
AI Forecasting Hackathon

Track 3: AI Progress Monitoring & Early Warning Systems:

Abdymazhit Toretay
AITU

Shoplikov Alisher
AITU

Dzholdybayev Zhan

Toleugazyuly Zhanmadi

With
Apart Research

Abstract

This project presents a machine learning system that analyzes historical AI industry data and forecasts future trends using LSTM neural networks. The system tracks 19 AI-related metrics, including model size, compute, costs, and hardware performance, generating 5-year forecasts. A FastAPI backend powers the prediction and visualization pipeline, while a React dashboard displays interactive Plotly charts. The project helps visualize the evolution of AI technology and anticipate future developments.

Keywords: AI, LSTM, forecast, RAG, LLM, models.

1. Introduction

The AI industry has grown rapidly over the past decade, driven by advances in model architecture, computation, and hardware. Understanding these trends is crucial for predicting future costs, capabilities, and research directions. The *AI Trends Forecasting Dashboard* provides an interactive tool for analyzing past developments and projecting future trends based on real data.

Using time series forecasting with LSTM models, the system predicts how core AI metrics — such as model parameters and GPU performance — will evolve in the next five years. It combines deep learning, data visualization, and web technologies to make complex forecasting accessible and interpretable.

2. Methods

Data and Processing

Data from 2010 onward were collected across several sources covering AI models, GPU clusters, and hardware. Each dataset was cleaned, aggregated yearly, and normalized using **MinMaxScaler**. Missing values were filled through interpolation to maintain continuity.

Modeling

Each of the 19 metrics is forecasted with an **LSTM neural network** trained on 3-year sliding windows.

Key parameters:

- 1 LSTM layer (32 units), 0.1 dropout
- Adam optimizer, MSE loss
- Early stopping for regularization

Models were trained in TensorFlow/Keras and saved as **.keras** files with their respective scalers.

System Design

The backend, built with **FastAPI**, loads the trained models, performs inference, and generates **Plotly** visualizations. The **React** frontend retrieves data from **‘/graphs’** and displays side-by-side charts for historical and forecasted trends.

An experimental module uses **ColBERT v2.0** for semantic search over PDF documents, enabling retrieval of relevant AI research text.

3. Results

To demonstrate the semantic search capabilities, the ColBERT-based RAG system was queried against the 2025 AI Index Report (456 pages) with the question: "Which companies will be most in demand in 2026?"

The system successfully retrieved contextually relevant passages and identified key organizations projected to be in high demand, including Code.org, CSTA, and ECEP Alliance (computer science education expansion), Ipsos (AI public opinion research), and The Raspberry Pi Computing Education Research Centre (global CS education research). Notably, the retrieval correctly disambiguated the query intent—recognizing that the report emphasizes educational and research organizations rather than commercial AI product companies.

```
query = "Which companies will be most in demand in 2026?"

Answer:

Okay, based on the provided text excerpts from the 2025 AI
Index report, here's a breakdown of the companies likely to be
in high demand in 2026, along with the source material:

**Key Companies in Demand (2026 - According to the Report):**

*   **Code.org, CSTA, and ECEP Alliance:** These organizations
are highlighted as crucial players in expanding and improving
computer science education, suggesting sustained demand for
their services and expertise. (Page 454, Appendix 2)

*   **Ipsos:** The report points to Ipsos as conducting a
survey to provide key insights and data on public opinion.
(Page 456)

*   **The Raspberry Pi Computing Education Research Centre:**
This group is involved in gathering data and expanding
research about computer science education around the world.
(Page 455, Appendix 4)

**Important Note:** The report doesn't offer a list of
specific companies in terms of *industry* or *product*
categories. Instead, it identifies organizations and research
groups heavily involved in the development and expansion of
AI-related skills and knowledge.

Do you want me to elaborate on any of these points, or would
you like me to investigate other aspects of the report (e.g.,
specific technologies, trends, or data sources)?
```

In AI Forecast Dashboard (Figure 1) presented data of some predicted metrics, collected from [1] (it contains data about models, ML hardware, GPU clusters etc.)

