AI Queuer

Requirements

Description

Without any setup in my computer I want to be able to query incoming AI prompt requests and answer them.

Things to note:

- My answer might take some time, users should not be halted while the response is being generated.

- Only admins should be able to query and respond.

- Users need to be authenticated in order to create prompt requests. (future)

Architecture

Database: PostgresSQL, Ent (ORM)

API: Golang, Gin

Infrastructure: Docker, NGINX, Raw VPS

Database

Table	Column Names		
PromptRequest	identifier - string	prompt - string	is_queued
PromptResponse	prompt_request - relation	response - string	

API

Endpoint	Method	Body	Response
/prompt	POST	prompt : string	queue_id : string

/prompt/:id	GET		prompt : string , state : enum , response? : string
/prompt/:id	PUT	<auth>	prompt : string , state : enum
/prompt/:id	POST	<auth> , response : string	prompt : string , state : enum , response : string

States: `awaiting` , `queued` , `responded`

Middleware

- Authentication (admin + user?)

- Check State (enforce when necessary)