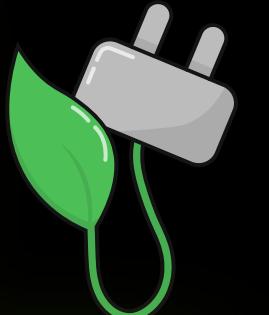


AGENTIC AI TO MEASURE AI'S CARBON FOOTPRINT IN BUSINESS ENVIRONMENTS



Name: Durva Dhanraj Patil

Institute: K.B.Joshi Institute of
Information Technology, Pune

E-mail: patildurva2003@gmail.com

Problem Statement:

Businesses run multiple AI/ML models but do not know how much energy or CO₂ each model consumes, leading to:

- higher cloud costs
- unnecessary GPU usage
- increased carbon emissions
- no visibility or optimization



Proposed Solution (High-Level)

Build an Agentic AI system that automatically tracks, analyses, and optimizes the carbon footprint of all AI workloads.

The solution has 3 major layers:

- A) Data Collection Layer (Measurement & Monitoring)
- B) Agentic Intelligence Layer (Analysis & Recommendations)
- C) Visualization & Reporting Layer (Dashboard + Alerts)

A) DATA COLLECTION LAYER (MEASUREMENT & MONITORING)

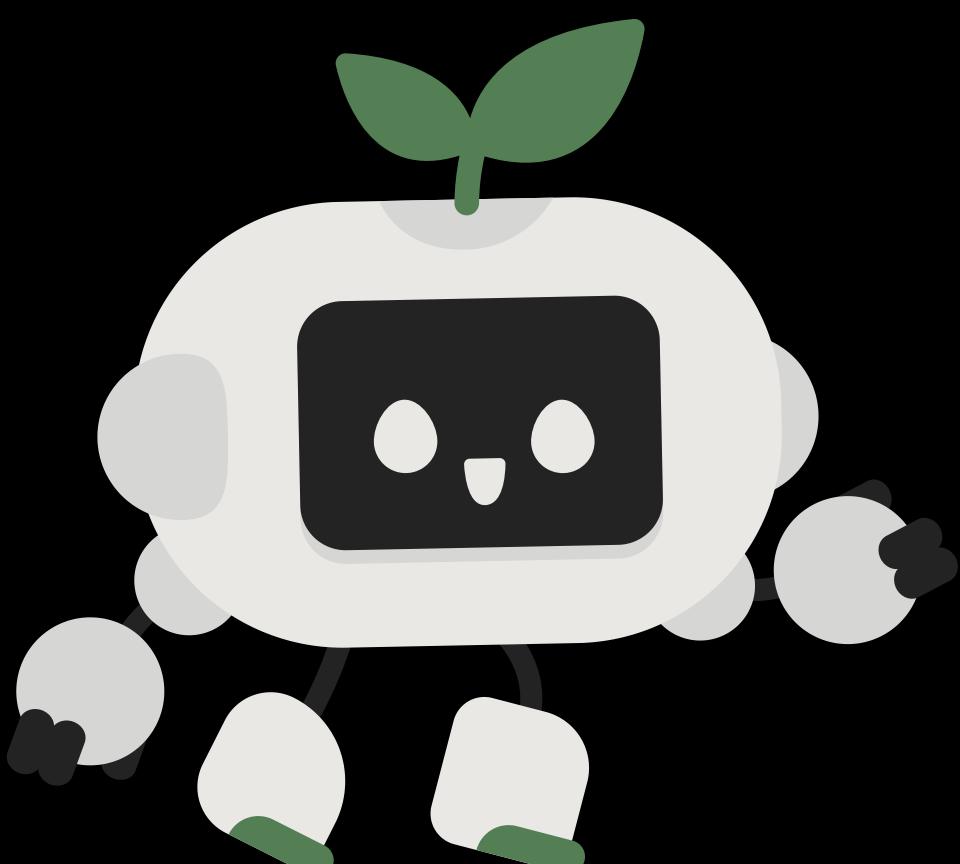
Goal: Measure real-time energy and CO₂ emissions per AI model/job.

How it works (Technical):

- **Wrap ML training/inference jobs using Python energy trackers**
 - CodeCarbon, psutil, nvidia-smi, CPU/GPU utilization APIs
- **Track:**
 - runtime
 - GPU/CPU usage
 - power consumption
 - kWh used
 - CO₂ emitted (kg CO₂e)
- **Export metrics using:**
 - Prometheus exporters
 - Docker containers for isolation
- **Store emissions in a lightweight database (SQLite or PostgreSQL)**

Outcome (Functional):

- Every ML model automatically logs its carbon footprint.
- CO₂ per model, per job, per hour becomes visible.



B) AGENTIC INTELLIGENCE LAYER (ANALYSIS & RECOMMENDATIONS)

Goal: Use an AI agent to analyze emissions and suggest optimization strategies.

How it works (Technical):

- **Use LangChain to build an agent that can query emission metrics.**
- **Agent reasons over:**
 - high-consumption jobs
 - idle GPU time
 - peak vs off-peak grid carbon intensity
- **Suggests optimizations such as:**
 - switch to mixed precision
 - reduce batch size
 - use efficient model architectures
 - schedule jobs to low-carbon hours
 - move workloads to greener cloud regions

- **Optional automation:**
 - scale down unused GPU nodes
 - schedule training at night
 - auto-stop long-running jobs
- **Outcome (Functional):**
 - Business gets actionable insights or auto-executed steps for greener AI usage.

C) VISUALIZATION & REPORTING LAYER (DASHBOARD + ALERTS)

Goal: Give business teams a simple UI to track and improve emissions.