In Figure 1 you can see plots of forecast and raw data, that are named by (where it taken [.csv] and what feature taken [ex. 'Parameters']).

The forecast was made for 5 years, 2025-2030 (by using LSTM).

RMSE ≈ 0.2 .

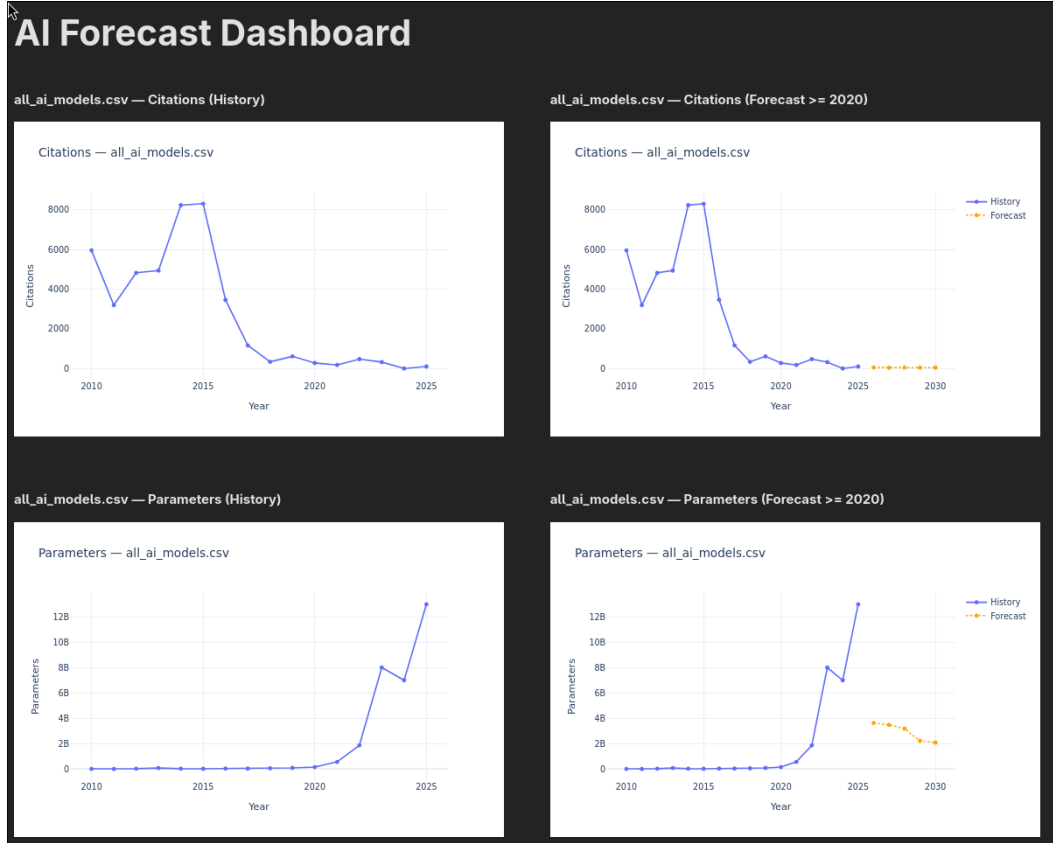


Figure 1 – AI Forecast Dashboard by data from [Epoch](#) (other graphs in Appendix)

4. Discussion

The dashboard demonstrates how deep learning can forecast trends in complex technological domains. LSTM models effectively capture long-term dependencies, revealing insights into the pace and direction of AI progress.

The modular structure allows easy integration of new datasets and model updates. The combination of forecasting and visualization supports decision-making for researchers and industry analysts, offering a clear view of AI's projected trajectory.

Overall, the system provides a flexible and interpretable framework for monitoring and forecasting AI development using real-world data.

