

# Fine-Tuning LLMs For Proposal Writing

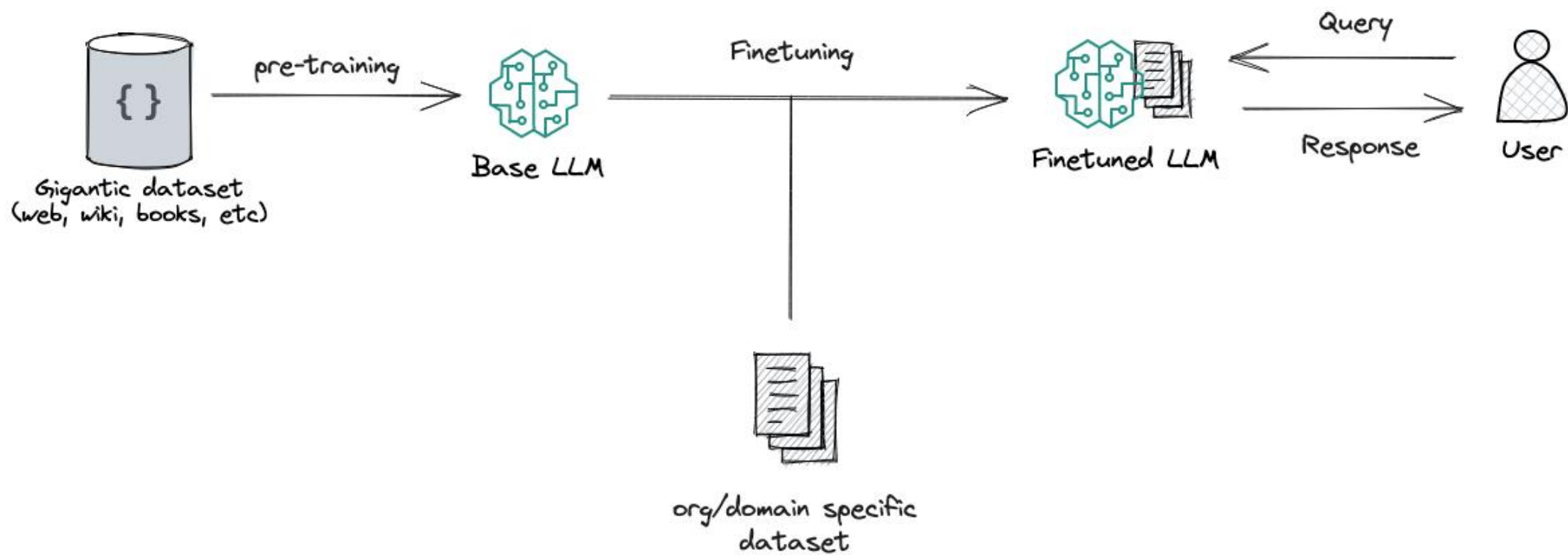
PPC group

21/6/2024

LLM Team

# Introduction

## Fine-tuning LLMs



# Why on-premises LLMs?

## Pros:

- **Data Privacy:** Ensures sensitive data remains on-premises
- **Customization:** Allows for specific fine-tuning to meet local needs.
- **Reduced Latency:** Faster response times without relying on external servers.
- **Cost Savings:** Eliminates recurring costs associated with cloud-based solutions.

## Cons:

- **Resource Intensive:** Requires significant computational resources.
- **Maintenance:** Needs regular updates and maintenance.
- **Scalability:** Limited by local hardware capabilities.
- **Complex Setup:** Initial setup can be complex and time-consuming.

# EUPCD Server Specifications

**GPU:** RTX 4060 Ti  
16GB

**CPU:** i5-14600KF  
3.5GHz

**RAM: 32GB**

**GPU VRAM:**

Crucial for efficient  
model fine-tuning  
and inference.

Models exceeding  
16GB VRAM fall back  
to slower CPU and  
RAM computation.

