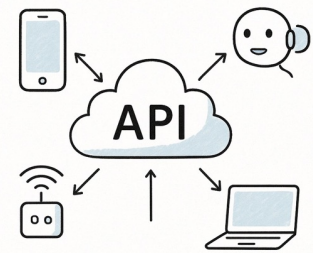


Learning Graph-Based API Models from REST Traffic for Intelligent Web Caching

Motivation

APIs are the driving force behind the digital economy—powering everything from web applications and IoT devices to next-generation AI agents. The demand for fast, efficient, and sustainable data communication is growing exponentially, and innovative solutions are needed. We are working on a startup project to build the first AI-driven web proxy for semantic API optimization. Leveraging cutting-edge research and state-of-the-art AI technologies, we offer a unique opportunity to actively shape a transformative solution.



Shaping the Future of API Communication

Your Task

Develop an AI-based solution that generates a graph-based API model from observed REST API traffic and/or an OpenAPI spec. The model allows caching of arbitrary REST APIs on a Content Delivery Network

- Related-work analysis and evaluation of suitable ML approaches. One approach could be reinforcement learning on a small LLM, see [1]
- Create or find a dataset for training and testing the approach (from public datasets, public APIs like [2], or a self-hosted example server)
- Implement an automatic feedback loop by creating a model, deploy it, collect cache statistics, derive an improved model, repeat
- Compare the effectiveness to an existing system and doing simple prompt engineering with the OpenAI API

[1] <https://www.youtube.com/watch?v=C4HxJQ2QzWo>

[2] <https://openapisearch.com/search>

Requirements

- Experience in deep learning, full-stack development, REST API design
- Excellent skills
- Motivation :)

Contact

Markus Sosnowski sosnowski@net.in.tum.de
Kilian Glas glask@net.in.tum.de

