How should the PII be masked?

After a brief research, I ended up deciding between 3 options:

- entirely custom solution
- using Scrubadub
- using Presidio

I assumed the following limitations:

- challenge implementation should take 10-15h (in reality took ~18h ☺)
- challenge implementation will be done in overtime and during the weekend (so with limited energy, but actually I had more than I imagined☺)

Because of this I ruled out designing and implementing my own algorithms from scratch, because of the mentioned assumptions. Both scrubadub and presidio packages looked OK (stars on GH, documentation), so I've chosen one that I assumed will be easier to use.
I ended up in using custom detectors and post processors in this framework and it went very smoothly, so I'm happy with the decision.

I was also thinking about the most ambitious approach – training a local LLM that would mask PIIs before sending to the remote LLM, the one that should not see them at any time. This might be the best approach in terms of what can be masked and how accurately, but it would definitely exceed the time limit considerably (although I would greatly benefit from the acquired knowledge, looking forward to try this approach in the future).

What web framework to use?

I was given choice between Flask and FastAPI. Regardless that I'm more experienced with Flask, I have chosen FastAPI, because:

- it is better in streaming responses,
- it offers out-of-the box documentation required by the challenge,
- I'd wanted to learn FastAPI better.

How to integrate the LLM?

Here I had two ideas – using an OpenAI client, or leveraging the fact that our company has a Slack plugin called AI Buddy which is capable to talk to all top OpenAI GPT models, so there was a possible option of integrating a Slack client that would proxy messages to LLM. I discarded the latter idea as too complex and created a paid account on platform.openai.com and added some quota.

What LLM client to use?

You use langchain and I do not have experience in it so the choice here was very easy – I've chooses langchain Python's client.