

Running An Ai GPT Model Locally on Raspberry Pi With Web UI

G.ashwath¹, S.K.balachandar², Mrs.S.P.Audilne Benna³, Dr.Rajiniginath⁴

^{1,2}Dept of computer science and Engineering

³Asst.professor, Dept of computer science and Engineering

⁴professor, Dept of computer science and Engineering

^{1,2,3,4}Sri Mthukumaran Institute of Technology,Mangadu,Chennai-69

Abstract- This project aims to deploy an AI-based Generative Pre-trained Transformer (GPT) model locally on a Raspberry Pi, enabling on-device natural language processing capabilities. Leveraging advancements in edge computing, we propose a solution that integrates a GPT model into the Raspberry Pi environment, eliminating the need for constant internet connectivity and ensuring data privacy. The deployment includes a user-friendly web interface, allowing seamless interaction with the GPT model.

I. INTRODUCTION

Generative Pre-trained Transformer (GPT) models, pioneered by OpenAI, represent a significant breakthrough in natural language processing (NLP). These models excel at generating coherent and contextually relevant text based on input prompts, making them invaluable for a wide range of applications, including chatbots, text generation, and language translation.

This project focuses on the local deployment of an AI GPT model on the Raspberry Pi, coupled with a web-based user interface (UI) for seamless interaction. By bringing AI capabilities directly to the edge, we aim to empower users with enhanced privacy, reduced latency, and improved accessibility to AI-powered services.

II. LITERATURE SURVEY

M. Satyanarayanan et al. This comprehensive review provides insights into the convergence of edge computing and AI technologies, discussing the challenges, opportunities, and emerging trends. It explores the potential of deploying AI models on edge devices like the Raspberry Pi and the implications for various domains.

This paper presents recent advancements in Generative Pre-trained Transformer (GPT) models, include designated field on the website or application. If the OTP is valid, indicating that the user has access to the email associated with the account, the user is prompted to create a

new password. If the OTP is invalid or expired, the user is prompted to try again, ensuring the security of the process. Once a new password is created and confirmed, the user can log in using their new credentials.

III. DISADVANTAGE

Raspberry Pi devices have limited computational resources compared to high-end servers or cloud platforms. GPT models are computationally intensive, requiring significant processing power and memory. Running such models locally on a Raspberry Pi may result in slower inference speeds and reduced model capacity due to resource constraints.

IV. PROPOSED SYSTEM

The proposed system aims to enable the local deployment of AI GPT models on Raspberry Pi devices, coupled with a web-based user interface (UI) for seamless interaction. Leveraging the computational capabilities of the Raspberry Pi and advancements in AI optimization techniques, the system provides users with on-device natural language processing capabilities while ensuring data privacy and minimizing reliance on external servers.

Real-time Chat Functionality: Integrate a live chat feature that allows users to connect with lawyers instantly for real-time assistance. **Appointment Scheduling:** Provide users with the option to schedule consultations or appointments with lawyers directly through the chat interface.

V. SOFTWARE OVERVIEW

5.1 OVERVIEW OF SOFTWARE

Operating System: The Raspberry Pi typically runs on a Linux-based operating system such as Raspbian (now known as Raspberry Pi OS). This operating system provides the foundation for running software applications and managing hardware resources.

Raspberry Pi: Ensure you have a Raspberry Pi board with sufficient resources to run the AI model. A Raspberry Pi 4 with 4GB or 8GB of RAM would be suitable for most AI tasks.

Storage: You'll need sufficient storage space for the AI model and related files.
AI Framework: Choose an AI framework compatible with Raspberry Pi, such as TensorFlow Lite, PyTorch, or ONNX Runtime, depending on your AI model.

Web Framework: Select a web framework for building the user interface. Flask and Django are popular choices for Python-based projects.

5.2 EXISTING SYSTEM



OpenAI's Codex is a powerful AI model based on GPT technology that can be deployed locally on Raspberry Pi devices. Using Codex, developers can create custom applications with natural language understanding enabling local AI inference without relying on external servers. While Codex does not offer a ready-to-use web UI out of the box, developers can integrate it into web applications running on the Raspberry Pi to provide a user-friendly interface.

5.3 Registration for Lawyers to Offer Pro Bono Services:

Pro Bono Registration Portal: Create a dedicated registration portal for lawyers interested in offering pro bono services. **Verification Process:** Implement a verification process to ensure the eligibility of lawyers and validate their credentials before they can participate in pro bono cases. **Case Matching Algorithm:** Develop a system that matches pro bono cases with registered lawyers based on their expertise, availability, and location.

5.4 AI Reporting/Warn Police for Dangerous Crime Queries:

Sentiment Analysis: Integrate sentiment analysis to detect potentially dangerous or illegal queries from users. **Warning System:** Implement a warning system that flags and reports suspicious or alarming queries to the appropriate

authorities, such as law enforcement. **Emergency Hotline Integration:** Provide users with emergency contact information or hotlines for immediate assistance in case of dangerous situations.

5.5 Negative Prompt Handling by AI:

Contextual Understanding: Train the AI to recognize negative prompts or inquiries related to dangerous activities or crimes. **De-escalation Responses:** Provide pre-defined responses that aim to de-escalate the situation and discourage harmful behavior. **Referral to Support Services:** Offer resources and referrals to mental health services, legal aid organizations, or crisis hotlines for individuals in distress.

VI. RESULTS AND ANALYSIS

Start by setting up your Raspberry Pi with an operating system like Raspberry Pi OS (formerly Raspbian). Follow the official instructions for installation.

Ensure your Raspberry Pi is connected to the internet and accessible on your local network.

Deploy your web application on the Raspberry Pi and test it locally to ensure everything is working correctly.

VII. SUMMARY

Begin by setting up your Raspberry Pi with the necessary operating system and dependencies. Develop a web user interface using frameworks like Flask or Django, allowing users to interact with the GPT model through their browser. Deploy the web application on the Raspberry Pi, testing it locally and considering optimizations for production use, such as deploying behind a web server like

REFERENCES

- [1] Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models. ArXiv (Cornell University).
- [2] Ramuš Cvetkovič, Iva, and Marko Drobnjak. "As Above so Below: The Use of International Space Law as an Inspiration for Terrestrial AI Regulation to Maximize Harm Prevention." *Artificial Intelligence, Social Harms and Human Rights*. Cham: Springer International Publishing, 2023. 207-238.
- [3] Prentice, Catherine. "Demystify Artificial Intelligence." *Leveraging Emotional and Artificial Intelligence for*

Organisational Performance. Singapore: Springer Nature Singapore, 2023. 25-40.

- [4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 1-9.
- [5] Kurzweil, R. The singularity is near : when humans transcend biology. Penguin Books, 2005, 10.