# Phi-3 Models

**Phi-3** is a family of open AI models developed by Microsoft.

## Parameter Sizes:

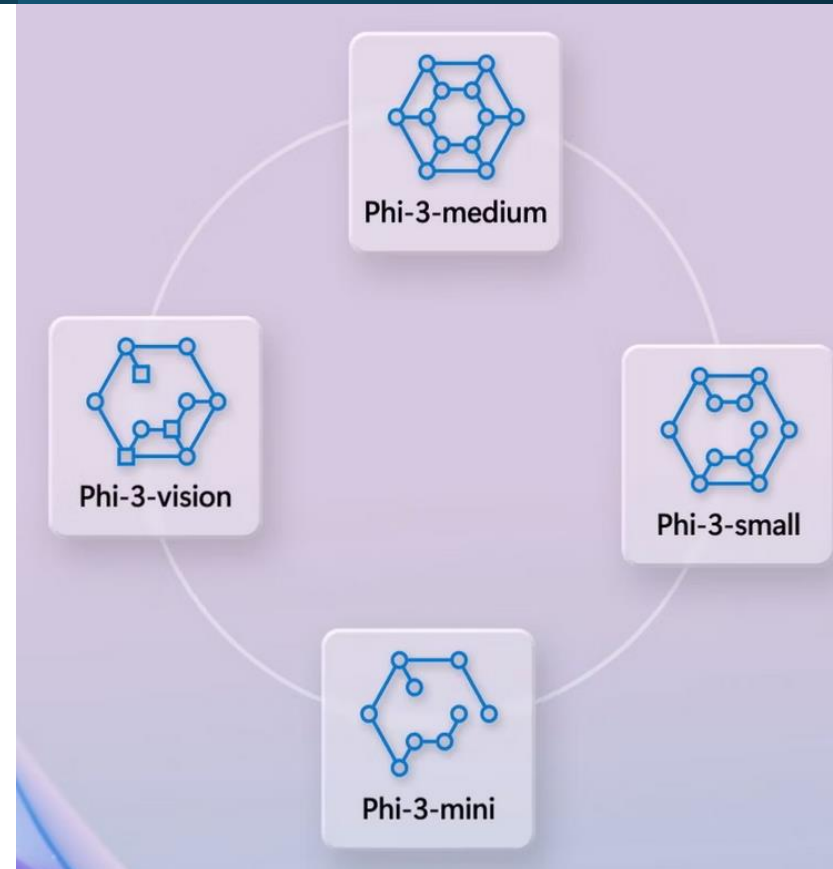
- **Phi-3 Mini:** 3B
- **Phi-3 Medium:** 14B

## Context Window Sizes:

- 4k
- 128k

## Performance:

**Phi-3** models outperform models of the same size and even the next size up in various benchmarks.



# Data Preprocessing

**Dataset:** Collection of 58 proposals from 2021, 2022, and 2023.



## Preprocessing:

**Cleaning:** Removed unnecessary headers and footers from the proposals.

**Sectioning:** Split the proposals into subsections.

**Saving:** Saved each subsection into separate files for easier processing and analysis.

# Data Preparation

Paired proposal's abstract with its subsections

Formatted the data into a template suitable for training the model.

## Template:

<b>System</b>	You are a proposal writer. According to the following abstract, you will write the text for the requested subsection:
<b>User</b>	Abstract: [abstract_text] Subsection: [subsection_name]
<b>Assistant</b>	[the proposal subsection]

# Proposal Template

## Excellence

- **1.1: Objectives and Ambition**
  - 1.1.1: Objectives
  - 1.1.2: Ambition
  - 1.1.3: State of the Art
  - 1.1.4: Relation to the Work Programme
- **1.2: Methodology**
  - 1.2.1: Overall Concept and Methodology
  - 1.2.2: Linked Research and Innovation Activities
  - 1.2.3: Activities Linked to the Project
  - 1.2.4: Interdisciplinary Approach and Compliance
  - 1.2.5: Gender Dimension
  - 1.2.6: Open Science
  - 1.2.7: Integration of Social Sciences and Humanities
  - 1.2.8: Research Data Management

