INTERNATIONAL UNIVERSITY OF RABAT

Higher School of Computer Science and Digital Technology

Engineering Cycle : Big Data and Artificial Intelligence

FINAL YEAR PROJECT REPORT

Forgery Detection In Digitized Documents Using Artificial Intelligence

Completed by: Aya KARBICH

Defended on July 4, 2025, before the jury :

Prof. Mehdi ZAKROUM	*Academic Supervisor*
Prof. Youssef GAHI	*Examiner*
Mr. Mohammed TOUHAMI OUAZZANI	*Professional Supervisor*

Academic Year : 2024 – 2025

ACKNOWLEDGMENTS

The completion of this project owes its success to the essential contribution of several people, whose decisive support deserves my deepest recognition. First and foremost, I wish to thank God for giving me the strength, patience and will necessary to carry out this project. I express my deepest gratitude to my parents. Their love, sacrifices, support and constant encouragement have been my greatest source of motivation. May they find here the expression of my eternal gratitude. I thank the entire faculty and administrative staff of the Higher School of Computer Science and Digital Technology at the International University of Rabat. The quality of the education provided has been the foundation upon which this project could be built. I wish to extend my warmest thanks to my academic supervisor, Professor **Mehdi ZAKROUM**. I thank him for his invaluable guidance throughout this internship, for the relevance of his advice that guided my approach, as well as for his great kindness and availability. His scientific rigor has been a major asset for the success of this project. My thanks also extend to my company supervisor, Mr. **Mohammed TOUHAMI OUAZZANI**. His trust and the quality of his supervision within Attijariwafa Bank allowed me to integrate quickly and work under optimal conditions. I thank in advance Professor **Youssef GAHI** for agreeing to evaluate this work. Finally, I do not forget my friends and classmates for their mutual support and the moments we shared.

SUMMARY

Faced with the resurgence of sophisticated document fraud, generated by the digital transformation of the banking sector, manual control methods prove largely insufficient. This final year project, conducted within Attijariwafa Bank, addresses this challenge by developing an artificial intelligence system for automated detection of falsifications in digitized documents.

The objective is to precisely localize alterations by combining two complementary approaches. The visual methodology relies on a *Vision Transformer* model (SegFormer), optimized for segmentation of suspicious zones. In parallel, semantic analysis employs a robust OCR pipeline and a large language model (LLAMA-3) to identify textual and logical inconsistencies.

Experimental results validate the system's high performance, demonstrating its ability to identify the vast majority of frauds. The project results in a functional application prototype, based on a microservices architecture, confirming the solution's viability in a professional environment.

In conclusion, this work provides a technical and industrializable solution enabling significant reduction of financial risk, optimization of operational efficiency and improvement of client journeys.

Keywords

Document fraud, SegFormer (Vision Transformer), hierarchical OCR, LLaMA-3 (LLM), explainable AI (XAI), LLM Whisperer.

ABSTRACT

Face à la montée de la fraude documentaire sophistiquée, alimentée par la transformation numérique du secteur bancaire, les méthodes de révision manuelle s'avèrent largement insuffisantes. Ce projet de fin d'études, mené au sein d'Attijariwafa Bank, relève ce défi en développant un système d'intelligence artificielle pour la détection automatisée des altérations dans les documents numérisés.

L'objectif est de localiser avec précision les modifications en combinant deux approches complémentaires. La méthodologie visuelle s'appuie sur un modèle Vision Transformer (SegFormer), optimisé pour la segmentation des régions suspectes. Parallèlement, une analyse sémantique emploie un pipeline OCR robuste et un grand modèle de langage (LLAMA-3) pour identifier les incohérences textuelles et logiques.

Les résultats expérimentaux valident la haute performance du système, démontrant sa capacité à identifier la grande majorité des fraudes. Le projet aboutit à un prototype d'application fonctionnel basé sur une architecture de microservices, confirmant la viabilité de la solution dans un environnement professionnel.

En conclusion, ce travail fournit une solution technique et industrialisable capable de réduire significativement le risque financier, d'optimiser l'efficacité opérationnelle et d'améliorer les parcours clients.

Mots-clés

Fraude documentaire, SegFormer (Vision Transformer), OCR hiérarchique, LLaMA-3 (LLM), IA explicable (XAI), LLM Whisperer.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ASPP	Atrous Spatial Pyramid Pooling (Module of DeepLabv3+ architecture)
BCE	Binary Cross-Entropy
CPU	Central Processing Unit
DL	Deep Learning
ELA	Error Level Analysis (JPEG error level analysis)
FN	False Negative
FP	False Positive
GAN	Generative Adversarial Network
GIMP	GNU Image Manipulation Program
GPU	Graphics Processing Unit
IoU	Intersection over Union
JWT	JSON Web Token
KPI	Key Performance Indicator
LLM	Large Language Model
MiT	Mix Transformer (SegFormer model encoder)
ML	Machine Learning
MLP	Multi-Layer Perceptron
OCR	Optical Character Recognition
RAG	Retrieval-Augmented Generation
RNN	Recurrent Neural Network
ROI	Return on Investment
RVL-CDIP	Ryerson Vision Lab Complex Document Information Processing
TP	True Positive
ViT	Vision Transformer
XAI	eXplainable AI

GENERAL INTRODUCTION

The digital transformation of the banking sector has fundamentally redefined interactions between financial institutions and their clients. This revolution, accelerated by the health crisis and driven by the emergence of neobanks, has resulted in massive digitalization of traditional banking processes. Procedures once exclusively physical are transforming into entirely digitized journeys, redefining the contemporary banking experience.

This evolution, while considerably improving client experience and operational efficiency, generates new security challenges. Identity verification, account opening and credit granting now rely on the analysis of digital documents transmitted remotely. This digitalization opens the way to sophisticated forms of document fraud, where criminals exploit digital editing technologies to subtly alter identity documents, bank statements and income justifications.

Study justification

Document fraud now represents one of the most critical threats to the banking sector. This problem is experiencing particularly concerning expansion in emerging economies, where rapid financial inclusion clashes with control infrastructures still under development. Global financial losses reach alarming levels, transforming this operational challenge into a strategic imperative.

Traditional detection methods, based on human expertise, show their structural limitations when faced with the increasing sophistication of falsifications and the document volumes processed daily. Manual inspection suffers from temporal constraints, inter-operator variability and cognitive fatigue that affects detection precision. Faced with increasingly subtle modifications, the human eye proves insufficient to guarantee the security of digitalized banking processes.

This problem presents particularly complex technical challenges. The extreme imbalance between authentic and falsified zones complicates machine learning, while the need for explainability imposes specific architectural constraints. Real-time processing requirements and document format diversity make industrialization particularly delicate.

Objectives and expected contribution

This final year project aims to design and develop an automated document fraud detection system, capable of identifying and precisely localizing modifications made to banking documents. The main objective consists of exploiting recent advances in computer vision and deep learning to create a solution that is both performant and explainable.

The expected scientific contribution focuses on developing neural network architectures adapted to the specificities of document falsification detection, notably managing the extreme imbalance between authentic zones and altered zones. On the industrial level, this research aims to provide Attijariwafa Bank with a significant competitive advantage by automating a critical process while maintaining the highest security standards.

General methodology

The adopted approach is structured around a rigorous experimental approach, combining fundamental research and application development. The methodology relies on computer vision and deep learning techniques, with particular attention paid to result explainability to guarantee the operational acceptability of the solution.

The approach integrates the constitution of a representative dataset, the development and comparative evaluation of several model architectures, as well as the design of a demonstration interface allowing visualization of detection results. A thorough experimentation phase will validate system performance.

Report structure

This report is organized into five main chapters that trace the scientific and technical approach adopted :

The **first chapter** establishes the strategic context and research framework by analyzing the macro-banking environment, document fraud trends and presenting the host organization. It develops the research problem and specifies the project's scientific and industrial objectives.

