AMOS SS08 Project cloud-native-llm

# Template

Please use the following template for creating your project report.

|  |  |
| --- | --- |
| Project name | Amos2024ss08-cloud-native-llm |
| Project mission | To create a Large Language Model on Cloud Native Technologies our work was focused mainly on the following topics:  1. Automated Dataset Preparation: Develop an automated system that prepares structured datasets from CNCF landscape documentation.  2. LLM Fine-Tuning: Enhance a Large Language Model specifically tailored to the prepared dataset for precise and relevant responses.  3. Open-Source Contribution: Release and open-source the fine-tuned model and dataset preparation tools to the community, fostering collaboration and innovation. |
| Industry partner | Kubermatic |
| Team logo |  |
| Project summary | DeepCNCF is a specialized Large Language Model designed to provide context-aware answers to questions related to Cloud Native technologies. To create this model, we collected data from various sources, including documentation sites on Cloud Native technologies and Stack Overflow. This data was then processed into a Q&A format.  Using this curated dataset, we retrained the gemma2 model to improve its ability to answer questions specifically about Cloud Native technologies. Evaluations indicate that our best-trained model, with 9 billion parameters, significantly outperforms the original gemma2 model in this domain. |
| Project illustration |  |
| Team photo |  |
| Project repository | <https://github.com/amosproj/amos2024ss08-cloud-native-llm.git>  <https://huggingface.co/Kubermatic> |