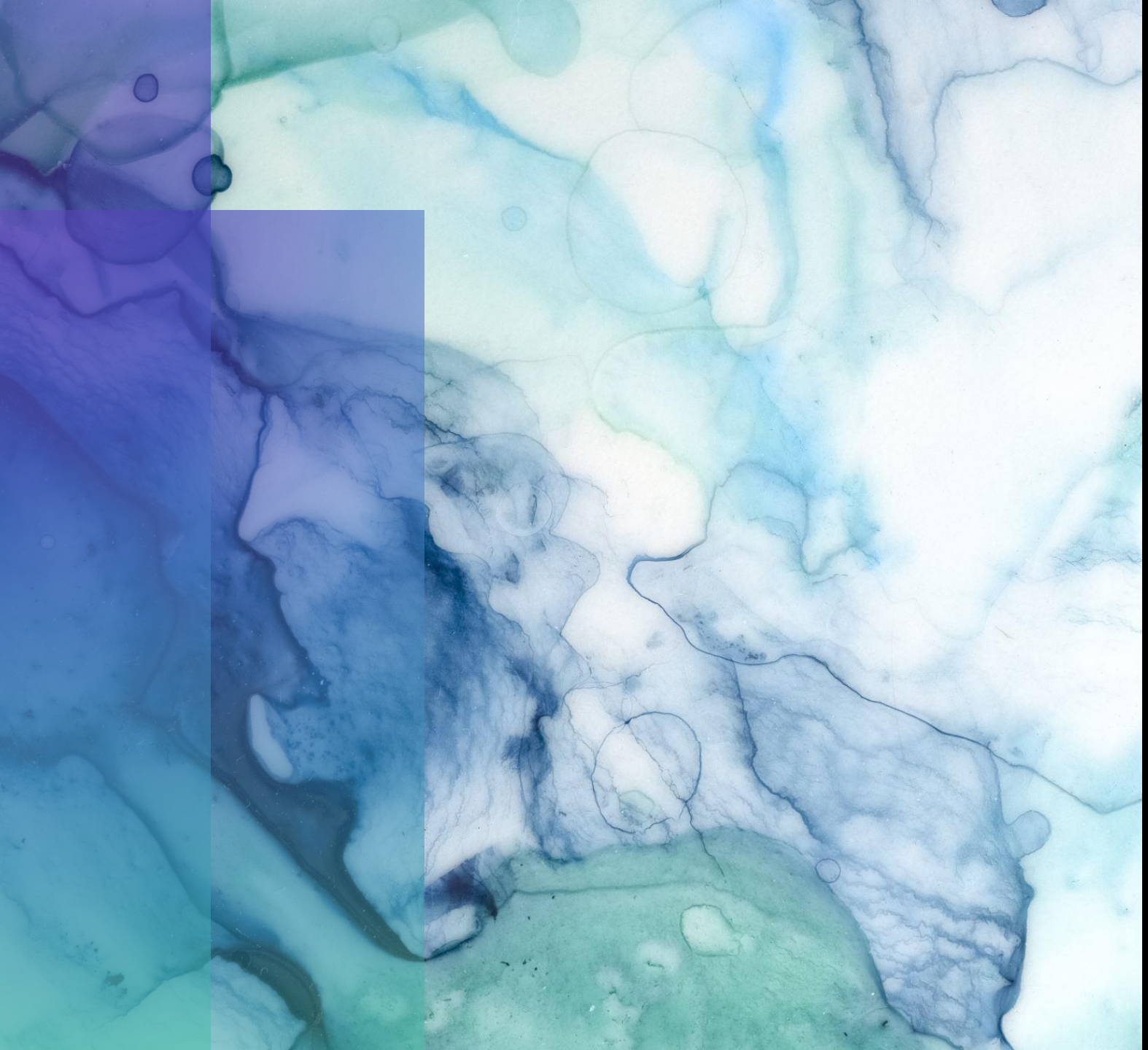


Accessing Large Language Models

Greg Witt

Overview

- Platforms – Mistral AI, Openai
- API – Access, Keys, LangChain
- RAG – When / Why
- Example - Project

Mistral AI

- Founded in April 2023
- Headquarters in France
- Focused on Producing Open AI models
- Founders
  - Arthur Mensch (Google Deepmind)
  - Guillaume Lample (Meta Platform)
  - Timothee Lacroix (Meta Platform)
- Models:
  - Mistral 7B – uses grouped-query attention
  - Mistral 8x7B – 8 group mixture of experts
  - Mistral Large 2 – 123 billion parameters