The **second chapter** presents a comprehensive state of the art of document falsification detection techniques, from traditional methods to the most recent approaches based on artificial intelligence. This critical review allows identifying scientific barriers and positioning our contribution.

The **third chapter** details the adopted methodology, including dataset constitution, choice and

adaptation of neural network architectures, as well as retained evaluation metrics. Aspects related to explainability and interpretability of results are also addressed.

The **fourth chapter** presents experimental results obtained, with comparative analysis of different tested approaches. Model performance is evaluated according to several criteria : detection precision, localization capacity, processing time.

The **fifth chapter** discusses practical implications of results, presents the developed prototype and evaluates its integration potential in Attijariwafa Bank's technological ecosystem. Improvement perspectives and future research directions are also presented.

A general conclusion synthesizes this work's contributions and opens perspectives on future challenges in combating document fraud in the banking sector.

This project, conducted within Attijariwafa Bank's Digital Center, is part of the group's digital transformation strategy and contributes to strengthening its position as banking leader in Morocco and Africa.

1
CHAPTER

STRATEGIC CONTEXT & RESEARCH FRAMEWORK

Chapter Introduction

In a constantly evolving financial environment, the massive digitalization of banking processes has transformed client interactions while generating new security risks. This chapter analyzes how the evolution of the global banking sector has created the conditions for the emergence of document fraud as a major strategic threat.

1 Global banking context and emergence of documentary risks

1.1 Digital transformation of the banking sector

1.1.1 Post-2008 crisis transformation : towards digital banking

The 2008 financial crisis accelerated a structural transformation of the banking sector, creating the conditions for massive digitalization that now exposes institutions to new fraudulent risks.

Client process digitalization This transformation is characterized by massive technological investments (McKinsey Global Banking Review 2023) :

— Technology spending has increased by approximately **38%** since 2013 and now represents **10.6%** of net banking revenues.

— Up to **87% of account openings** at American megabanks and fintechs are conducted through digital channels, radically transforming the client relationship.

— Credit applications are increasingly analyzed by AI, automating document verification.

New operational paradigm Digitalization fundamentally transforms the banking relationship :

— **Explosive document volume** : billions of users generate millions of documents daily that need to be controlled.

— **Decision immediacy** : account opening in **10–15 minutes** (some neobanks announce < 10 min) versus several days historically.

— **Physical disintermediation** : more than one-third of branches have closed in Europe and North America since 2015.

Security consequences This transformation generates new critical vulnerabilities :

— **Expanded attack surface** : each digitized document becomes a potential fraud vector.

— **Time pressure** : constrained deadlines limit the depth of manual controls.

— **Facilitating standardization** : format uniformization simplifies fraudulent reproduction.

1.2 Overview of modern banking cyber threats

1.2.1 Digital risk ecosystem

Banking digitalization has created an environment conducive to the emergence of multiple attack vectors, transforming cybersecurity into a critical strategic issue.

Main cyber threat typology According to cybersecurity authorities (CISA, EBA), the banking sector faces five major threat categories :

— **Denial of service attacks** : peaks at $>$ **71 million requests/second** (2023).

— **Banking ransomware** : average cost of approximately **4.9 million $** per incident.

— **Phishing and social engineering** : involved in **68–90 % of data breaches** according to studies.

— **Digital document fraud** : **+ 244 %** between 2023 and 2024 ; now $\approx 57\,\%$ of document fraud is digital.

Sector specificities The banking sector presents particular vulnerabilities linked to its operational constraints : transaction volume of several trillions daily, multiplicity of access channels and regulatory obligations for service continuity.

1.2.2 Document fraud : a specific expanding threat

Position in the fraudulent ecosystem Document fraud is distinguished by its **execution subtlety**, its direct targeting of business processes (onboarding, credit) and its **temporal persistence** : the effects of a false document can manifest several months after its validation.

Specialized typology of document fraud (Sources : Entrust *Identity Fraud Report 2025*, LexisNexis Risk Solutions 2024)

Identity document falsification – 40% of detected cases

— **Basic alterations** : manual modifications still common.

— **Semi-professional reproductions** : high-resolution counterfeits in strong progression.

— **Industrial falsifications** : criminal networks integrating holograms and RFID chips.

Financial document alteration

— Reproduction of legitimate pay slips.

— Creation of fictitious companies.

— Manipulation of standardized banking interfaces.

1.2.3 Regional and economic impact

Francophone Africa : sustained progression Document and identity fraud now shows double-digit growth :

— **+20 % per year** : in several countries, falsification rates increase by approximately 20 % each year ;

— **40 %** of official documents used in criminal activities are false ;

— **15 %** of national identity cards examined in Cameroon in 2023 proved to be falsified, a proportion even higher in border areas ;

— **30,000** fraudulent documents originating from West Africa were detected in international visa applications in 2021 ;

— **20 %** of applications for Senegalese civil service positions controlled in 2022 contained falsified documents.

Economic and security costs

— Losses linked to loans obtained with false documents exceed **150 million $ per year** in West Africa ;

— In 2022, more than **5,000** illegal border crossings in the Sahel were facilitated by falsified identity documents ;

— The introduction of detection equipment in Burkina Faso allowed neutralizing **85 %** of false documents presented at administrative counters in 2023.

2 Presentation of the host organization

This project was conducted within the **Attijariwafa Bank** (AWB) group, in collaboration with the teams from the **Data division** of the **Digital Center**, attached to the **Transformation, Innovation, Technologies and Operations** (TITO) division.

FIGURE 1 – Attijariwafa Bank group logo

2.1 History and positioning

Resulting from the merger announced at the end of **2003** and finalized in **2004** between the *Banque Commerciale du Maroc* (founded in 1911) and *Wafabank*, Attijariwafa Bank is today the leading banking group in the Kingdom — and remains among the very first in Africa by Tier 1 capital.

As of December 31, 2024, the Group relies on **7,223 branches**, **20,782 employees** and a presence in **27 countries** (Maghreb, West and Central Africa, Europe, Middle East). It serves a portfolio of more than **12 million clients** :contentReference[oaicite :0]index=0. In the Moroccan market, its shares reach **26.97% of credit outstanding** and **25.21% of deposits** at the end of 2023.

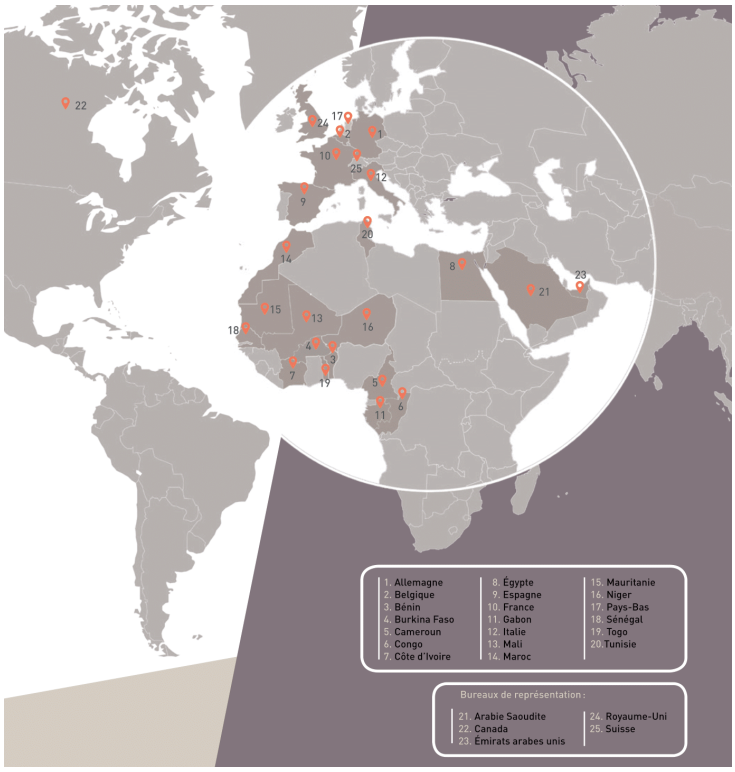


FIGURE 2 – Geographic presence of Attijariwafa Bank.

