Take home project – Generative AI back-end

Story

As an iBanFirst customer, I would like to extract the invoice details from a pdf file.

Description

You will find enclosed three examples of invoice.

The expected delivery is a command which takes the path to the pdf file as an argument and outputs a valid json with the following keys:

- Reference: (string) A unique ID to identify the invoice
- Beneficiary: (string) Beneficiary name
- Account ID: (string) The beneficiary account ID
- Amount: (decimal) Amount to send
- Currency: (string) 3 characters to identify the currency to send (e.g. EUR, USD)
- Due date: (date) Payment due date

Please note that all fields are optional, and the invoice may be partial.

Delivery

The goal is to assess your ability to use libraries effectively and implement the required functionality.

The delivery includes:

- A **private** GIT repository. The sources should be documented, tested and ready to deliver in production (clean and without useless code).
- A README file to explain the build and execution instructions.

The programming language must be Python, but the choice of framework and implementation is flexible.

Don't ignore the packaging (code easy to run), think about the code readability (style, structure and comments), error handling and add tests to cover the edge cases.

Command sample

```
$ parseInvoice ~/invoices/invoice1.pdf
```

```
{
  "reference": "287605fd-a",
  "beneficiary_name": "ABC Corp",
  "account_id": "FR7630006000011234567890189",
  "amount": 36.0,
  "currency": "EUR",
  "due_date": "2025-01-17"
}
```