

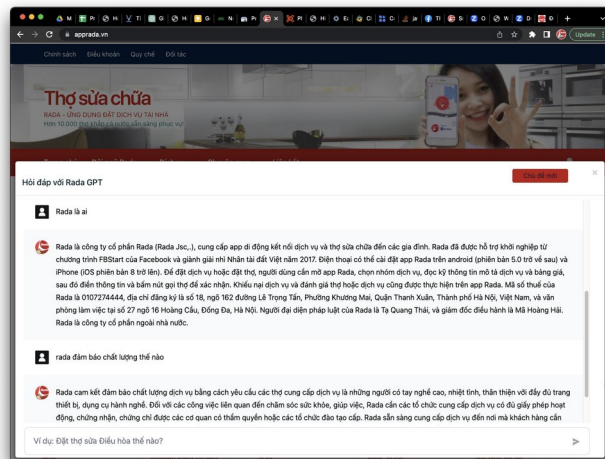
# myGPT: Giải pháp huấn luyện dữ liệu dựa trên mô hình ngôn ngữ lớn

## Thông tin về giải pháp

**Ứng dụng:** Triển khai ChatBot GPT phục vụ tư vấn và chăm sóc khách hàng; ứng dụng đào tạo nội bộ; ứng dụng dịch vụ công hoặc trợ lý hoạt động trong tổ chức.

**Nơi ứng dụng:** Website hoặc ứng dụng dùng riêng của các Tổ chức

**Mục đích:** Thay thế con người để trả lời tự động tất cả các câu hỏi của khách hàng hoặc người dùng nội bộ dưới dạng Chat tìm hiểu về Công ty, sản phẩm, dịch vụ hoặc các kiến thức được cung cấp trước khi quyết định mua, sử dụng sản phẩm hoặc tác nghiệp.



## Ưu điểm của giải pháp:

- Chỉ Sử dụng dữ liệu riêng của Đơn vị hoặc tổ chức ứng dụng để tương tác với khách hàng hoặc người dùng sản phẩm tìm hiểu thông tin, không trả lời các kiến thức bên ngoài khác.
- Ứng dụng mô hình ngôn ngữ phổ biến (LLM) để nói chuyện (chat) như người tư vấn với ngôn ngữ tiếng Việt giao tiếp tự nhiên, chuẩn mực.
- Có thể đào tạo, tái đào tạo, đào tạo bổ sung kiến thức của trường thường xuyên theo nhu cầu thực tế của các Tổ chức.
- Có thể dạy AI hiểu với hàng chục ngàn đến hàng trăm ngàn tài liệu liên quan đến Tổ chức trong thời gian ngắn và giá thành hợp lý.
- Hiểu biết ngữ cảnh nói chuyện theo chủ đề giống như ChatGPT.
  - Có thể chỉ dẫn người dùng đến các trang thông tin chi tiết bằng hình ảnh, video hoặc các trang gửi đăng ký đặt hàng, chốt đơn hoặc các tài liệu chi tiết.

## Chuẩn bị dữ liệu đầu vào:

- Tài liệu chứa thông tin tiếng Việt dạng văn bản text không có format.
- Tài liệu chứa được chia thành các khối thông tin, mỗi khối có tối đa từ 2.000 ký tự hoặc nhỏ hơn tùy dạng dữ liệu theo đặc tính của từng ứng dụng cụ thể.
- Yêu cầu viết chính xác, xúc tích đầy đủ thông tin liên quan đến sản phẩm, công ty, không sử dụng các câu từ trả lời theo văn nói của tư vấn viên. Khuyến khích sử dụng từ ngữ thông dụng dễ hiểu, trong sáng mà người dùng hay nói về sản phẩm hoặc kiến thức.
- Tài liệu dạng bảng, các dạng dữ liệu có cấu trúc trên các cơ sở dữ liệu quan hệ hoặc các dạng file pdf

**Lưu ý:** Các tài liệu được biên soạn đầy đủ, chi tiết sẽ cho kết quả xử lý tốt và hiệu quả cao khi triển khai training chatbot.

# Giá thành triển khai

## Chi phí dịch vụ triển khai myGPT

Nội dung	Giá thành	Ghi chú
Chatbot trả lời thông tin hoặc tài liệu của tổ chức, sản phẩm và dịch vụ	150tr – 200tr	Trọn gói đã bao gồm việc xử lý dữ liệu giới thiệu Công ty/Tổ chức và thông tin của tối đa 200 sản phẩm/dịch vụ hoặc 50 trang tài liệu.
Chi phí triển khai ứng dụng tương tác người dùng (tích hợp vào các ứng dụng đang có)	355tr – 2ti	Bao gồm cả việc xây dựng cấu trúc dữ liệu tiền xử lý AI; Ứng dụng tương tác phản hồi người dùng; Cơ sở dữ liệu lưu trữ AI dùng riêng
Chi phí tích hợp vào các ứng dụng đang khai thác hiện có	Theo thực tế	Tính toán dựa trên Usecase
Chi phí xử lý & nhập dữ liệu huấn luyện AI	300 đồng/15 ký tự (tổ chức nhà nước)  Trường hợp dữ liệu kỹ thuật chứa công thức, logic tính toán có độ vênh lớn với mô hình ngôn ngữ sẽ tính riêng theo thực tế.	Áp dụng quy định mức chi phí Nhập dữ liệu có cấu trúc tại Điểm 1 - Điều 4 của Thông tư số 164/2012/TT-BTC hướng dẫn mức chi tạo lập thông tin điện tử nhằm duy trì hoạt động thường xuyên của cơ quan, đơn vị sử dụng ngân sách nhà nước. Áp dụng trong trường hợp này là nhập dữ liệu từ cấu trúc hiện tại (dữ liệu đang lưu trên CSDL, file hoặc các hệ thống lưu trữ điện tử khác), sang một cấu trúc khác (dữ liệu chuẩn bị để AI xử lý). <b>Lưu ý:</b> Đối với các format dữ liệu đầu vào khác thì có thể phải áp dụng các công việc như số hoá, nhận dạng, đánh máy, chỉnh lý... được quy định ở thông tư 164 này
Chi phí hosting máy chủ riêng ứng dụng Chatbot	1.2tr – 4.6tr/tháng	- Không GPU
Máy chủ riêng có hoặc không có GPU	Theo thực tế	Tùy thuộc vào nhu cầu phục vụ của khách hàng sẽ tính riêng