2.2 Mission, values and strategic axes

— **Mission :** sustainably finance economic development and financial inclusion in its geographies, through a responsible and innovative banking experience.

— **Values :** leadership, integrity, team spirit, client proximity, societal commitment.

— ***@MBITIONS 2025* Plan :**

1. Acceleration of digital transformation (mobile banking, open-banking, data & AI).
2. Strengthening of pan-African influence through targeted acquisitions and bank-insurance synergies.
3. Reinforced extra-financial ambition : green finance, inclusion, ESG governance.

2.3 Technical organization : the TITO division and the Digital Center

The *Transformation, Innovation, Technologies and Operations* (TITO) division orchestrates IS modernization, data/AI strategy and *Agile-at-scale* change management. It includes :

— **Transformation Office** : governance of strategic programs and enterprise architecture ;

— **Digital Center** : multidisciplinary « factory » (UX, full-stack, Agile coaches) responsible for delivering new client journeys (*onboarding* 100% mobile, super-app *Attijari Mobile*, open-API…). It also hosts the *Fintech Catalyst* program and an *AI Lab* led by *Attijariwafa Ventures* to industrialize POCs ;

— **Operations & IT** : 24/7 operations, hybrid cloud, data-lake, DevSecOps, RPA.

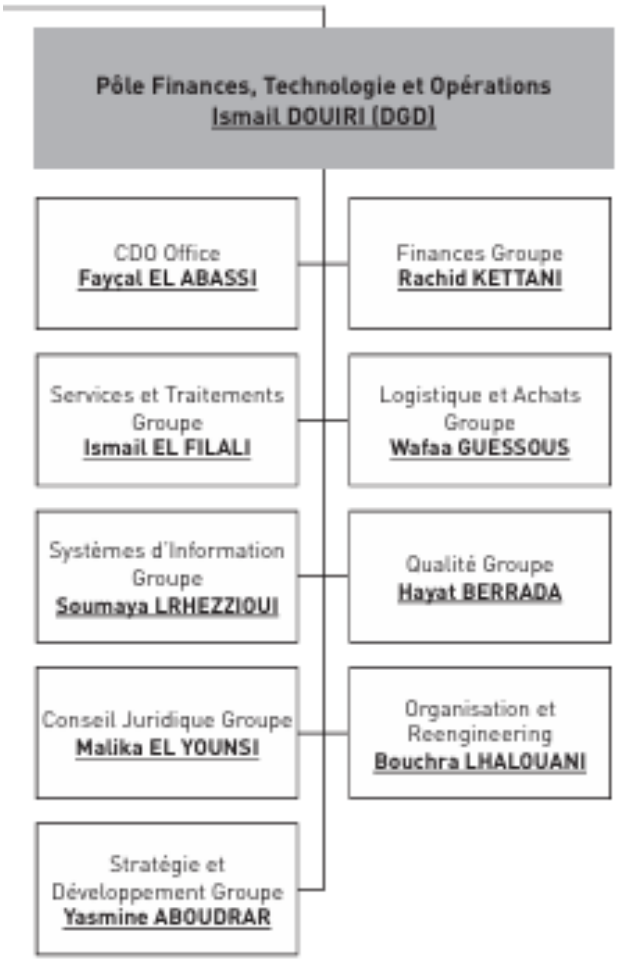


FIGURE 3 – Extract from the organizational chart of the Transformation, Innovation, Technologies and Operations division.

2.4 Cybersecurity challenges

Digitalization exposes AWB to growing threats (targeted phishing, ransomware, data leaks, document fraud). To address this, the Group deploys :

— A state-of-the-art **Security Operations Center** (SOC) and a SIEM correlating security

events *in real time*.

— A *Zero-Trust* strategy : generalized MFA, network micro-segmentation, end-to-end encryption.

— AI fraud detection models (real-time scoring for web, mobile and ATM).

— An awareness program led by CISO **Mehdi Tazi** (Cyber Security Day, e-learning, simulated phishing campaigns).

— Continuous ISO 27001 and GDPR-like compliance, with regular audits and penetration tests.

2.5 Vision and perspectives (2025 horizon)

1. Become the reference pan-African banking platform and serve **30 million clients**.
2. Continue intelligent automation (sovereign cloud, micro-services, generative AI for advisory and compliance).
3. Achieve the **net-zero 2030** objective on direct operations and finance **2 billion €** of green projects, while integrating one million new micro-entrepreneurs.
4. Expand the fintech ecosystem through equity stakes and regional innovation hubs (Dakar, Cairo, Abidjan).

These objectives conclude the *@MBITIONS 2025* plan and pave the way for a 2030 program focused on the *platform-bank* model, asset tokenization and open-finance on the continent.

Solid profitability trajectory. AGR projections (*House View* January and April 2023) anticipate an average annual growth rate of net income attributable to the group (RNPG) of **+7.2%** over 2023-2026E, bringing earning capacity from **7.5 to 9.2 MMDH**. This scenario, supported by Attijariwafa bank's leadership position, justifies an increase in the banking sector weighting in portfolios and validates the relevance of the ambitions set for the 2025 horizon.

3 Problematic, objectives & method

3.1 Research problematic

3.1.1 Context and challenges of banking digitalization

The digital transformation of the banking sector is accompanied by massive digitalization of traditional processes. Client *onboarding* procedures, identity verification (Know Your Customer –

KYC) and credit granting are now carried out mainly remotely, generating a daily flow of **tens of thousands of scanned documents or PDFs**. These documents cover a varied typology : bank statements, official identity documents, pay slips, employment contracts, proof of residence, accounting statements, etc.

This document massification, while improving operational efficiency and client experience, opens the way to new fraudulent risks : selective deletion of amounts, insertion of signatures, reproduction of official stamps, modification of dates, falsification of personal data. Hidden in such a voluminous flow, these manipulations threaten the integrity of banking decisions.

3.1.2 Limitations of traditional detection approaches

Insufficiency of human control Although still constituting the main defense in two-thirds of banks, manual inspection has become an operational bottleneck and a source of vulnerability. This approach presents critical weaknesses :

— **Operational and cost constraints :** Manual processing of thousands of documents is time-consuming, delaying key decisions such as loan granting, and represents a prohibitive cost for limited processing volume.

— **Human reliability and fatigue :** Vigilance declines rapidly (15 % less performance from the first half-hour) and the risk of error can be **multiplied by three when the shift exceeds 12 h** compared to an 8-hour duration. Furthermore, inter-analyst variability frequently oscillates between 20 and 30%, compromising the homogeneity of controls.

— **Overwhelmed by fraud sophistication :** Human expertise is now powerless against modern falsifications. Documentary *deepfakes* and modifications affecting only **less than 1% of a page's pixels** are practically invisible to the naked eye.

Failure of primitive technological solutions The first waves of automation have also proven insufficient to counter increasingly complex threats :

— Traditional OCR, designed for text extraction, is ineffective for analyzing the structural and visual integrity of documents.

— Metadata analysis is easily circumvented by fraudsters, who systematically modify or delete it.

This double failure, human and technological, makes it essential to design a new generation of detection systems, capable of responding to the challenges posed by modern document fraud.

3.1.3 Associated technical and operational challenges

— **Extreme statistical imbalance** : altered pixels are ultra-minority, hence a risk of model bias towards the majority class.

