:** Next.js, React.js
- **Server:** Debiant 10 & up
- **Type:** OpenSource

# Huấn luyện dữ liệu riêng của myGPT

LLM Foundation Model

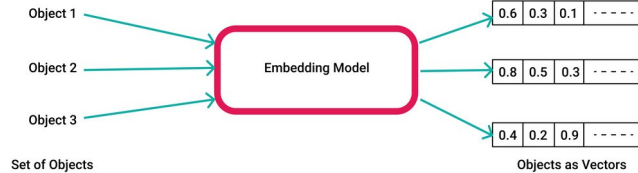


OpenAI  
**ChatGPT**  
GPT3.5  
Or Private LLM



APIs

RAG



NLP

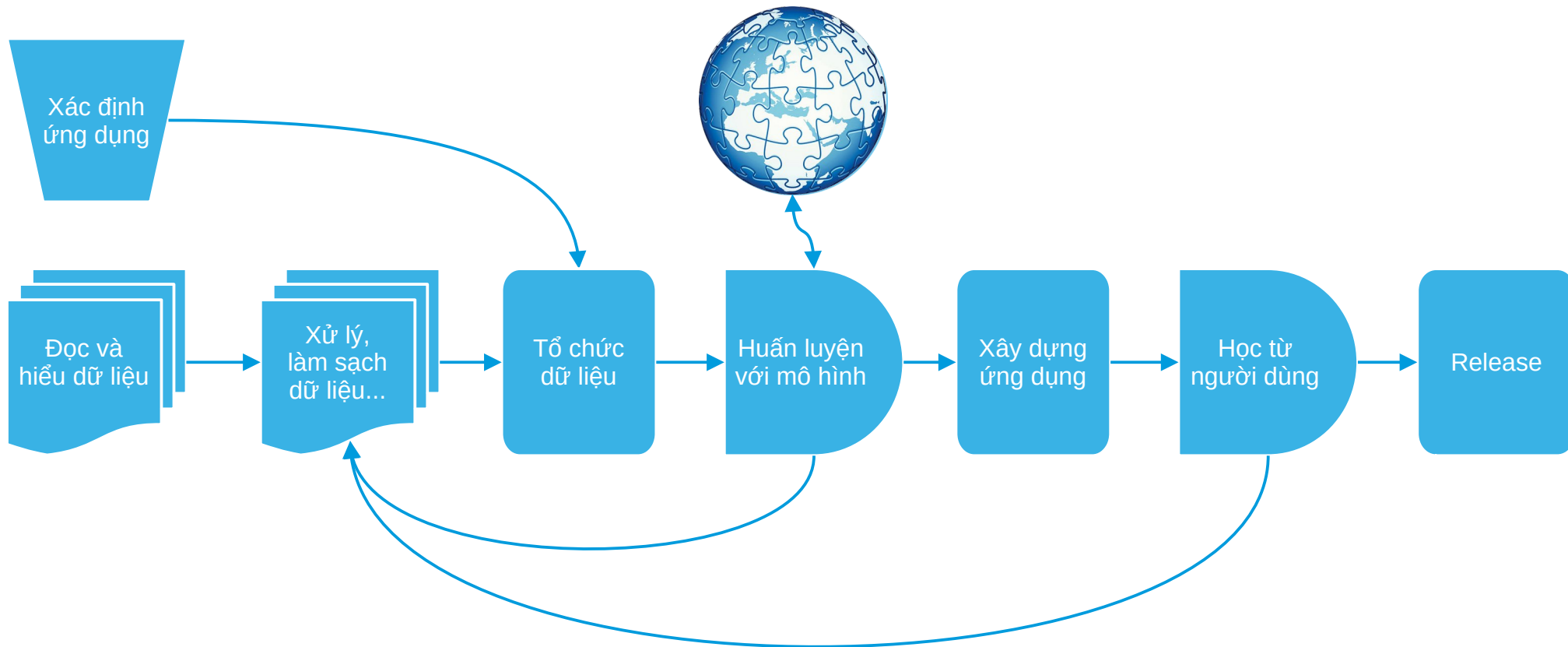


myGPT  
AI Data

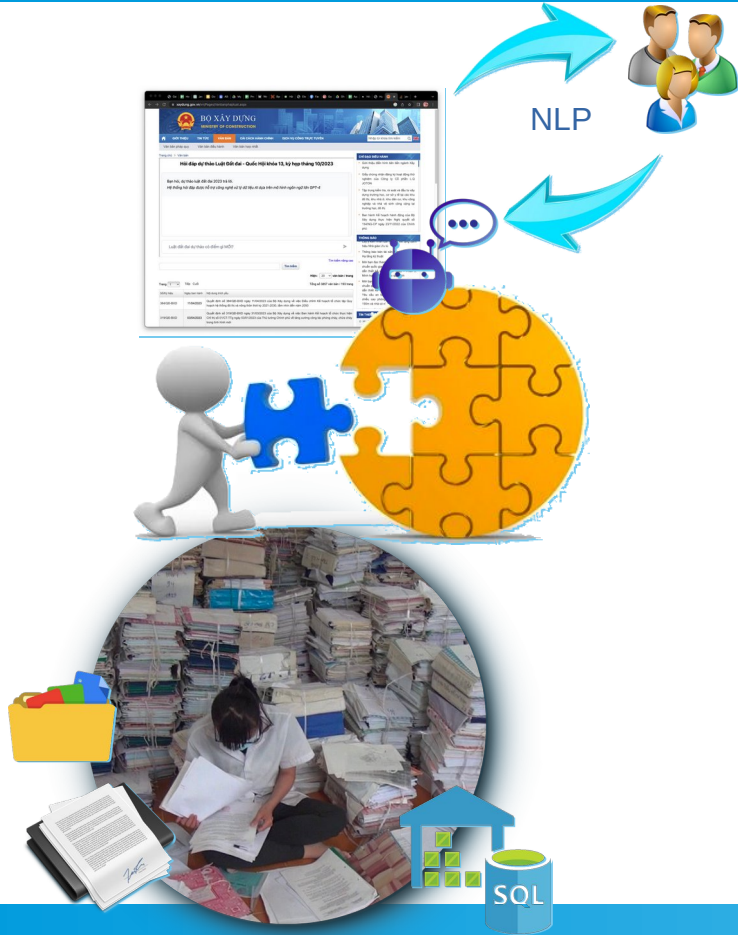


mygpt.vn

# Huấn luyện dữ liệu riêng của myGPT



# Ưu điểm của giải pháp



- Hiểu câu hỏi bằng ngôn ngữ tự nhiên
- Tìm kiếm nhanh trên mạng neuron với hàng triệu dữ liệu
- Sử dụng mô hình ngôn ngữ phù hợp với mục tiêu
- Dữ liệu lưu trữ riêng
- Đáp ứng nhiều cách thức tổ chức dữ liệu khác nhau
- Trả lời bằng ngôn ngữ tự nhiên, dễ hiểu
- Có khả năng tích hợp vào nhiều ứng dụng khác nhau

# Ưu điểm của giải pháp



## Giá thành phù hợp

### Tham khảo với huấn luyện mô hình gốc Llama 2:

- Năng lượng điện tiêu thụ: 13.244.246KWh
- Phát thải carbon: 539 tấn
- Máy tính sử dụng: 1 cụm siêu máy tính phục vụ nghiên cứu của Meta (không có thông số) và 1 cụm máy tính hội tụ dùng riêng. Cụm máy tính dùng riêng dùng Card đồ họa A100 của nVidia với khoảng 2000 GPU (giá GPU A100 khoảng 20-30k \$US/cái).
- Nhân sự: Không có công bố chính thức tuy nhiên Meta sử dụng rất nhiều team khác nhau và trong đó có team xác định rủi ro với quy mô là 350 người.