5. Conclusion

This project demonstrates a practical approach to monitoring and forecasting AI industry progress through machine learning-driven time series analysis combined with intelligent document retrieval. By training 19 independent LSTM models on historical data spanning over a decade, the system successfully generates 5-year

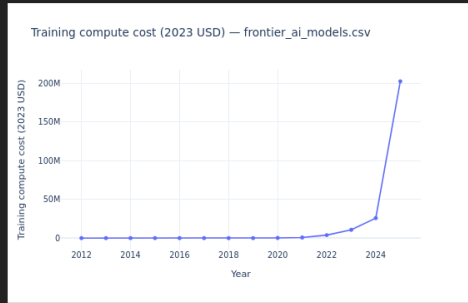
forecasts across critical metrics including model scale, computational requirements, hardware capabilities, and economic factors. The integration of deep learning forecasting with interactive visualization provides stakeholders—researchers, policymakers, and industry analysts—with an accessible tool for understanding AI trajectory and identifying potential inflection points in technological development. Beyond quantitative forecasting, the system incorporates a Retrieval-Augmented Generation (RAG) pipeline using ColBERT v2.0 for semantic search over authoritative sources such as the AI Index Report. This enables users to query complex questions (e.g., "Which companies will be most in demand in 2026?") and receive contextually grounded answers extracted from comprehensive industry analyses, bridging the gap between numerical predictions and qualitative insights.

6. References

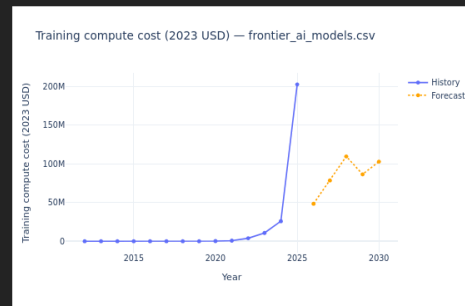
- [1] <https://epoch.ai/>
- [2] <https://github.com/stanford-futuredata/ColBERT>

7. Appendix

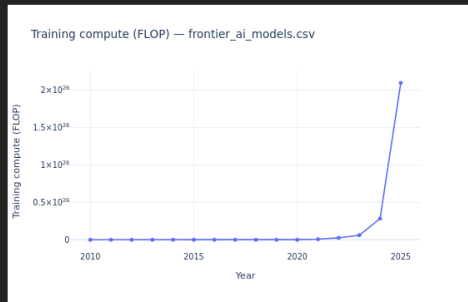
frontier_ai_models.csv — Training compute cost (2023 USD) (History)



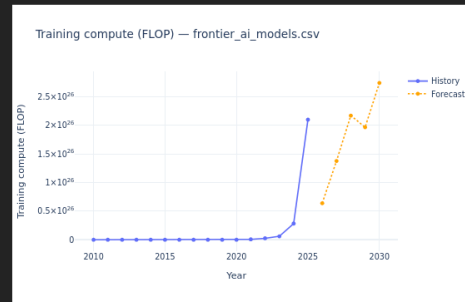
frontier_ai_models.csv — Training compute cost (2023 USD) (Forecast >= 2020)



frontier_ai_models.csv — Training compute (FLOP) (History)



frontier_ai_models.csv — Training compute (FLOP) (Forecast >= 2020)



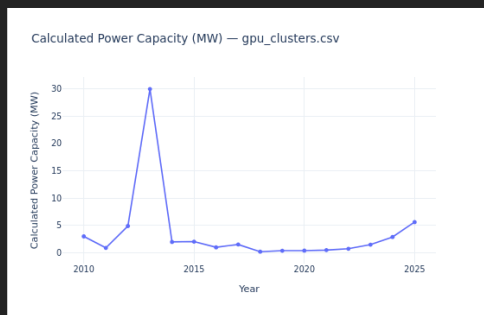
gpu_clusters.csv — Calculated Cost (History)