— **Multi-format and multi-resolution robustness** : the system must remain performant on native PDFs, color or B&W scans, and various resolutions.

— **Interpretability and validation** : alerts must be explicitly localized to enable rapid decision-making by Compliance teams.

— **Real-time performance constraints** : detection must be performed without degrading operational deadlines.

Central Problematic

How to design and deploy an automated system capable of accurately detecting and localizing minute modifications in high-resolution documents, while respecting constraints of processing time, diagnostic precision and explainability ?

3.2 Scientific & industrial objectives

3.2.1 General objective

Design an artificial intelligence solution capable of quickly and reliably detecting fraudulent modifications in digitized documents, while remaining explainable and performant.

3.2.2 Scientific objectives

1. Develop a vision model capable of identifying both textual and structural alterations in a document ;
2. Develop robust learning strategies in the face of marked imbalance between falsified and authentic zones ;
3. Produce readable visual outputs so that analysts immediately understand why a document is flagged.

3.2.3 Industrial objectives

Accelerate client journeys and lighten the human control burden, without compromising reliability or traceability of decisions. The solution must remain compatible with common document formats and satisfy compliance and data protection requirements.

Key deliverables

— A reproducible scientific prototype accompanied by code and a dataset conforming to model needs.

— A demonstration interface illustrating detection and localization of falsifications.

— Detailed documentation facilitating transfer to technical teams.

Success criteria

Detection precision, seamless integration into banking workflows, measurable reduction in processing time and compliance with regulatory obligations.

Chapter Conclusion

This first chapter has established the foundations of our study. We have explored the general context of the project and defined the main objectives to be achieved.
The next section will examine the state of the art in falsification detection technologies, which will allow us to justify our technological choices for the continuation.

2
CHAPTER

THEORETICAL FOUNDATIONS & STATE OF THE ART

Chapter Introduction

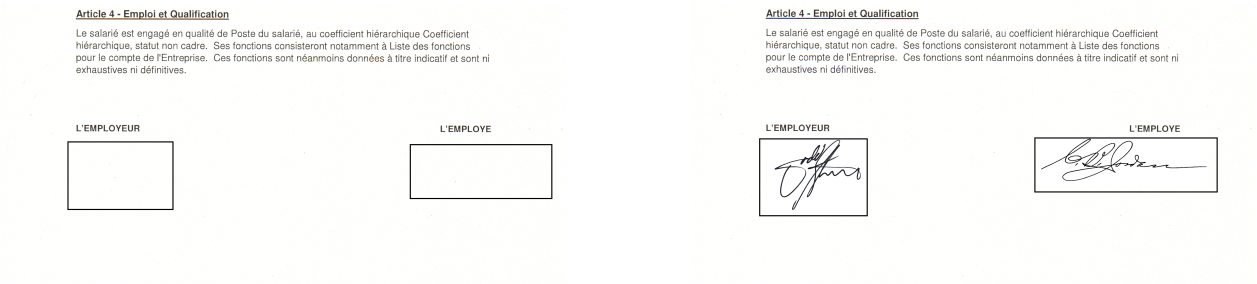
After establishing the strategic importance of fraud detection, this chapter delves into the heart of the technical problem. We will begin by dissecting the anatomy of modern falsifications to understand what we need to detect. We will then lay the theoretical foundations of artificial intelligence in vision, before reviewing state-of-the-art approaches. This critical analysis will allow us to justify the choice of architecture that will serve as the foundation for our solution.

1 Anatomy of falsification

1.1 Alteration typologies

Splicing

Splicing consists of assembling fragments from different documents to create a new falsified document. This technique involves cutting and pasting authentic elements (signatures, stamps, logos, texts) from distinct legitimate sources. The forger can, for example, take the header of an official document, add the body of another document, then insert false information to create an apparently coherent whole. The final result seems authentic because it uses real elements, but their combination creates an entirely falsified composite document.



(a) Original document (b) Falsified document

FIGURE 4 – Splicing process : assembly of elements from distinct sources

Example : pasting a person's photo on a blank passport template. *Composite variant :* assembling the header of a legitimate statement, the body of another and altered amounts.

Copy-Move

Copy-move consists of duplicating a zone of the document and relocating it elsewhere in the same document to hide or modify information. This technique is commonly used to erase troublesome text by covering it with a "clean" part of the same document. For example, one can copy a blank zone of the document and paste it over an amount to modify, or duplicate a signature to place it in another location. Unlike splicing which mixes multiple sources, copy-move only uses elements internal to the original document, which can make the falsification less detectable to the naked eye.

(a) Original document (b) Falsified document

FIGURE 5 – Copy-Move process : duplication of an intra-document region

Retouching / Removal (Inpainting)

Inpainting locally modifies the content of a document by removing or altering elements so that the treated zone appears natural and coherent. This technique uses sophisticated algorithms that "guess" and automatically reconstruct missing content based on surrounding pixels. For example, one can completely erase a digit or letter, and the algorithm will intelligently fill the zone with background texture that harmonizes perfectly with the rest of the document. Unlike copy-move which duplicates existing zones, inpainting generates new content quasi-invisibly, making the falsification particularly difficult to detect.

(a) Original document (b) Falsified document

FIGURE 6 – Inpainting process : local modification by automatic filling

Emerging techniques

— **AI generation** : complete creation of an artificial document with GAN or diffusion models.

— **Documentary deepfakes** : realistic variation of the same document (dates, amounts) produced at large scale.

Comparative synthesis

Typology	**Fraudster's goal**	**Banking example**	**Detection clues**
Splicing	Assemble legitimate pieces	Photo pasted on passport	Noise breaks, lighting, double compression
Copy–Move	Duplicate or hide	Stamp copied to another page	Pairs of identical regions, cloned texture
Retouching	Erase or modify subtly	Transform « 2024 » → « 2025 »	Zone too smooth, weakened noise, blurred contours
AI Generation	Create complete fake	Credible synthetic statement	Lack of natural noise, global inconsistencies

TABLE 2 – Comparative synthesis of main textual alteration typologies and corresponding detection indices

1.2 Physical & statistical signatures

Even the most skillful retouching disturbs the internal balance of the file. These « fingerprints » are distributed across several levels :

1.2.1 Pixel-level traces

Key idea : An authentic image has a natural and coherent digital texture. A retouch is like grafting a piece of skin : even if the color is good, the texture betrays the operation.

In an original image, neighboring pixels have very similar colors and brightness, creating predictable statistical relationships. A copied-pasted or retouched zone breaks this harmony. Mathematical tools (like Fourier analysis or wavelets) act as « developers » that highlight these breaks in the image texture, where the eye sees nothing.

1.2.2 Lighting and perspective anomalies

Key idea : In a real scene, everything is governed by the same physical laws. An added object will not necessarily respect these laws.

If a photo is taken with a light source coming from the left, all shadows must project to the right. An inserted object with a poorly oriented shadow is an indication of falsification. Similarly, perspective (the way distant objects appear smaller) must be coherent for the entire image. An added element may seem to « float » or have an incorrect angle relative to the rest of the scene.

1.2.3 Compression artifacts

Key idea : JPEG compression, used by most cameras and scanners, leaves a specific signature on the entire image. Modifying the image disturbs this signature in a detectable way.

— **Misaligned JPEG blocks :** To compress, JPEG divides the image into an invisible grid of 8x8 pixel blocks. If you copy part of an image and paste it onto another, the grid of the pasted part will almost never be perfectly aligned with the background image grid. This « break » in the grid is formal proof of manipulation.

— **Heterogeneous quality (ELA analysis) :** Each time an image is saved in JPEG, it loses a little quality. Error analysis (ELA) allows visualizing the « wear levels » of the image. A zone that has been modified and re-recorded will have a different wear level from the rest of the image, which will make it stand out during ELA analysis.