## Các chi phí khách hàng phải trả khác trong quá trình hoạt động

- **Chi phí tài khoản ChatGPT 4.0:** 20\$/tháng (trả trực tiếp cho OpenAI)
- **Chi phí Tokens API GPT dùng cho huấn luyện** (sử dụng model Ada, Davinci... tùy thuộc vào dữ liệu và yêu cầu ứng dụng cụ thể để quyết định):
  - **Ada:** 0.0004\$/ 1k tokens
  - **Babbage:** 0.0006\$/ 1k tokens
  - **Curie:** 0.003\$/ 1k tokens
  - **Davinci:** 0.03\$/ 1k tokens
  - **Lưu ý:** Chi phí này sẽ trả nhiều lần trong quá trình huấn luyện và điều chỉnh nhằm thu được kết quả phù hợp. Phí này chỉ phải trả cho lần huấn luyện tiếp theo. Thông thường để đạt được một mô hình hoàn thiện sẽ phải training và hiệu chỉnh trong khoảng từ 5 đến 50 lần tùy độ phức tạp của dữ liệu. Đối với một model thông dụng với tổng mức xử lý khoảng 1 triệu từ thì chi phí này được đo đếm giao động trong khoảng từ 5.000.000 đến 20.000.000đ
- **Phí Tokens chat gpt-3.5-turbo:** 0.002\$/1K tokens chat

- 1k token khoảng 750 từ, quân bình 1 người hỏi đáp thoại và trả lời đơn giản sẽ trong khoảng từ 200-300 từ cho 1 câu thoại (hỏi và trả lời, bao gồm cả prompt hướng dẫn của hệ thống), 1k tokens sẽ đáp ứng khoảng 3-4 câu thoại → 20\$ sẽ đáp ứng được khoảng 30.000 – 40.000 đoạn hội thoại.
- Phí này trừ trong quá trình người dùng chat với hệ thống. Khi nào hết cần phải nạp thêm để sử dụng. Có thể nạp trước.

*Ghi chú: Các chi phí này OpenAI sẽ trừ trực tiếp trên tài khoản thẻ Visa của tổ chức, đơn vị triển khai không thu.*

### **Thời gian & các bước triển khai:**

- **Triển khai phần ứng dụng:** 1 tuần - 2 tháng ngày sau khi nhận máy chủ
- **Chính lý và hiệu chỉnh tài liệu:** Tùy thuộc vào độ phức tạp của tài liệu, đơn vị thực hiện hiệu chỉnh dưới sự hướng dẫn của bên triển khai.
- **Mô phỏng dữ liệu:** Tùy thuộc vào khối lượng dữ liệu
- **Phân tích, trích xuất dữ liệu:** Tùy thuộc vào thực tế dữ liệu
- **Mã hoá dữ liệu theo mô hình:** 1 ngày cho khoảng 1 triệu từ
- **Kiểm tra & hiệu chỉnh mô hình:** Sau khi đưa dữ liệu đã huấn luyện lên máy chủ online tiến hành kiểm tra, hiệu chỉnh và train lại sẽ trong khoảng 1 tuần – 5 tháng tùy thuộc thực tế.

### **Xem dự án mẫu đã triển khai công cộng:**

- <https://apprada.vn> tư vấn về Công ty, app Rada, giá sửa chữa
- <https://homeid.asia> tư vấn Công ty, ứng dụng HomeID & kiến thức về nhà chung cư
- <https://rangdongstore.vn> tư vấn lựa chọn sản phẩm đèn
- <https://dauthau.net> tư vấn lựa chọn gói sản phẩm của Dauthau.INFO

## **Thông tin đơn vị triển khai giải pháp**

1. Công ty First software Solution
2. Công ty CP phát triển nguồn mở & dịch vụ FDS
3. Công ty CP Giải pháp Thông tin iWay
4. Công ty CP EcoIT
5. Cty CP giải pháp phần mềm GTO
6. Công ty TNHH SmartBooks
7. Công Ty CP Tư Vấn và Tích Hợp Công Nghệ D&L
8. Công ty CP phát triển nguồn mở Việt Nam (VINADES)
9. InnoDigi Technology Vietnam
10. Công ty CP Công nghệ BPO (BPOTech JSC)