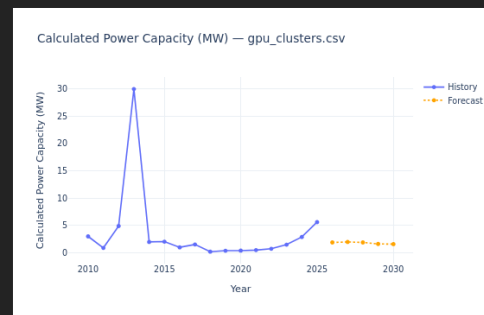
gpu_clusters.csv — Calculated Cost (Forecast >= 2020)



gpu_clusters.csv — Calculated Power Capacity (MW) (History)



gpu_clusters.csv — Calculated Power Capacity (MW) (Forecast >= 2020)



gpu_clusters.csv — Hardware Cost (History)



gpu_clusters.csv — Hardware Cost (Forecast >= 2020)



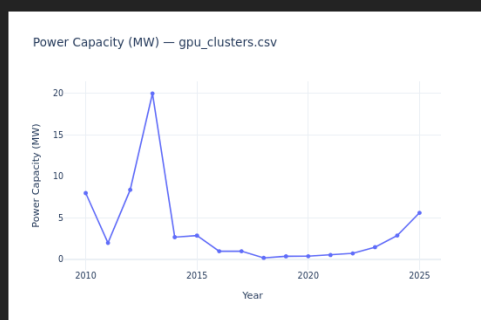
gpu_clusters.csv — % of largest cluster when first operational (History)



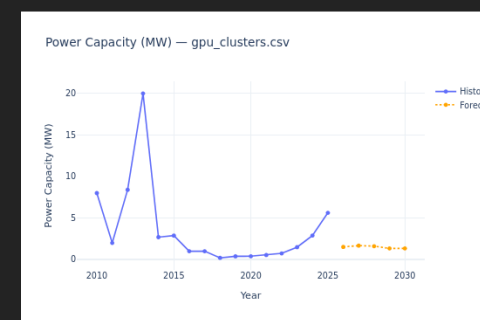
gpu_clusters.csv — % of largest cluster when first operational (Forecast >= 2020)



gpu_clusters.csv — Power Capacity (MW) (History)



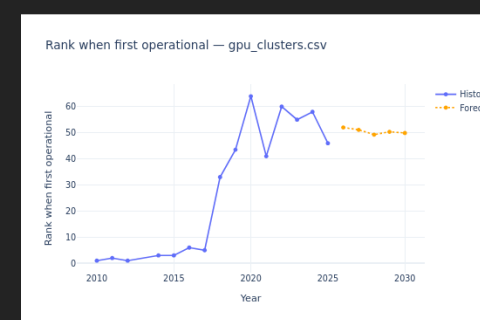
gpu_clusters.csv — Power Capacity (MW) (Forecast >= 2020)



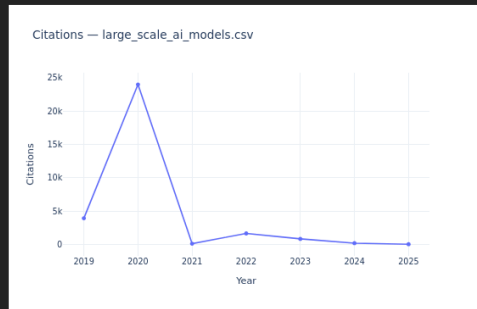
gpu_clusters.csv — Rank when first operational (History)



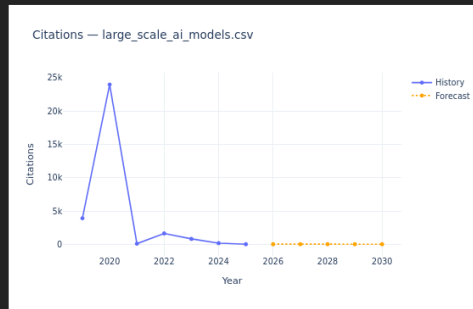
gpu_clusters.csv — Rank when first operational (Forecast >= 2020)



large_scale_ai_models.csv — Citations (History)



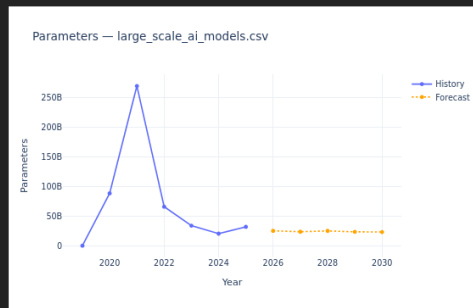
large_scale_ai_models.csv — Citations (Forecast >= 2020)