1.2.4 Noise inconsistencies (PRNU)

Key idea : Each camera or scanner has a unique and invisible « fingerprint » that it imprints on each photo.

Due to tiny manufacturing imperfections, a device's sensor leaves a noise pattern (called PRNU) that is unique to that device. It's its signature. To detect fraud, the PRNU is extracted from the image. If a zone of the image doesn't have this PRNU (or has a different one), it means it was added from another source. This is one of the most reliable techniques because it's almost impossible to fake this sensor noise.

2 Deep Learning foundations for computer vision

2.1 ML vs DL for images

The distinction between traditional machine learning and deep learning marks a true paradigmatic break in image analysis.

Traditional ML approach This methodology relies on **manual feature extraction** (*feature engineering*). The system's performance depends on an expert's ability to design relevant descriptors (e.g., color histograms, textures via Gabor filters, SIFT interest points). These characteristics are then submitted to a classifier (SVM, Random Forest). This approach, while powerful, presents a major bottleneck : its effectiveness is intrinsically limited by the relevance of predefined characteristics, a laborious and often sub-optimal process in the face of falsification diversity.

DL (Deep Learning) approach Deep Learning, conversely, **automates representation learning**. By relying on deep neural networks, the model learns the most discriminating characteristics directly from raw pixels. It builds a **hierarchical representation** of the image : the first layers detect simple patterns (contours, gradients), while subsequent layers assemble them to form increasingly abstract concepts (textures, shapes, and *ultimately*, manipulation artifacts). This end-to-end learning capability explains its superiority and adoption as the de facto standard in computer vision.

2.2 Architectures : CNN, RNN, ViT

Deep Learning's power is embodied in specific architectures, designed to exploit the spatial properties of images.

CNN (Convolutional Neural Networks) CNNs are the **canonical** architecture for vision. Their fundamental building block, the **convolution layer**, applies a set of filters on the image to create feature maps. This mechanism is powerful because it respects two key principles : it operates locally (a neuron is only connected to a small region, its **receptive field**) and it shares filter weights across the entire image, making it efficient and capable of detecting a pattern regardless of its position. They excel in identifying local artifacts (texture breaks, compression traces).

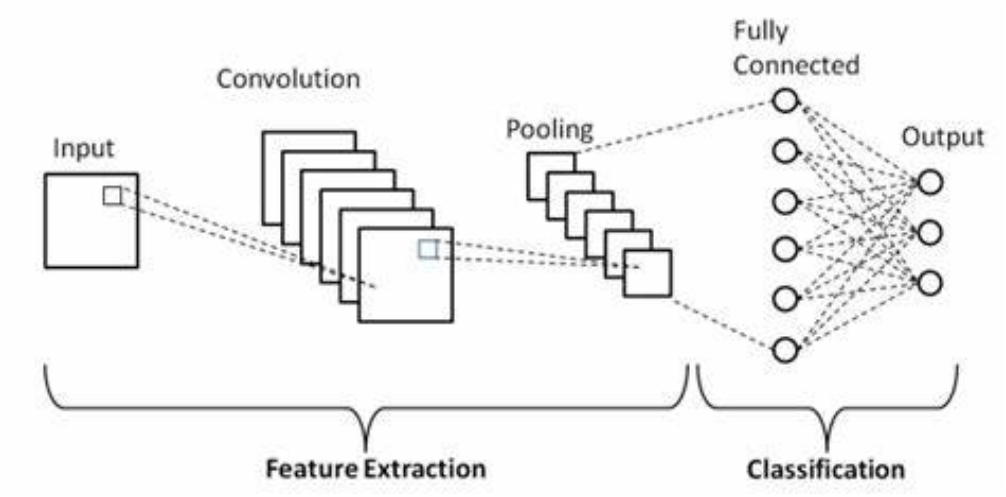


FIGURE 7 – Example of CNN pipeline for local artifact detection.

RNN (Recurrent Neural Networks) Designed for **sequential data**, RNNs are less central in static image analysis. Their use remains relevant for specific tasks like tracking contours of an altered zone or analyzing sequential metadata, but they are rarely at the heart of modern falsification detection systems.

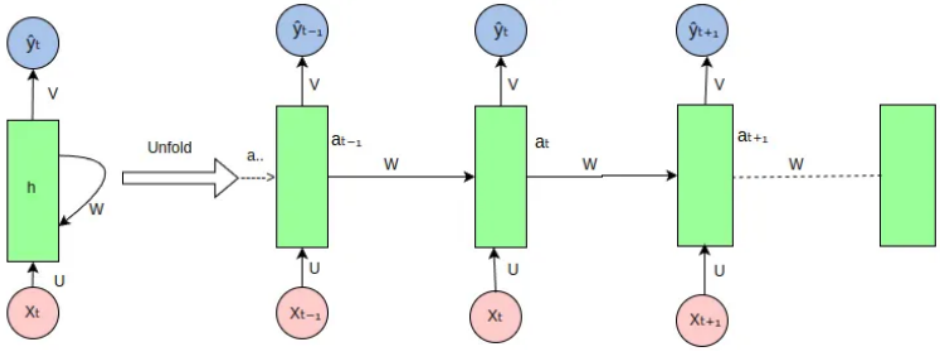


FIGURE 8 – Principle of an RNN applied to a sequence of image patches.

ViT (Vision Transformers) ViTs represent a powerful alternative to CNNs. They divide the image into a sequence of *patches* and, thanks to **self-attention mechanisms**, evaluate relationships between each patch and all others. Their major asset is a **global context understanding** of the image, allowing them to detect long-distance inconsistencies (e.g. non-uniform global lighting). This performance gain comes at a cost : ViTs are data and computational resource hungry for training.

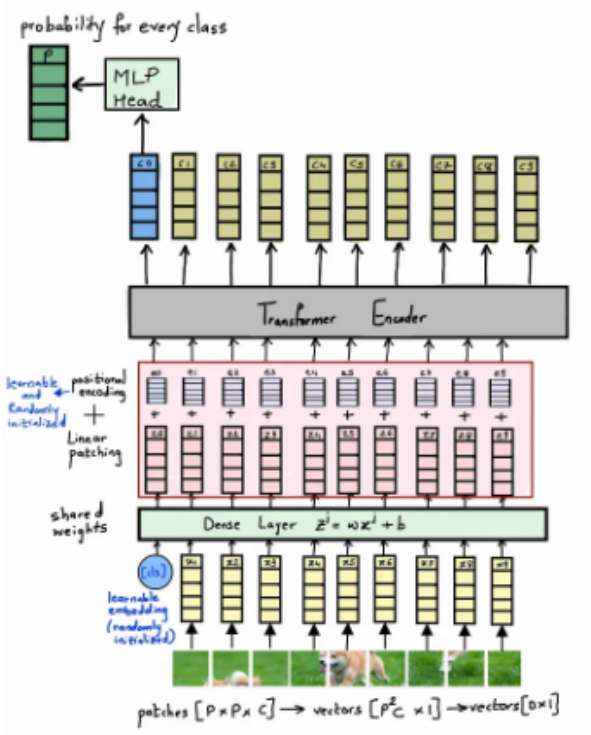


FIGURE 9 – Patch division and self-attention scheme of a Vision Transformer.

2.3 Optimization & regularization

Training a Deep Learning model is a balancing exercise. One must find the best parameters for the task (optimization) while ensuring that the model remains capable of generalizing to unknown data (regularization).

Optimization

This process adjusts the network weights (θ) to minimize the loss function $J(\theta)$. The most common **optimizer** algorithms are **SGD** (with *momentum* γ to smooth convergence, where weight update θ is done with velocity v_t and learning rate η : $v_t = \gamma v_{t-1} + \eta \nabla_\theta J(\theta)$ and $\theta \leftarrow \theta - v_t$) and **Adam**, an adaptive optimizer often preferred for its fast convergence.