## Impact

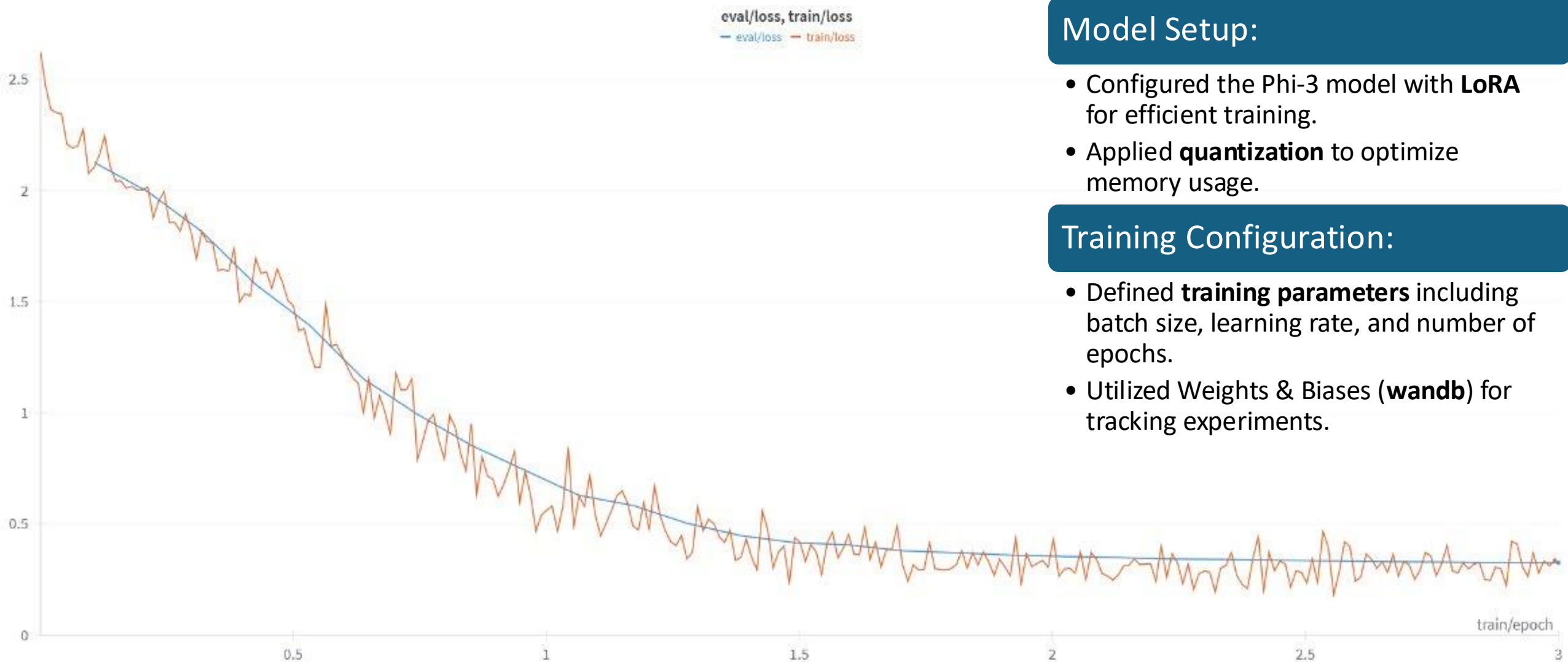
- **2.1: Project's Pathways Towards Impact**
  - 2.1.1: Contribution to Expected Outcomes and Impacts
  - 2.1.2: Requirements, Barriers, and Wider Impacts
  - 2.1.3: Scale and Significance
- **2.2: Measures to Maximize Impact: Dissemination, Exploitation, and Communication**
  - 2.2.1: Dissemination and Communication Strategy
  - 2.2.2: Exploitation
  - 2.2.3: Intellectual Property Management
- **2.3: Summary**

## Quality and Efficiency of the Implementation

- **3.1: Work Plan and Resources**
  - 3.1.1: Overall Work Plan Structure and Strategy
  - 3.1.2: Detailed Work Plan Description
  - 3.1.3: Work Packages, Deliverables, and Milestones
  - 3.1.4: Risks and Resources
  - 3.1.5: Critical Risks for Implementation
  - 3.1.6: Summary of Staff Effort
  - 3.1.7: Cost Justification
- **3.2: Capacity of Participants and Consortium as a Whole**
  - 3.2.1: Project Management Principles and Structure
  - 3.2.2: Complementarity Between Participants
  - 3.2.3: Resources to be Committed
  - 3.2.4: Communication amongst Partners



