

# Big Data Project

## Past Projects and Project Ideas

Jonathan Fürst

# Past Projects

# Projects from last year

Projects from last year:

- <https://github.com/orgs/BDP25/repositories>
- <https://bdp25.github.io/>

# Project Ideas

# Project Idea: Agenic AI

## Size Matters: Energy Trade-offs between Small and Large LLMs in Agenic AI

- Agenic AI refers to autonomous artificial intelligence systems that can make independent decisions to achieve specific goals.
- Nowadays people refer to this in context of LLMs and tool calling and frameworks such as Langchain, Langgraph and protocols such as MCP
- Problem: High processing times, cost/energy and latency → does not scale for Big Data
- Idea:
  1. Implement an agentic workflow for a data science problem
  2. Integration of EnergyMeter
  3. Measure and potentially optimize (e.g., small vs. large LLM)

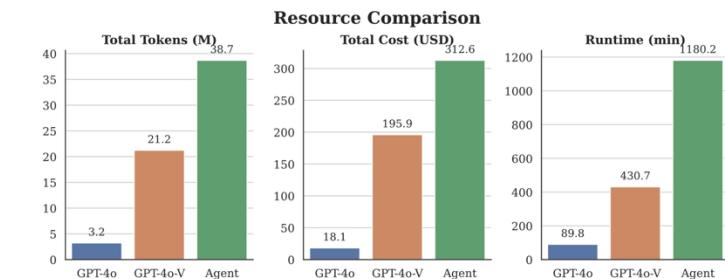


Figure 7: Resource usage. The agentic system incurs significantly higher costs compared to static baselines.

# Project Idea: Athletes Network

Optimierung eines KI-basierten Recruiting-Matching-Algorithmus

- **Der Status Quo (Pipeline):**
  - Automatisches **Parsing** von CVs (Athlet:innen) und Job-Descriptions mittels KI.
  - Berechnung eines **Matching-Scores** (0-100).
- **Human-in-the-Loop:** Manuelle Prüfung aller Matches mit Score > 75.
- **Das Problem:**
  - Aktuelle Genauigkeit ist unzureichend: **80-90% False Positives**.
  - Großteil der KI-Vorschläge wird vom Vermittlungsteam als „Kein Match“ verworfen.
- **Das Projekt-Ziel:**
  - **Mittelfristig:** Signifikante Erhöhung der „Match OK“-Rate (Präzision).
  - **Langfristig:** Vollautomatisierung des Matchings ohne manuelle Prüfung.
- **Die Datenbasis:**
  - Zugriff auf **reale CRM-Daten** (anonymisiert, unter NDA).
  - Features: CV-Daten, Job-Requirements, historische Matching-Entscheidungen (Labels).