Regularization

This set of techniques prevents **overfitting**, the phenomenon where a model memorizes the training dataset at the expense of its ability to generalize.

— **Dropout :** Randomly deactivates neurons at each training step, forcing the network to develop more robust representations less dependent on each other.

— **Batch Normalization :** Stabilizes and accelerates training by normalizing activations within each mini-batch.

— **Data Augmentation :** The most effective regularization technique in vision. It consists of artificially increasing the dataset size by applying realistic transformations to images : rotations, zooms, but also simulation of folds, stains, lighting variations or scanning, to make the model robust to real-world imperfections.

3 Overview of existing segmentation approaches

3.1 Segmentation : losses & metrics

For document fraud, knowing that an image is fake is insufficient ; the manipulation must be localized. The problem is therefore treated as a **semantic segmentation** task, where the objective is to assign each pixel a class : "authentic" or "manipulated". The success of this task crucially depends on the choice of loss function and evaluation metrics.

Loss Functions

The major challenge being **class imbalance** (very few manipulated pixels), loss functions must be chosen carefully to prevent the model from ignoring the minority class.

— **Binary Cross-Entropy (BCE) :** The basic loss for pixel-by-pixel binary classification. For a ground truth $y \in \{0, 1\}$ and a prediction $\hat{y} \in [0, 1]$, it is defined by :

$$L_{\text{BCE}} = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})] \tag{1}$$

It is often weighted to give more importance to the "manipulated" class.

— **Dice Loss :** Very popular in segmentation, it directly maximizes the overlap between prediction P and ground truth G.

$$L_{\text{Dice}} = 1 - \frac{2|P \cap G|}{|P| + |G|} \tag{2}$$

— **Tversky Loss :** A generalization of Dice Loss that offers fine control over the precision/recall trade-off by penalizing false positives (FP) and false negatives (FN) differently via parameters α and β. This is a major asset when the business cost of a false negative is much higher than that of a false positive.

$$L_{\text{Tversky}} = 1 - \frac{\text{TP}}{\text{TP} + \alpha\text{FN} + \beta\text{FP}} \tag{3}$$

— **Focal Loss :** Improves BCE by adding a modulation factor $(1 - p_t)^\gamma$ that reduces the impact of easy examples and forces the model to focus on difficult cases, where p_t is the model's probability for the correct class.

$$L_{\text{Focal}} = -\alpha_t(1 - p_t)^\gamma \log(p_t) \tag{4}$$

Evaluation Metrics

They quantify the model's performance by focusing on its ability to precisely localize fraud.

— **Intersection over Union (IoU) :** The reference metric in segmentation, it measures the degree of overlap between prediction and ground truth.

$$IoU = \frac{|P \cap G|}{|P \cup G|} \tag{5}$$

— **F1-Score :** The harmonic mean of precision and recall, it provides a balanced measure, particularly reliable in case of class imbalance.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{8}$$

3.2 Algorithms : U-Net, DeepLab, SegFormer, hybrids

Foundational blocks of segmentation

— **U-Net** : symmetric encoder-decoder architecture where the encoder captures semantic context by reducing spatial resolution, while the decoder restores precise localization by climbing back to original resolution. The *skip-connections* directly transfer feature maps from the encoder to the corresponding decoder, avoiding loss of spatial information. This fusion allows the

network to combine high-level features (*what*) with low-level details (*where*), making U-Net particularly effective for precisely segmenting object boundaries. The architecture remains a reference because it perfectly balances generalization capacity and fine detail preservation.

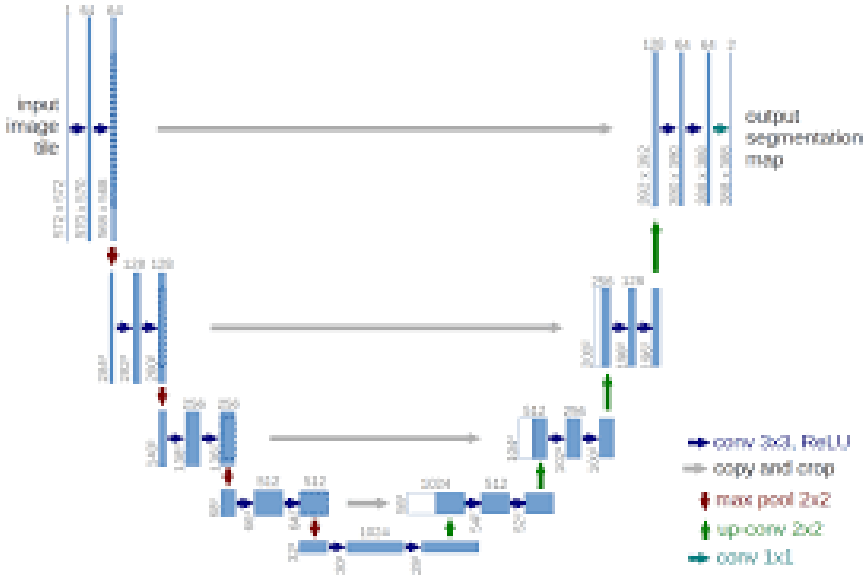


FIGURE 10 – Simplified U-Net architecture : contracting encoder (left) and expansive decoder connected by *skip connections*.

— **DeepLabv3+** : architecture that uses dilated convolutions (*atrous*) to capture context at different scales without losing spatial resolution. The ASPP (*Atrous Spatial Pyramid Pooling*) module applies parallel convolutions with different dilation rates, allowing simultaneous capture of fine details and global context. A lightweight decoder then fuses these rich semantic features with contour information from low-resolution encoder layers. This combination produces particularly sharp and precise segmentation boundaries, surpassing traditional approaches in object delimitation.

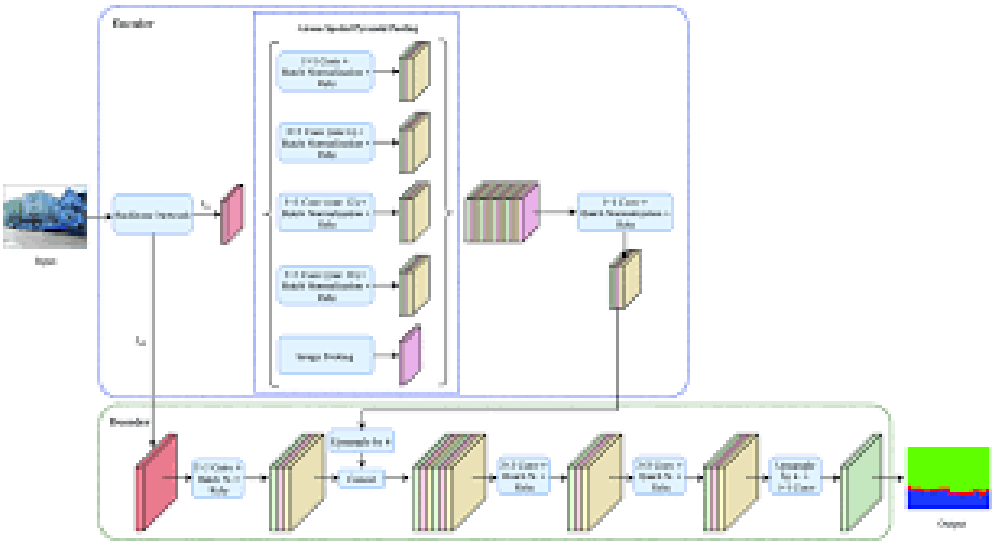


FIGURE 11 – DeepLabv3+ scheme : dilated convolution encoder (*atrous*) and ASPP module, followed by a lightweight decoder.

