UNIVERSITY OF SARGODHA

DEPARTMENT OF COMPUTER SCIENCE

FACULTY OF COMPUTING & IT

Bachelor's Degree in Computer Science

Area: AI & ML

Physics-Informed LLM-Based Weather Prediction System

FYP Co-Ordinator:

Dr. Muhammad Ilyas

Supervisor:
Dr Fahad Maqbool

Candidates:

M Abubakar (BSCS51F22S026) (TL)

M Awais Ilyas (BSCS51F22S018)

M Zeeshan Asif (BSCS51F22S017)

BSCS SELF SUPPORT I

2022/2026

1 ABSTRACT

Traditional Numerical Weather Prediction (NWP) models resolve complex physical equations for atmospheric dynamics. However, it requires immense computational resources and often lack accuracy. On the other hand, data-driven Machine Learning (ML) and Large Language Models (LLMs) offer quicker prediction ignoring the physical consistency. This project introduces a hybrid framework that combines Physics-Informed Neural Networks (PINNs) with ML and LLMs to develop a physics-consistent and interpretable weather prediction system. PINNs embed governing equations like Navier–Stokes and thermodynamic laws into the learning process, also ensure that the forecasts must follow fundamental physics. The ML component improves feature learning from large-scale weather datasets. In contrast, the LLM interprets and translates model outputs into human-readable forecasts. The proposed framework aims to attain accurate, physically valid, and computationally effective short-term weather predictions. It also covers the gap between traditional physics-based models and modern AI-driven tools.

2 Background and Justification

2.1 Background

Weather forecasting has been one of the most crucial scientific challenge. There are traditional methods which rely on atmospheric observations like cloud patterns , wind direction and speed. But the main concern remains its accuracy which is every low and on large scale predictability becomes hard.

In 19th century when telegraph invented, Data started to transfer on long distances. It allowed us to make first weather map and storm warnings. However, these methods rely on highly factual data and shortage of mathematical accuracy.

With Numerical Weather Prediction (NWP) invention in the 20th century, which allowed models atmospheric behavior by computing physical equations like the Navier–Stokes, thermodynamic, and radiation equations across a 3D grid. NWP systems require supercomputers for accuracy, which are expensive, and often struggle with quicker forecasts.

Recent development in **Machine Learning (ML)** and **Artificial Intelligence (AI)** has transformed weather prediction. Models like **FourCastNet (NVIDIA, 2022)** and **GraphCast (Google DeepMind, 2023)** use deep learning to generate high-resolution forecasts faster and more efficiently than traditional NWP. However, ML models have some drawbacks such as:

- Limited understanding of physical laws (risking unrealistic outputs),
- Poor generalization in data-scarce regions, and
- Lack of interpretability.

To confront these challenges, **Physics-Informed Neural Networks (PINNs)** have emerged as a bridge between physics and ML. PINNs integrate fundamental physical equations into their training process directly. It also ensures that the predictions follow conservation laws and remain physically consistent.

Contrary to it , Large Language Models (LLMs) like GPT and LLaMA exhibit extraordinary proficiency to analyze, describe, and address complex data in natural language.In weather forecasting, they can refine complex outputs into clear reports , integrate data from text, numbers, and satellite imagery, and enhances user accessibility. In spite of the progress, no comprehensive tool currently integrates **PINNs**, **ML**, and **LLMs** for weather prediction. A **hybrid physics-informed LLM model** offers the potential to gain physically consistent, computationally efficient, and human-readable forecasts.

2.2 Justification

The proposed **Physics-Informed LLM Weather Prediction System** incorporates the strengths of physics-based and AI-driven methods to overcome current forecasting limitations.

1. **Physical Characteristics:**
   All the predictions follow Laws of Physics such as conservation of mass, energy, and momentum etc. And this is ensured by PINNs which prevents unrealistic results such as temperature exceeding more than it can exist is natural environment.
2. **Computational Efficiency:**
   Unlike NWP systems that depend on supercomputers, this hybrid tool can generate accurate short-term forecasts faster on moderate hardware.
3. **Accuracy in Sparse Data Regions:**
   Physics constraints assist PINNs infer missing values in under-observed areas, while LLMs help in contextual reasoning when data is limited.
4. **Understanding and Communication:**
   Complex outputs of model are converted into human-readable form by LLM module. This helps non-technical users such as farmers, policymakers etc. to easily access the system.
5. **Public Effects:**
   Proper and on time forecasts allow the disaster managers and other authorities to prepare in best possible way. This results in decreased economic losses, and strengthens the implementation of safety measures in vulnerable regions.

3 Methodology

For producing precise, human understandable and which follows physics laws, making a system which combines both Physics-Informed Neural Networks (PINNs) with Large Language Models (LLMs).

3.1 Data Collection and Preprocessing

Weather data will be obtained from **ERA5** and **WeatherBench** datasets, containing temperature, pressure, humidity, and wind information. Data will be normalized and reduced to a manageable grid (≈1°×1°), and split into training, validation, and test sets.
Tools: xarray, numpy, pandas.

