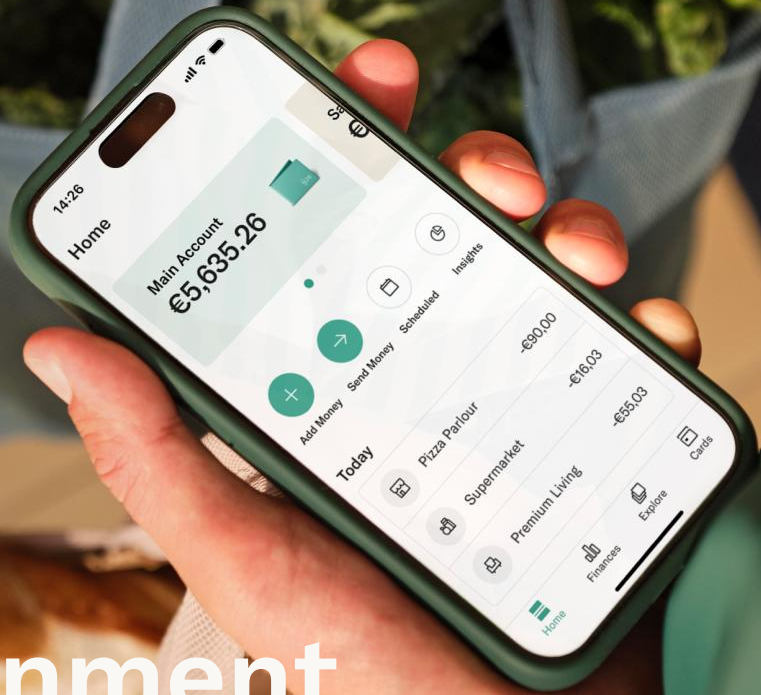


Love your bank

# Generative AI - Take Home Assignment

N26 Data Science - Generative AI  
April 2025



# Motivation

N26 processes a large volume of diverse documents from customers, as well as public and private institutions. These documents play a crucial role in the bank's financial and regulatory operations.

Managing these documents manually is time-consuming and complex, posing challenges to operational efficiency and scalability.

With advancements in Language Model technology, we now have the opportunity to streamline document handling through automation, saving time and costs for all parties involved.



# Your Task

Extract specific information from a set of documents. The extracted data will then be passed to another service for further processing.

The information you need to extract falls into two categories:

- General Fields - These fields must be extracted from every document.
- Document-Specific Fields - These depend on the document type, which is identified within the general fields.

“

**This is your chance to  
show your AI  
whispering skills.**

”

# Provided Datasets

- 4 different document types in 5 different languages (English, German, French, Spanish, Italian)
  - Personal Account
  - Credit
  - Garnishment
  - Investment
- JSON file containing the required extraction fields for the documents
- CSV file containing the field descriptions (reference for the field contents / dictionary)

# Requirements

- Commented, clean analysis in a Jupyter notebook. Please be explicit about your testing and evaluation strategy.
- Python script (document\_extractor.py) that accepts documents as input and provides the output ready for the downstream process
- For the downstream process, ensure to standardize your output:
  - Dates should follow the format “dd.mm.yyyy”
  - Names should follow the format “first\_name last\_name”
  - Addresses should follow the format “street\_name street\_number, city zipcode, country”
  - Monetary amounts should follow the format “amount currency\_symbol” using a full stop as decimal separator for the amount
  - Additional fields must follow the same format as originally found in the document.
- Properly packaged source code with running instructions. It should run on Mac / Linuxlike OS
- A few slides summarizing assumptions, experimentation steps and final results. Outline steps which would be required to integrate this solution into live/production.

# Any questions?

Feel free to send any questions you might have to your Tech Recruiter.

We want to ensure that everyone has an equal shot at this task, so we will respond to all applicants with compiled task clarifications.

All the best for the task! 🚀

Thank you for  
your time.

