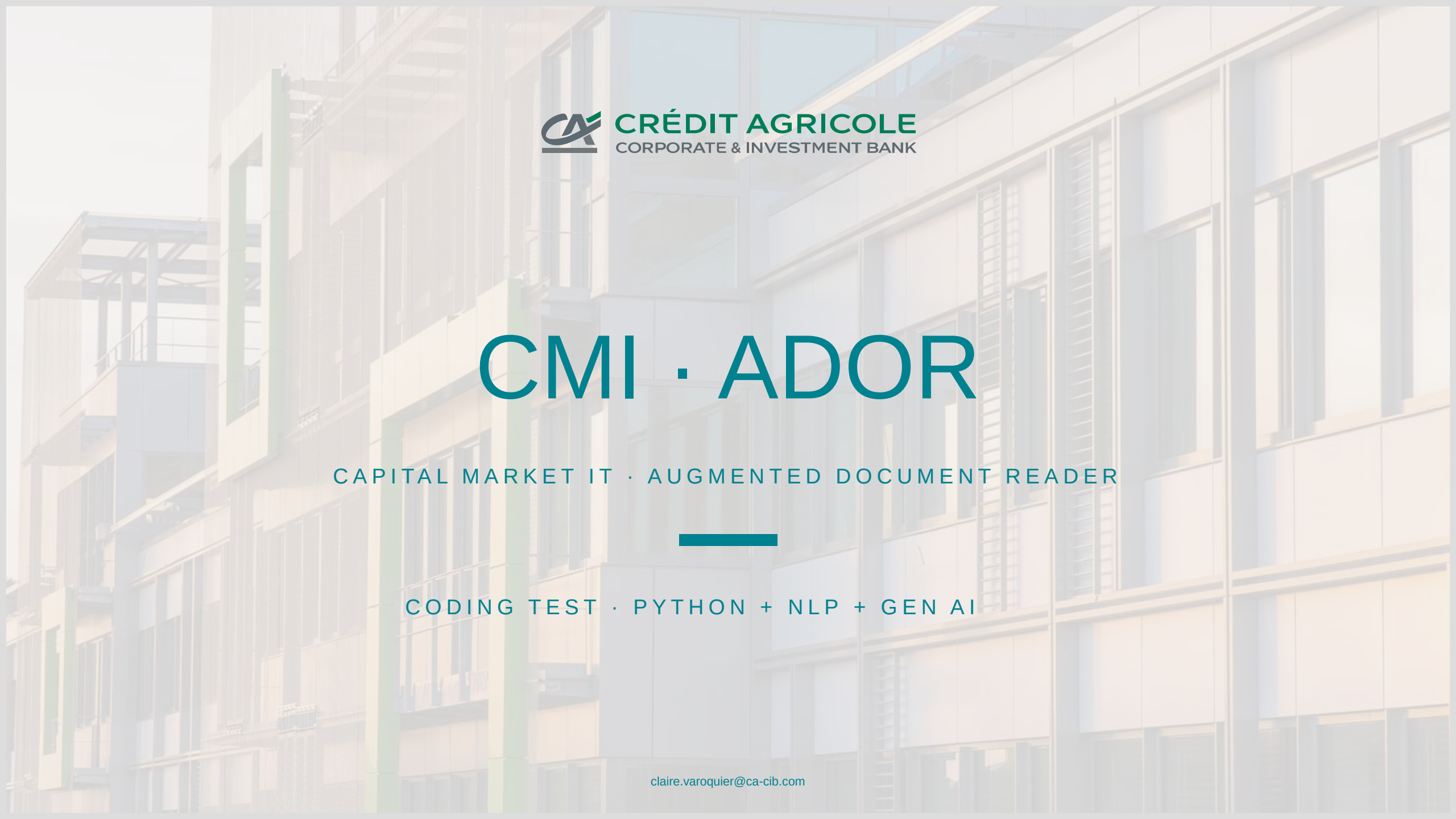

CRÉDIT AGRICOLE
CORPORATE & INVESTMENT BANK

CMI · ADOR

CAPITAL MARKET IT · AUGMENTED DOCUMENT READER

CODING TEST · PYTHON + NLP + GEN AI

claire.varoquier@ca-cib.com

Overview

Context & Objectives

Context ► Welcome to CMI Architecture & Innovation team. As an AI engineer your first mission is to build a financial document reader tool augmented by IA. The final product will be able to classify or summarize documents, discover predefined topics, recognize named entities or answer questions related to the provided document.