3.2 PINN Module

PINNs incorporates physical laws (e.g., **Navier–Stokes** and **thermodynamic equations**) into neural network training. The model decreases both data error and physics residuals to make sure the realistic and physics-based predictions.

Tools: PyTorch, DeepXDE

3.3 LLM Module

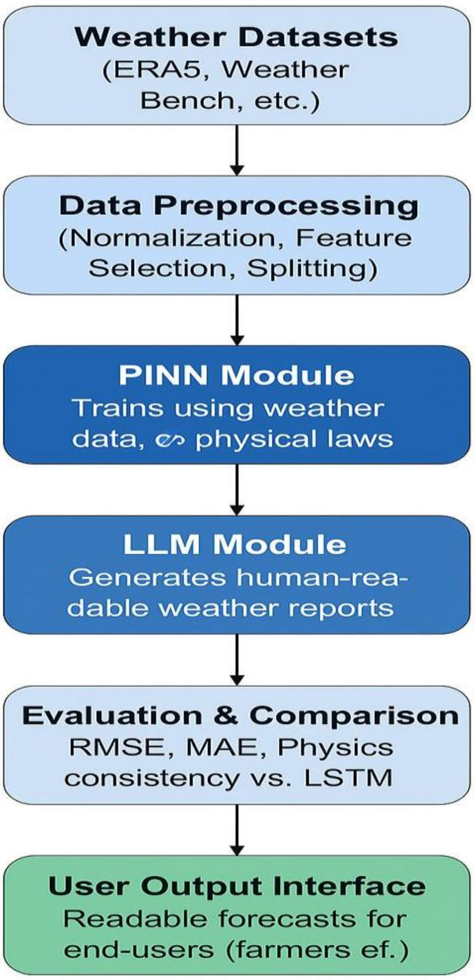
A pre-trained **LLM** (e.g., **GPT-2** or **LLaMA-2**) analyze PINN outputs and generates readable weather summaries for users.

Tools: Hugging Face Transformers

3.4 Hybrid Integration and Evaluation

Workflow: Weather data → PINN forecast → LLM explanation → User output.

Performance will be examined using **RMSE**, **MAE**, and **physics consistency**. Its performance is compared with baseline ML models (LSTM, CNN).



4 Project Scope

The goal of the proposed project is to develop a hybrid weather prediction system by fusing Large Language Models (LLMs) and Physics-Informed Neural Networks (PINNs). Below is a description of the scope:

4.1 In-Scope

- **Processing Data:** Utilize publicly accessible datasets like ERA5 and WeatherBench , and preprocess important variables for analysis, such as temperature, wind speed, air pressure, humidity, and rainfall.
- **PINN-Based Forecasting:** Using PINNs, create a short-term prediction model (for 1-3 days only) while making sure the outcomes adhere to accepted physical laws of the atmosphere
- **LLM Integration:** Pre-Trained LLMs such as GPT-2,LLaMA will be used to convert numerical outputs into human-readable forecasts and desired outputs.
- Hybrid System: PINN and LLM modules will be integrated or combined into one manner or method for providing expert and user outputs.
- **Analysis and Judgement:** Results will be integrated with ML baselines such as LSTM, CNN by using RMSE, MAE, and physical consistency checks.
- **Product or Deliverables:** A functional Prototype, visualized results and findings in form of final report.

4.2 Out of Scope

- High-resolution global NWP simulations.
- Long-term climate forecasting.
- Training LLMs from scratch (only pre-trained models used).
- Supercomputer-level computation.
- Deployment as a public or operational forecasting system.

5 High Level Project Plan

Phase	Duration	Activities
Phase 1: Research & Setup	Weeks 1–3	Literature review (NWP, ML, PINNs, LLMs), dataset collection, environment setup
Phase 2: Baseline Models	Weeks 4–6	Implement ML models (LSTM/Transformer) for comparison

Phase	Duration	Activities
Phase 3: PINN Development	Weeks 7–10	Train PINN with weather data and physics constraints
Phase 4: LLM Integration	Weeks 11–13	Use pre-trained LLM for forecast explanation
Phase 5: Hybrid Model Testing	Weeks 14–15	Combine PINN + LLM, evaluate vs. baselines
Phase 6: Documentation & Presentation	Week 16	Final report, demo, and presentation

6 References

[1] Richardson, L. F. (1922). *Weather Prediction by Numerical Process.* Cambridge University Press.

[2] Moreno Soto Á. et al. (2024). *Physics-informed neural networks for high-resolution weather reconstruction.* Open Research Europe

[3] PINNS for weather reconstruction

[4] Pathak, J., et al. (2022). *FourCastNet: Accelerating global high-resolution weather forecasting using adaptive Fourier neural operators.*

[5] Lam, R., et al. (2023). *GraphCast: Learning skillful medium-range global weather forecasting.* Google DeepMind.

[6] Karniadakis, G. E., et al. (2021). *Physics-informed machine learning.* Nature Reviews Physics.

[7] Hugging Face. (2024). *Transformers Library for LLMs.* https://huggingface.com