

REST API with PII Masking & LLM Integration**

Objective

Build a REST API using **FastAPI** or **Flask** that:

1. **Input:** Accepts a user prompt and contextual text.
2. **PII Detection & Masking:** Detects and masks PII (e.g., names, phone numbers, emails) in both the prompt and context.
3. **LLM Processing (NO PII Exposure):** Ensures the LLM **never** receives or processes any PII. Only the masked input should be sent to the LLM.
4. **Unmasking & Streaming:** Streams the response token-by-token while unmasking PII before sending the final answer back.

API Specification:

- **POST /generate-answer**
 - Accepts `prompt` and `context` as input.
 - Detects and masks PII before sending to LLM.
 - Streams the response with PII unmasked.

Example Request:

```
{
  "prompt": "What is John's phone number?",
  "context": "John's phone number is 555-123-4567. He lives at 123 Maple Street in San Francisco, and his email is john.doe@example.com."
}
```

Response (Streamed):

```
{
  "response": "John's phone number is 555-123-4567"
}
```

Key Rule:

The LLM should never see any PII.

Evaluation Criteria:

- Correct PII masking/unmasking.
- LLM receives no PII.
- Efficient token streaming.
- Code quality and documentation.