# Ứng dụng trong thực tế



- **Chatbot tư vấn, trợ lý:** ứng dụng trong các doanh nghiệp/ tổ chức để trả lời tự động người dùng thông tin về tổ chức, sản phẩm, giá cả hoặc các chiến dịch marketing, bán hàng, tuyển sinh...
- **Hỗ trợ nội bộ trong tổ chức:** Hỗ trợ đào tạo hội nhập, trợ lý kiến thức (KB), quy trình, nội quy, biểu mẫu...
- **Khai thác tài liệu lưu trữ:** Tri thức hoá các tài liệu lưu trữ để phục vụ khai thác kho dữ liệu, tàng thư...
- **Dịch vụ công:** Hỗ trợ trả lời thủ tục hành chính, dịch vụ công, hỏi đáp văn bản, chính sách, hướng dẫn hành chính...
- **Text to SQL, Function Calling:** Hỏi đáp với Database bằng ngôn ngữ tự nhiên (thay thế cho SQL), GraphQL, API
- **Tóm tắt, tổng hợp sách, báo, tạp chí, ấn phẩm:** Áp dụng trong thư viện, gợi ý tài liệu, ấn phẩm...



# Ứng dụng trong thực tế

lắp máy điều hoà 9000btu chạy 12m ống đồng tổng cộng bao nhiêu tiền

Công lắp máy 9.000btu: 150.000đ - 300.000đ Công chạy ống đồng và quấn bảo ôn 9.000btu: 140.000đ/m dài ống tiết diện dây Tổng cộng giá lắp đặt điều hoà 9000btu chạy 12m ống đồng là từ 1.830.000đ đến 1.980.000đ.

tôi muốn lắp trọn gói máy điều hoà 9000btu, ống đồng dài 8m thì hết bao nhiêu tiền

Giá thành lắp đặt máy điều hoà 9000btu là từ 150.000đ đến 300.000đ, công chạy ống đồng và quấn bảo ôn 9.000btu là 140.000đ/m dài. Vậy nếu lắp trọn gói máy điều hoà 9000btu với ống đồng dài 8m thì giá là:  
Giá lắp máy điều hoà 9000btu: 150.000đ - 300.000đ Giá công chạy ống đồng và quấn bảo ôn 9.000btu 140.000đ/m dài x 8m = 1.120.000đ  
Tổng giá trọn gói là từ 1.270.000đ đến 1.420.000đ.

2023-05-08T06:01:22.618Z: \* Tìm thấy 0: lắp đặt điều hoà, đơn vị tính, giá thành, ghi chú công lắp máy 9.000btu, bộ, 150.000đ - 300.000đ, công chạy ống đồng và quấn bảo ôn 9.000btu, m dài, 140.000đ, ống tiết diện dây công chạy ống đồng và quấn bảo ôn 9.000btu máy inverter, m dài, 150.000đ, ống tiết diện dây giá đỡ cục nóng 9000btu, bộ, 90.000đ - 110.000đ, công lắp máy 12.000btu, bộ, 150.000đ - 300.000đ,

## Hỏi Đáp - Nghị Định 53 - An ninh mạng

Cơ quan nào thẩm định hồ sơ hệ thống thông tin

Cơ quan thẩm định hồ sơ hệ thống thông tin để đưa vào Danh mục hệ thống thông tin quan trọng về an ninh quốc gia được quy định tại khoản 1, khoản 2, khoản 3. Bộ Công an có trách nhiệm thẩm định hồ sơ để nghị đưa hệ thống thông tin vào danh mục này nếu có đủ căn cứ về cấp độ an toàn thông tin. Hồ sơ đề nghị được gửi về Cục An ninh mạng và phòng, chống tội phạm sử dụng công nghệ cao thuộc Bộ Công an hoặc Bộ Tư lệnh Tác chiến không gian mạng thuộc Bộ Quốc phòng tùy theo loại hệ thống thông tin.

\*\*\*Chi tiết 5 điểm tham chiếu trong văn bản

### #Trang 4

Trường hợp cần xác thực thông tin trong hồ sơ và hiện trạng thực tế của hệ thống thông tin được nêu trong hồ sơ, cơ quan thẩm định quy định tại khoản 1, khoản 2, khoản 3 Điều này tổ chức khảo sát, kiểm tra thực tế để thẩm định đề nghị đưa hệ thống thông tin vào Danh mục hệ thống thông tin quan trọng về an ninh quốc gia

### #Trang 4

Thẩm định hồ sơ để nghị đưa hệ thống thông tin vào Danh mục hệ thống thông tin quan trọng về an ninh quốc gia 1