# Fine-Tuning Process



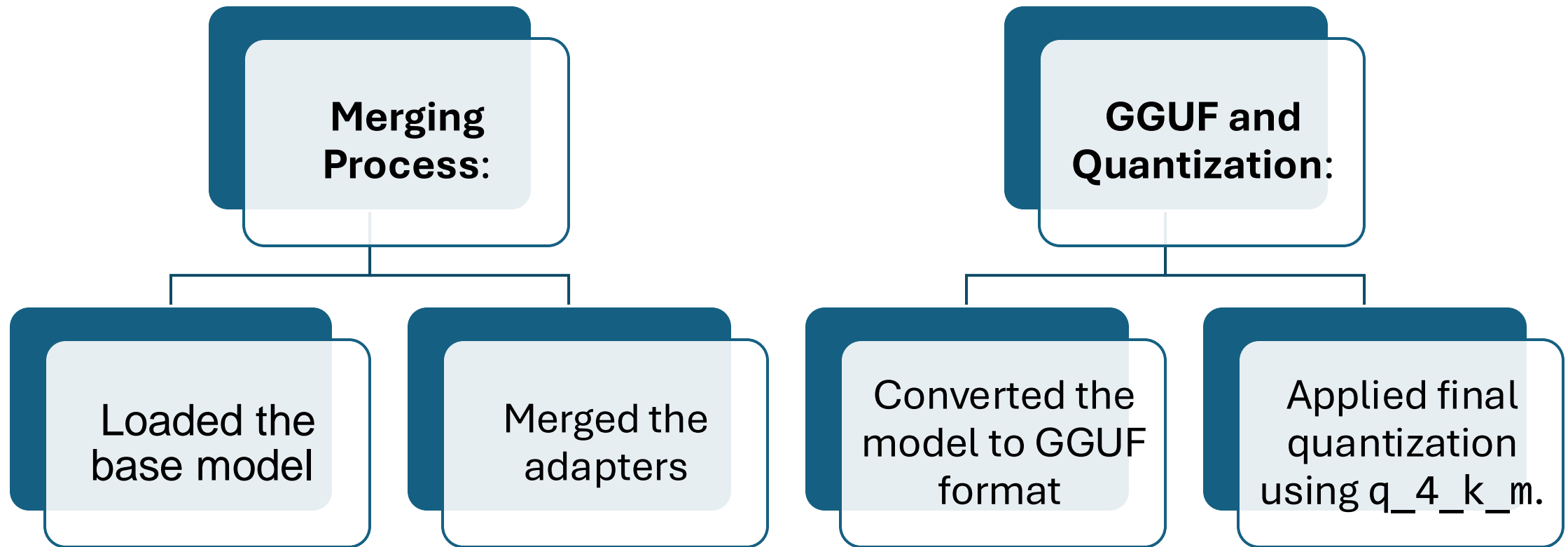
## Model Setup:

- Configured the Phi-3 model with **LoRA** for efficient training.
- Applied **quantization** to optimize memory usage.

## Training Configuration:

- Defined **training parameters** including batch size, learning rate, and number of epochs.
- Utilized Weights & Biases (**wandb**) for tracking experiments.

# Merge and Convert to Quantized GGUF



# Inference

## **Prompt:**

“You are a proposal writer for a consortium that consists of Public Power Corporation as the energy provider, TU Eindhoven as the stream processing experts, Athena Research Center and Uni Athens as AI, ApeiroPlus as risk assessment expert, Uni Trento as Earth Observation expert, and EuroControl as a use case provider.

**The topic of the proposal is the following:**

## **ExpectedOutcome:**

Projects’ results are expected to contribute to all of the following outcomes:

Critical infrastructure operators are more resilient to threats and natural and human-made hazards;

Improved monitoring, risk assessment, forecast, mitigation and modelling techniques aimed at increasing the resilience of critical infrastructures, validating multi-hazard scenarios, creating interactive hazard maps supported by Earth Observation and other data sources.

## **Scope:**

Under the Open Topic, proposals are welcome to address new, upcoming or unforeseen challenges and/or creative or disruptive solutions for increasing the resilience of critical infrastructure, that are not covered by the other topics of Horizon Europe Calls Resilient Infrastructure 2021-2022, Resilient Infrastructure 2023 and Resilient Infrastructure 2024.

Adapted to the nature, scope and type of proposed activities, proposals should convincingly explain how they will plan and/or carry out demonstration, testing or validation of developed tools and solutions. Proposals should also delineate the plans to develop possible future uptake and upscaling at local, national and EU level for possible next steps after the project.

In this topic the integration of the gender dimension (sex and gender analysis) in research and innovation content should be addressed only if relevant in relation to the objectives of the research effort.