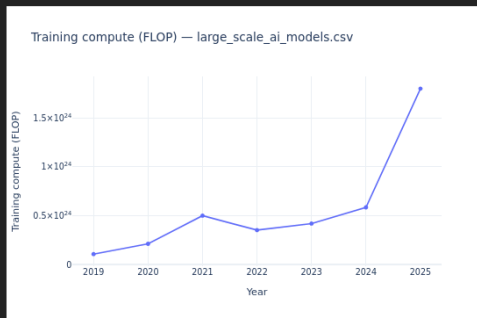
large_scale_ai_models.csv — Parameters (History)



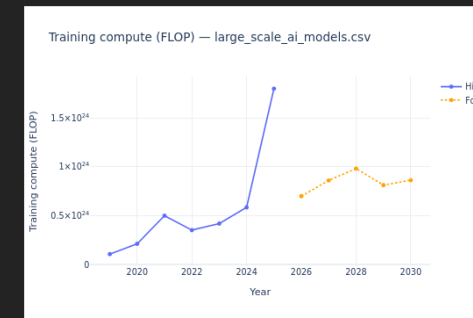
large_scale_ai_models.csv — Parameters (Forecast >= 2020)



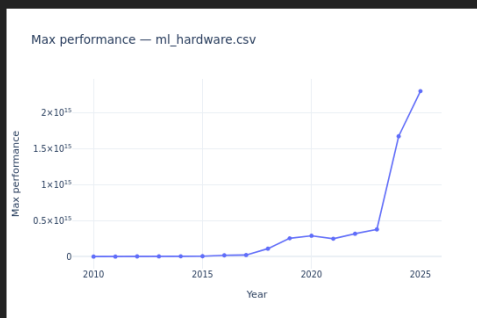
large_scale_ai_models.csv — Training compute (FLOP) (History)



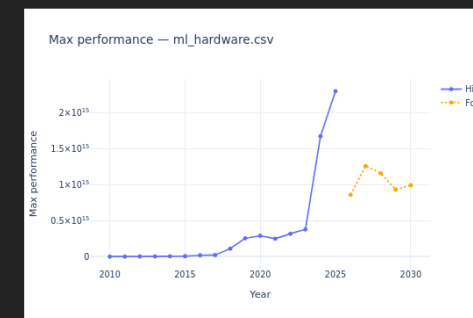
large_scale_ai_models.csv — Training compute (FLOP) (Forecast >= 2020)



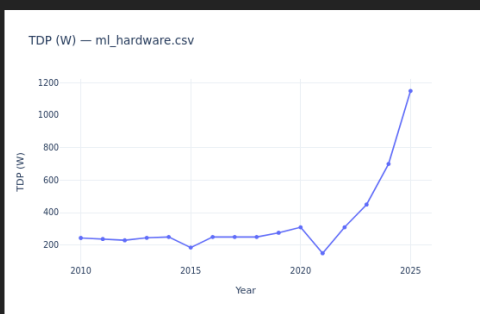
ml_hardware.csv — Max performance (History)



ml_hardware.csv — Max performance (Forecast >= 2020)



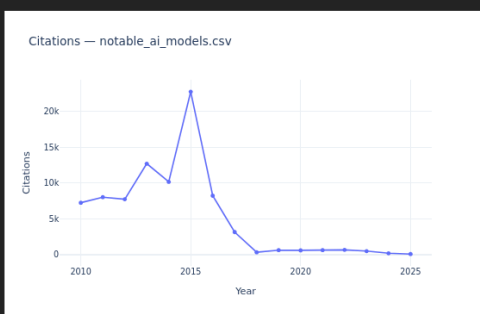
ml_hardware.csv — TDP (W) (History)



ml_hardware.csv — TDP (W) (Forecast >= 2020)



notable_ai_models.csv — Citations (History)



notable_ai_models.csv — Citations (Forecast >= 2020)