— **SegFormer** : innovative architecture that replaces traditional CNNs with a hierarchical Transformer encoder capable of processing images at different resolutions without complex positional encoding. Multi-head self-attention efficiently captures long-term dependencies and global image context, enabling rich semantic understanding of the scene. The decoder uses only lightweight MLP (*Multi-Layer Perceptron*) layers to fuse multi-scale features and generate high-resolution segmentation masks. This Transformer-based approach offers a powerful alternative to classic CNN architectures while being more computationally efficient thanks to the absence of heavy positional encoding.

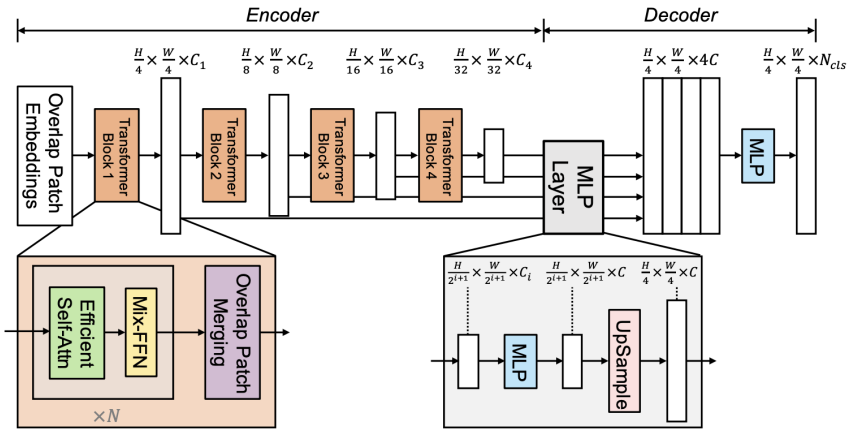


FIGURE 12 – SegFormer overview : hierarchical Transformer encoder (patches) and lightweight MLP decoder.

3.3 Advanced research trends

1. *Frequency exploitation* : bimodal architectures (DTD 2023, FFDN 2024) directly reading DCT coefficients or wavelets to spot compression scars.
   - The **DTD (Dual-domain Transformer)** model uses a two-branch Transformer architecture (spatial and frequency/DCT). Its key innovation is a cross-domain attention mechanism that allows each domain to inform the other at multiple network levels, thus improving detection of subtle correlations between visual and statistical artifacts.
   - **FFDN (Frequency-aware Forgery Detection Network)** goes further by using specialized neural blocks, designed to model the statistical distribution of DCT coefficients. It focuses on identifying sharp breaks in compression signatures (e.g. via estimated JPEG quality maps), which is a very reliable indicator of *splicing*.
2. *Structural / semantic analysis* : Graph-OCR + GNN (2022) or DCLNet (2025) model the document as a word graph to detect layout inconsistencies.
   - **Graph-OCR + GNN (2022)** approaches established the viability of this pipeline by training GNNs (typically Graph Attention Networks) in a supervised manner to classify document graphs. The graph edges represent spatial relationships (proximity, alignment) that are broken by manipulation.
   - **DCLNet (Document Contrastive Learning Network, 2025)** constitutes a major advance by applying self-supervised contrastive learning. It pre-trains the GNN to distinguish structural variations of the same document from those between different documents. This method allows learning very robust layout representations with little or no labeled falsification data, solving a major bottleneck.
3. *Multimodal explainability* : FakeShield (ICLR 2025) couples Vision Transformer and LLM to provide mask + textual explanation directly understandable by an analyst.
   - **FakeShield**'s architecture is designed end-to-end. The **Vision Transformer** doesn't just classify the image ; its internal attention maps are used to generate the fraud localization mask.
   - The central contribution is the coupling module that projects visual features extracted by the ViT into the semantic space of the **LLM**. The LLM is thus "conditioned" by visual evidence to generate relevant technical explanation (e.g. : "compression noise

inconsistency", "abnormal borders"), transforming a "black box" into an interactive analysis tool.

These axes show increasing specialization : from simple pixel-by-pixel mask towards multi-frequency, structured and explained understanding.

3.4 Benchmarks

Reference datasets

— **CASIA**, **COVERAGE** : historical, natural image-oriented ; useful for validating low-level robustness but not representative of documents.

— **DocTamper** (2023) : 170,000 synthetic pages covering copy-move, splicing and inpainting on PDF ; current reference for document fraud segmentation evaluation.

Observed limitations

— *Gap with reality* : variable scan quality, messaging-related artifacts, stains and folds absent from academic datasets.

— *Public corpus shortage* : confidentiality (GDPR, KYC) prevents sharing of real banking documents, hampering open-source research.

3.5 Industrial solutions

— **Ocrolus Detect** : AI + human review (*human-in-the-loop*) dedicated to finance.

— **Resistant AI – Documents** : combined analysis of PDF binary and visual rendering.

— **Inscribe AI** : risk score + natural language synthesis.

— **Mitek Digital Fraud Defender**, **Jumio Doc Proof**, **Klippa DocHorizon** . . . : identity verification extended to supporting documents.

These offerings are accessible via API, provide a score and *heat-map*, but generally remain poorly customizable and weakly explainable black boxes.

3.6 Chosen Technology : Justification for SegFormer Adoption

After a comparative evaluation phase including reference architectures like **UNet** and **DeepLabV3+**, the technical choice for the segmentation module was **SegFormer**. This decision is not

based solely on its results during our tests, but on a strategic analysis of its capabilities to meet current and future project needs at **AWB**.

1. **Superior Performance on Non-Local Falsifications** : During our benchmarks, SegFormer demonstrated superior segmentation performance, particularly on complex falsification cases. Its self-attention mechanism, which builds global document understanding, proved particularly effective for detecting non-local manipulations where purely convolutional models (CNN), due to their limited receptive field nature, showed their limits.
2. **Computational Efficiency for Deployment** : Although based on Transformer architecture, SegFormer is designed to be lightweight and efficient. It offers one of the best performance/computational cost ratios, making it compatible with realistic industrial deployment constraints without requiring prohibitive hardware infrastructures.
3. **Evolution Potential and Strategic Flexibility** : The choice of SegFormer is also strategic. Its architecture constitutes an ideal foundational brick for evolving our solution in phase with advanced research trends, notably via future addition of a frequency analysis stream (bimodal approach).

Chapter Conclusion

This chapter has built the technical foundation of our project. Starting from the anatomy of falsifications, we identified specific traces to detect. Exploration of computer vision principles and state-of-the-art architectures then allowed us to select SegFormer as the most relevant technology, offering an optimal balance between performance and efficiency.

3
CHAPTER

SPECIFICATIONS, PLANNING & TARGET ARCHITECTURE

Chapter Introduction

The previous chapter justified our choice of SegFormer as the technological foundation. It is now time to translate our vision into a concrete engineering project. This chapter establishes the complete specification of the solution : we will define functional and non-functional requirements, analyze the technical and organizational constraints that frame our work, and present the planning adopted to successfully complete this project within the allocated timeframe.