### #Trang 3

3 Trường hợp hệ thống thông tin trong quá trình thẩm định về cấp độ an toàn thông tin mà xét thấy có đủ căn cứ để đưa vào Danh mục hệ thống thông tin quan trọng về an ninh quốc gia, Bộ Thông tin và Truyền thông có trách nhiệm chuyển hồ sơ cho Bộ Công an để thẩm định hồ sơ để nghị đưa hệ thống thông tin vào Danh mục hệ thống thông tin quan trọng về an ninh quốc gia 4

### #Trang 4

thiết kế, xây dựng, quản lý vận hành; kiểm tra,

Danh mục hệ thống thông tin quan trọng về an

nh mục ứng dụng công nghệ cao thuộc Bộ Công an,

Bộ Quốc phòng đối với các hệ thống thông

tin quan trọng cơ yếu thuộc Ban Cơ yếu Chính

tr府 có trách nhiệm phân phối ý kiến bằng văn

bản để thẩm định hồ sơ để nghị đưa hệ thống thông

tin vào Danh mục hệ thống thông tin quan trọng về an

ninh quốc gia. Bộ Thông tin và Truyền thông có

trách nhiệm chuyển hồ sơ cho Bộ Công an để

thẩm định hồ sơ để nghị đưa hệ thống thông tin

vào Danh mục hệ thống thông tin quan trọng về an

ninh quốc gia. Bộ Thông tin và Truyền thông có

trách nhiệm chuyển hồ sơ cho Bộ Công an để

thẩm định hồ sơ để nghị đưa hệ thống thông tin

vào Danh mục hệ thống thông tin quan trọng về an

ninh quốc gia. Bộ Thông tin và Truyền thông có

trách nhiệm chuyển hồ sơ cho Bộ Công an để

thẩm định hồ sơ để nghị đưa hệ thống thông tin

vào Danh mục hệ thống thông tin quan trọng về an

ninh quốc gia. Bộ Thông tin và Truyền thông có

trách nhiệm chuyển hồ sơ cho Bộ Công an để

thẩm định hồ sơ để nghị đưa hệ thống thông tin

vào Danh mục hệ thống thông tin quan trọng về an

ninh quốc gia. Bộ Thông tin và Truyền thông có

trách nhiệm chuyển hồ sơ cho Bộ Công an để

thẩm định hồ sơ để nghị đưa hệ thống thông tin

vào Danh mục hệ thống thông tin quan trọng về an

ninh quốc gia. Bộ Thông tin và Truyền thông có

trách nhiệm chuyển hồ sơ cho Bộ Công an để

thẩm định hồ sơ để nghị đưa hệ thống thông tin

vào Danh mục hệ thống thông tin quan trọng về an

**BỘ XÂY DỰNG**  
MINISTRY OF CONSTRUCTION

GIỚI THIỆU TIN TỨC **VĂN BẢN** CÁC CÁCH HÀNH CHÍNH DỊCH VỤ CÔNG TRỰC TUYẾN

Văn bản pháp quy Văn bản điều hành Văn bản hợp nhất

Trang chủ > Văn bản

### Hỏi đáp dự thảo Luật Đất đai - Quốc Hội khóa 13, kỳ họp tháng 10/2023

Ban hỏi, dự thảo luật đất đai 2023 trả lời.  
Hệ thống hỏi đáp được hỗ trợ công nghệ xử lý dữ liệu AI dựa trên mô hình ngôn ngữ lớn GPT-4

Luật đất đai dự thảo có điểm gì MỚI?

Tìm kiếm

Tổng số 3857 văn bản / 193 trang

Hỏi: 20 | Văn bản / Trang

Số/Kj hỏi	Ngày ban hành	Nội dung trích yếu
384/QĐ-BXD	11/04/2023	Quyết định số 384/QĐ-BXD ngày 11/04/2023 của Bộ Xây dựng về việc Điều chỉnh Kế hoạch tổ chức tập Quay hoạch hệ thống đo thị và nâng thông tin kỹ 2021-2030, lần chỉnh sửa năm 2020
319/QĐ-BXD	03/04/2023	Quyết định số 319/QĐ-BXD ngày 31/03/2023 của Bộ Xây dựng về việc Ban hành Kế hoạch tổ chức thực hiện Chỉ thị số 01/CT-TTg ngày 03/01/2023 của Thủ tướng Chính phủ về tăng cường công tác phòng cháy, chữa cháy trong tình hình mới

Hoạt động xây dựng

Quy định về bảo vệ quyền riêng...