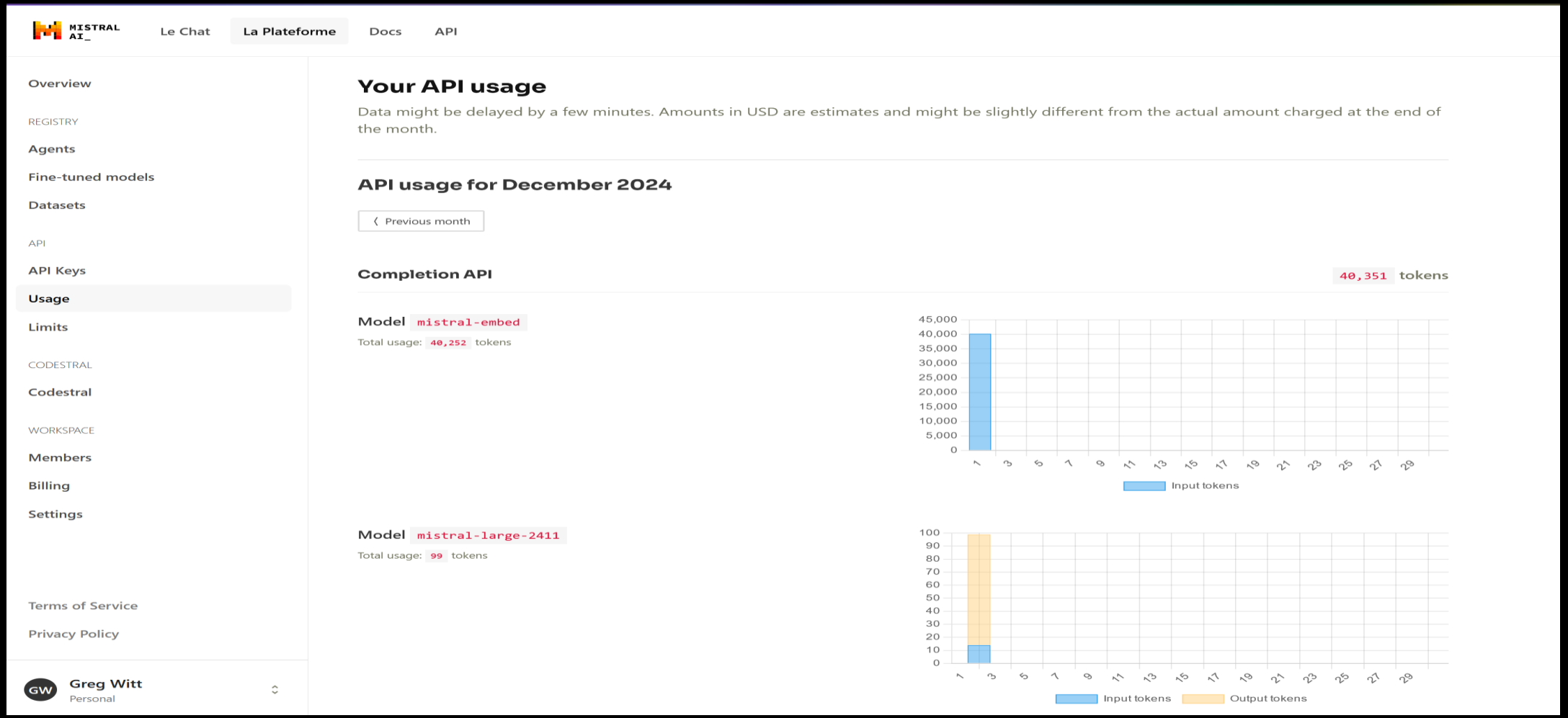
https://en.wikipedia.org/wiki/Mistral_AI

OpenAI

- Founded in Dec 2015
- Headquarters California, US
- Started in 2015 as a Non-Profit
- Shifted in 2019 as For-Profit
- Founders:
  - Elon Musk (Initial Investor)
  - Sam Altman (CEO)
  - Greg Brockman (President)
- Models:
  - ChatGPT
  - GPT 4
  - O1

https://en.wikipedia.org/wiki/OpenAI

Le Platforme



https://mistral.ai/news/la-plateforme/

OpenAI Platform

https://platform.openai.com/docs/overview

Configuration Time

API Access

Access Via Keys

Rate Limiting

Embedding Costs

Token Sizes

User Base

Hidding Your Keys.

Code Along

RAG Basics

- Upload Documents
- Chunk Documents
- Embed Documents
- Retrieve Documents

Architecture