Proposals should consider, build on if appropriate and not duplicate previous research, including but not limited to research by other Framework Programmes’ projects. When applicable, the successful proposal should build on the publicly available achievements and findings of related previous national or EU-funded projects.

Activities are expected to achieve TRL 6-8 by the end of the project – see General Annex B.

**Please provide the abstract of the proposal.”**

## Fine-tuned Model



## Pre-trained Model

The consortium aims to develop a holistic and integrated approach to enhancing the resilience of critical infrastructure through the development of advanced technologies, tools, and methodologies. The proposed solution is expected to address the challenges faced by critical infrastructure operators in the face of natural and human-made hazards, as well as the need for improved monitoring, risk assessment, forecasting, and mitigation techniques. The consortium's solution is expected to contribute to the following outcomes:

1. Critical infrastructure operators are more resilient to threats and natural and human-made hazards
2. Improved monitoring, risk assessment, forecast, and modelling techniques aimed at increasing the resilience of critical infrastructures
3. Validation of multi-hazard scenarios
4. Creation of interactive hazard maps supported by Earth Observation and other data sources

The consortium's solution is based on the integration of various technologies and methodologies, including stream processing, AI, Earth Observation, and risk assessment. The consortium's solution is expected to achieve the following results:

1. Improved real-time monitoring and risk assessment of critical infrastructure through the use of stream processing and AI
2. Enhanced forecasting and modelling of natural and human-made hazards through the integration of Earth Observation data
3. Development of a comprehensive risk assessment framework that considers various types of hazards and their potential impacts on critical infrastructure
4. Creation of interactive hazard maps that provide real-time information on the potential impacts of natural and human-made hazards on critical infrastructure
5. Demonstration of the solution's effectiveness through the use case provided by EuroControl

The consortium's solution is expected to contribute to the overall goal of increasing the resilience of critical infrastructure by providing critical infrastructure operators with advanced tools and methodologies to better prepare for, respond to, and recover from natural and human-made hazards. The proposed solution is expected to achieve TRL 6-8 by the end of the project, as outlined in General Annex B.

In an era marked by increasing threats to critical infrastructure from natural and human-made hazards, the proposed consortium project aims to revolutionize resilience strategies by leveraging the combined expertise of Public Power Corporation, TU Eindhoven, Athena Research Center, Uni Athens, ApeiroPlus, Uni Trento, and EuroControl. This innovative initiative seeks to develop and validate cutting-edge monitoring, risk assessment, forecasting, mitigation, and modelling techniques, thereby enhancing the resilience of critical infrastructures.

The project will focus on the integration of real-time data processing, advanced AI algorithms, and Earth Observation technologies to create a comprehensive, multi-hazard scenario modelling platform. By combining the strengths of each consortium member, the project will deliver a robust, scalable solution that addresses unforeseen challenges and disruptive solutions for critical infrastructure resilience.

TU Eindhoven will lead the development of a state-of-the-art stream processing framework, capable of handling diverse data sources, including Earth Observation data from Uni Trento and other relevant datasets. Athena Research Center and Uni Athens will contribute their AI expertise to develop intelligent algorithms for real-time hazard detection, risk assessment, and forecasting. ApeiroPlus will provide risk assessment expertise to validate the proposed solutions, ensuring their effectiveness in mitigating potential threats.

The project's innovative approach will result in a platform that supports interactive hazard maps, enabling critical infrastructure operators to make informed decisions based on real-time data and predictive modelling. The platform will also facilitate the development of future uptake and upscaling at local, national, and EU levels, paving the way for a more resilient and secure future.

In line with the Horizon Europe Resilient Infrastructure 2021-2022, Resilient Infrastructure 2023, and Resilient Infrastructure 2024 calls, this project will address novel challenges and disruptive solutions that are not covered by the existing topics. The consortium will ensure that the project's objectives align with the TRL 6-8 targets, as outlined in General Annex B.

The project will also incorporate a gender dimension, where applicable, to ensure that the research and innovation content is inclusive and addresses the needs of diverse populations. By building on previous research and achievements from other Framework Programmes and national projects, the consortium will deliver a comprehensive and impactful solution that contributes to the resilience of critical infrastructure.

# Next Steps

1

Fine-tuning with  
different parameters  
on LoRA

2

Exploring different  
quantization  
techniques

3

Adjusting various  
hyperparameters

4

Testing different  
models and sizes

5

Comparing fine-  
tuned models with  
base models across  
various benchmarks





Thank You