How it works (Technical):

- Build dashboard using Streamlit or Dash.
- Display:
 - total CO₂ emissions
 - emissions per model/job
 - energy usage charts
 - cost vs carbon metrics
- Alerts using:
 - Prometheus Alert Manager
 - Email/SMS/Slack notifications
- Deploy entire system using Docker containers.
- Outcome (Functional):
- Business teams can see:
 - which models are expensive
 - which jobs waste energy
 - how to optimize deployments
- Alerts warn if a job is consuming too much power.



END-TO-END FLOW SUMMARY

Step-by-step flow:

1. *ML job starts*
2. *Energy tracker measures power, runtime, CPU/GPU usage*
3. *CO₂ emission calculation (kWh × Carbon Intensity)*
4. *Prometheus collects & stores metrics*
5. *Dashboard visualizes emissions per model/job*
6. *Agentic AI analyses patterns & recommends or auto-takes actions*
7. *Alerts notify high-emission events*
8. *Business achieves lower cost & lower carbon footprint*

WHAT THIS SYSTEM DELIVERS

Functional Benefits :

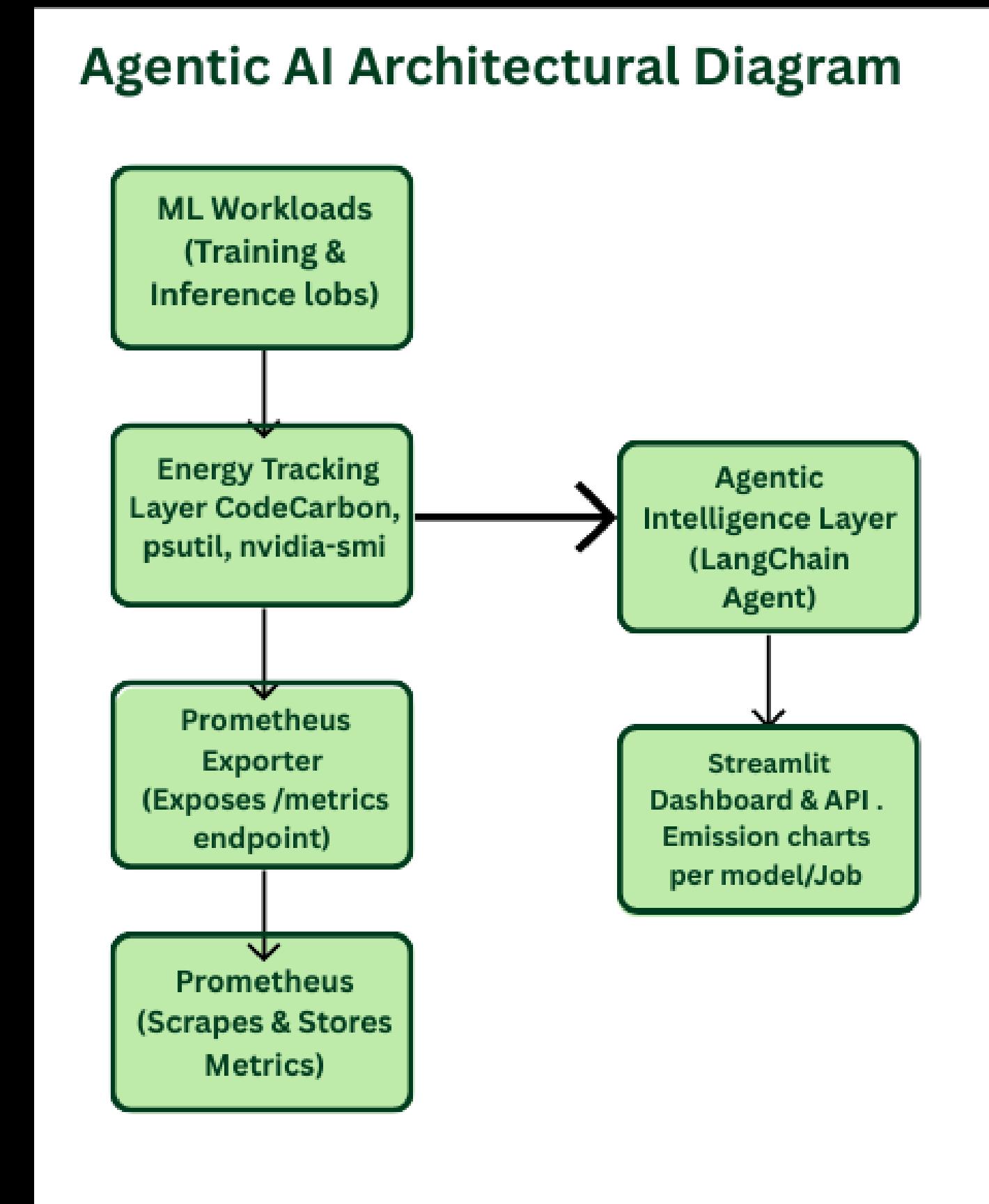
- *Real-time carbon tracking for all AI workloads*
- *Automatic insights to reduce cost and emissions*
- *Transparent reporting for sustainability teams*
- *Data-driven decision making for deployment and scaling*



Technical Strengths :

- *Python-based lightweight monitoring*
- *Prometheus metrics pipeline*
- *Docker for reproducibility*
- *Agentic AI powered by LangChain*
- *Fast dashboards with Streamlit*
- *Modular & expandable architecture*

A complete agentic AI system that monitors AI model emissions, analyzes energy usage, and offers intelligent optimization strategies — helping businesses run greener, cost-efficient AI.



Thank You