Gửi câu hỏi

## Collection: Đồng Tháp - TTCH

Phục vụ Hành chính công tỉnh Đồng Tháp (Số 85, đường Nguyễn Huệ, Phường 01, thành phố Cao Lãnh, tỉnh Đồng Tháp).

2. Nộp trực tuyến qua Cổng dịch vụ công tỉnh Đồng

**Nguồn**

Source (Doc id: c6dfde46-5d47-412d-b87d-86e72e4002e) Mã dịch vụ: BQLKKT\_DT21

Mã Dịch vụ Công Quốc gia: 1.009704.000.00...

Thủ tục để chính chủ dự án đầu tư trong trường hợp chia, tách, sáp nhập dự án đầu tư có mất phí không?

18:02:00

**Trả lời**

Không có mất phí khi thực hiện thủ tục để chính chủ dự án đầu tư trong trường hợp sử dụng quyền sử dụng đất, tài sản gắn liền với đất thuộc dự án đầu tư để góp vốn vào doanh nghiệp. Để thực hiện thủ tục này, bạn cần nộp 04 bộ hồ sơ bao gồm đất 0. Giấy chứng nhận quyền sử dụng đất, quyền sử hữu nhà và tài sản khác gắn liền với đất. Thời hạn xử lý là 50 ngày và không có phí dịch vụ. Bạn có thể nộp hồ sơ trực tiếp tại bộ phận Tiếp nhận và Trả kết quả tại Trung tâm Kiểm soát TTCH và Phục vụ hành chính công

**Nguồn**

Source (Doc id: 3506c36d-85c4-43c1-b7f4-90e14146646a) Mã dịch vụ: BQLKKT\_DT23

Mã Dịch vụ Công Quốc gia: 1.009766.000.00...

SEND



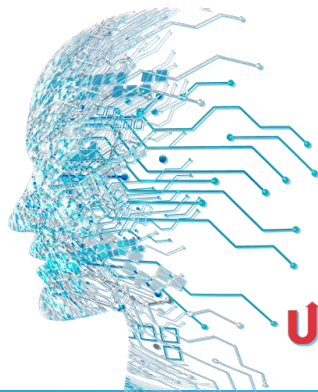
# Giá thành & Thời gian

Phân tích dữ liệu	Xử lý	Mã hoá	Fine-Tuning 2 w - 3 month	Ứng dụng	Tích hợp	Vận hành
<p>Đọc, hiểu, lên phương án tổ chức mô hình dữ liệu đáp ứng yêu cầu thực tế</p>	<p>Làm sạch dữ liệu, chuẩn hoá dữ liệu, phân hoạch, đánh nhãn và băm dữ liệu thành các đơn vị lưu trữ</p>	<p>Vector hoá các đơn vị dữ liệu, lưu trữ lên mạng neuron nội bộ</p>	<p>Kiểm tra, hiệu chỉnh, tối ưu hoá phản hồi, tiến hành xử lý và vector lại dữ liệu đến mức độ tối ưu theo yêu cầu</p>	<p>Hiệu chỉnh ứng dụng, phân tích &amp; nhận dạng câu hỏi, căn chỉnh kết quả search vector và phản hồi người dùng</p>	<p>Tích hợp vào các nền tảng đích (app, web hoặc các ứng dụng đặc thù)</p>	<p>Logs tracking</p>
<p><b>Tổ chức nhà nước</b> 300đ/15 ký tự hoặc tùy thuộc độ phức tạp/vềnh dữ liệu với mô hình</p>		<p>APIs pay</p>		<p>355tr-2tđ</p>	<p>Theo thực tế</p>	<p>APIs pay &amp; host</p>
<p><b>Doanh nghiệp</b> (dữ liệu thuộc nhóm myGPT đã trained)</p>		<p>5-30 days</p>		<p>150-200tr</p>	<p>Embedding code</p>	<p>APIs pay &amp; host</p>

# Demo - Q&A



- Công ty & Sản phẩm
- Dataset
- Văn bản pháp quy
- Thủ tục hành chính



Up next



**HiNET**  
Trust to Success

