**OPEA Comps**

**Difficulty:** Level 200

**Business** **Goal:**

The company wants you to explore the effort it would take to run the AI workloads completely on servers that will live in-house. The fractional CTO, suggests that its best practice to run workloads in containers or kubenetes. You as the AI Engineer have been tasked to determine how to learn to work with the building blocks to constructor your own GenAI workloads running on containers.

**Technical Uncertainty**

* Using OPEA does it serve the model via a LLM server?
* How do we orchestrate two services together?
* What is the quality of build across the various OPEA Comps?

**Technical Restrictions**

* [GenAIComps (GitHub Repo)](https://github.com/opea-project/GenAIComps)
* [OPEA Comps Project](https://opea-project.github.io/latest/GenAIComps/README.html)
* Docker Containers

**Homework Challenges**

* Orchestrate multiple services eg. 2 or 3 together
* Or Try and get a different comp working that Andrew did use not use.

**Homework Bonuses**

* Make a tutorial or technical doc that is public on LinkedIn, Medium, Hashnode, Your Blog
  + Tag Andrew or show off the work in the Discord show-and-tell.