|  |  |  |  |
| --- | --- | --- | --- |
| As part of today’s internship work, I made various improvements to make my artificial intelligence system for detecting anomalies more advanced and efficient. In particular, I made significant progress in the areas of autoencoder architecture and related prompt engineering.  The newly developed autoencoder code was updated to work with more flexible data structures and optimized so that the model files produced after training could yield more accurate results. In addition, I redesigned my prompt structures to make the outputs from the LLM model more meaningful and accurate.  Following these developments, I focused on running the LLM model on a CUDA-supported GPU to improve the system’s response time and increase overall processing efficiency. During this process, I conducted various research on running LLM models with CUDA cores in a Linux environment and successfully adapted some code sections to run in parallel with the GPU.  However, since data retrieval and reading operations are still CPU-based, I was not able to achieve the expected speed improvement in these parts. Therefore, I am continuing to research appropriate methods and configurations for running LLM models more effectively with CUDA in a Linux environment. The goal is to run all components of the model on the GPU in order to minimize response time and further increase system performance. | | | |
| **Sayfa No** | **Çalışmanın** | | **KONTROL** |
|  | Konusu :.........................................  ......................................................... | Yapıldığı Tarih  ...../..../202.. | ......................................  ...................................... |