

# Build your own LLM model for privacy

Tim Glauner

July 15, 2023

# Introduction

- ▶ There is much discussion about privacy of data when you integrate with GPT-4 directly
- ▶ So, we want to explore how easy and cost effective it is to build our own cloud based LLM where all data stays local to the developer

# Project Overview and Steps to Execute

- ▶ Identify pre-trained LLM
- ▶ Identify and utilize cost effective cloud environment like DigitalOcean with large storage and powerful graphics card
- ▶ Train model on large pdf files and validate power of local model

# Current Status and Areas to Explore

1. DigitalOcean seems not to have GPU infrastructure
2. What should be the architecture?
  - 2.1 Programming Language: Python
  - 2.2 User interface: Jupyter
  - 2.3 Repository: Github
  - 2.4 Instructions: <https://thomascherickal.medium.com/how-to-create-your-own-llm-model-2598615a039a>
  - 2.5 Cloud:
    - 2.5.1 General: [www.digitalocean.com](http://www.digitalocean.com)
    - 2.5.2 GPU: [www.runpod.io/about/gpu-cloud](http://www.runpod.io/about/gpu-cloud)
3. Can we use <https://www.runpod.io/about/gpu-cloud>?

# Conclusion

Thank you for your attention!