test typology

Python + NLP + Gen AI

Objectives ► In this coding test, you will only work on the Named Entity Recognition (NER) feature of the product. You will provide a Proof of Concept (PoC) to demonstrate how the tool can parse and extract financial entities from documents. Depending on the nature of the provided document you can use a rule-based parser, a NER model or a Large Language Model (LLM).

test language	*coding work items*	*test platform*
	★★☆☆	GitHub

work items

Architecture	30 min	
Parser	60 min	
NER	60 min	
LLM	30 min	Total **3 hours**

Data ► chat, docx and pdf files

Backend ► Python, API

Model ► open source NER model

Evaluation ► code and docs quality

Work Items (WI) Description

Architecture WI ► The first expected work item is a Global Architecture Document (GAD) that describes the interactions between the CMI Information System (IS) components and the document reader. The reader can be invoked programatically via APIs, and will also provide a User Interface (UI) enabling end users to upload a document and launch a classification, summarization, topic modelling, NER or Q&A feature. Documents will vary in size, format and level of confidentiality. They can be sent through different communication channels and processed in a synchronous or asynchronous way.

Coding WI ► Some kind of documents (e.g. docx files) can be processed by a rule-based parser coded in Python. For this work item, the expected artifact is a program that takes a document as input and returns a set of named entity values. The entities to extract are listed in the next slide. You can choose which Python packages to use and the format of the output files.

Coding WI ► Other kind of documents (e.g.chats) can be processed by a NER model. This work item is a combination of Python code and a Global Methodology Document (GMD). The Python code will give an overview of how to download and run a general-purpose NER model to extract named entities. You can choose which model to use. The methodology document will explain how this model can be fine-tuned to extract the financial entities listed in the next slide.

Methodology WI ► The last type of documents (e.g. pdf files) are more verbose, unstructured and require a more advanced language model. For this work item a GMD will explain how to build an entity extraction pipeline that relies on LLMs. The document will also include a description of the prompting and/or Retrieval-Augmented Generation (RAG) techniques to be used.

Data

11:49:05 I'll revert regarding BANK ABC to try to do another 200 mio at 2Y
FR001400QV82 AVMAFC FLOAT 06/30/28
offer 2Y EVG estr+45bps
estr average Estr average / Quarterly interest payment

chat

Counterparty ► BANK ABC

Notional ► 200 mio

ISIN ► FR001400QV82

Underlying ► AVMAFC FLOAT 06/30/28

Maturity ► 2Y EVG

Bid ► estr+45bps

Offer

PaymentFrequency ► Quaterly

Entities to extract

docx file

Counterparty ►

Initial Valuation Date ►

Notional ►

Valuation Date ►
Maturity ►
Underlying ►

Coupon ►
Barrier ►

Calendar ►

ZF4894, ALV, physical
Barrier 75%, 07 August 2026

Party A	BANK ABC
Party B	CACIB
Trade Date	31 January 2025
Trade Time	09:12:15
Initial Valuation Date	31 January 2025
Effective Date	07 February 2025
Notional Amount (N)	EUR 1 million
Upfront Payment	***TBD***% to be paid by Party B to BANK ABC on the Effective Date
Valuation Date	31 July 2026
Termination Date	07 August 2026
Underlying	Allianz SE (ISIN DE0008404005, Reuters: ALVG.DE)
Exchange	XETRA
Coupon (C)	0%
Barrier (B)	75.00% of Shareini

Interest Payments	
Interest Payments	None

Equity Payments	
Initial Price (Shareini)	Official closing price of the Underlying on the Initial Valuation Date on the Exchange
Sharefinal	Official closing price of the Underlying on the Valuation Date on the Exchange
Business Day	TARGET
Future Price Valuation	Not Applicable
Calculation Agent	Party B and Party A
ISDA Documentation	Option