1 Needs & requirements

1.1 Functional requirements

User Management & Authentication

ID	Description
BF-AUTH-01	Registration : validation (name, unique email, strong password) + confirmation
BF-AUTH-02	Complete authentication : login, refresh, password recovery, logout
BF-AUTH-03	User sessions via JWT

Table 3 – Requirements : Users & Authentication

Document Management

ID	Description
BF-DOC-01	Multi-file upload (PDF, PNG, JPG, TXT) : size limit + validation
BF-DOC-02	Organization by type (contracts, invoices, statements) with metadata
BF-DOC-03	Secure storage : encryption + automatic backups
BF-DOC-04	History and traceability of uploaded documents
BF-DOC-05	"Single document" or "document comparison" mode

Table 4 – Requirements : Document management

Visual Detection Pipeline (SegFormer)

ID	Description
BF-DETECT-VIS-01	Visual analysis to detect falsified zones
BF-DETECT-VIS-02	Generation of segmentation masks for suspicious regions
BF-DETECT-VIS-03	Calculation of confidence score per detected zone
BF-DETECT-VIS-04	Precise localization of falsified coordinates

TABLE 5 – Requirements : Visual detection pipeline (SegFormer)

Text Extraction & Analysis Pipeline

ID	Description
BF-OCR-01	Text extraction via *LLMWhisperer* (multilingual OCR)
BF-OCR-02	Preservation of original structure and layout
BF-ANA-01	Text analysis with *Llama* to detect inconsistencies
BF-ANA-02	Identification of suspicious elements (dates, amounts, names, signatures)
BF-ANA-03	Detection of logical inconsistencies

TABLE 6 – Requirements : Text extraction & analysis

Document Comparison Pipeline

ID	Description
BF-COMP-01	Acceptance of two documents : original + presumed falsified
BF-COMP-02	Text extraction from both docs via *LLMWhisperer*
BF-COMP-03	Comparison of extracted texts with *Llama*
BF-COMP-04	Identification & localization of textual differences
BF-COMP-05	Detailed report of detected modifications

TABLE 7 – Requirements : Document comparison

Orchestration & Result Combination

ID	Description
BF-ORCH-01	Coordination of different pipelines
BF-ORCH-03	Calculation of global falsification score
BF-ORCH-04	Prioritization according to confidence level
BF-ORCH-05	Management of conflicts between detection methods

TABLE 8 – Requirements : Orchestration and result combination

Interface & Visualization

ID	Description
BF-UI-01	Choice of analysis mode by user
BF-UI-02	Display of masks overlaid on document
BF-UI-03	Highlighting of textual differences
BF-UI-04	Dashboard with coordinates of falsification zones.

TABLE 9 – Requirements : Interface and visualization

1.2 Use Case Modeling

To model interactions and provide a clear overview of system functionalities, the use case diagram below has been developed. It identifies the main actor, the "User", as well as the major specialized subsystems ("Authentication System", "OCR System", "Detection System") with which they interact indirectly. The diagram details central use cases, such as document submission and analysis, and highlights the architecture's modularity through the different services involved.

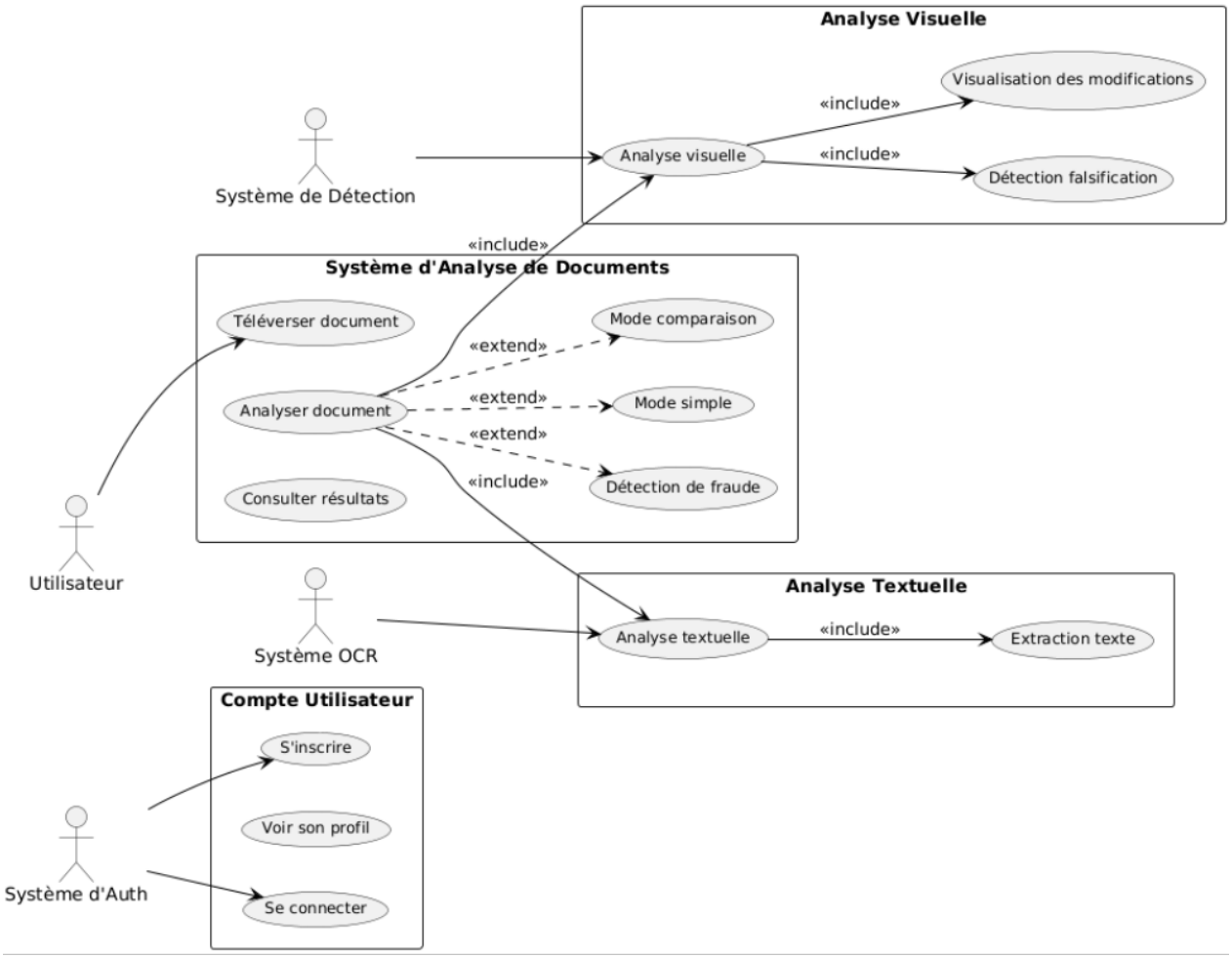


FIGURE 13 – Use case diagram of the document analysis system.

1.3 Non-functional requirements

Performance & Efficiency

— Complete processing < 2 min (SegFormer < 10 s)

— OCR > 95 %

— Optimized API calls and simultaneous processing supported

Reliability & Robustness

— Fault tolerance and backup mechanisms

— Automatic recovery of interrupted processing

Precision & Quality

— High SegFormer Dice Score

— OCR structural fidelity

— Explainable & coherent combined results

Scalability & Load Handling

— Load adaptability ; GPU-CPU distribution / optimization

— Efficient memory management

Security & Confidentiality

— Document & communication encryption

— Strict file validation and traceability

Integration & Interoperability

— Seamless integration (SegFormer / LLMWhisperer / Llama)

— API evolution support and standard format compliance

Availability & Maintenance

— Availability > 99 %

— Maintenance without interruption

2 Constraints & success indicators

To guarantee project alignment with the organization's strategic objectives and ensure its viability, it is imperative to clearly define the framework within which it must operate. This section details the technical, regulatory and organizational constraints that govern the project, as well as the success indicators and performance objectives that will measure its success.

2.1 Technical constraints

— **Computing resources :** The selected Deep Learning models, particularly `SegFormer` and `LLaMA-70B`, require significant computing power for their training and inference phases. The project must work with defined hardware resources, which imposes strategic choices on training times, batch sizes and model optimization for efficient execution. **Faced with this constraint, we opted to use the `Kaggle` cloud platform for intensive training phases, giving us access to powerful GPU accelerators (NVIDIA Tesla T4) essential to our experimental approach.**

— **Service latency :** For the solution to be practical, a maximum response time is defined. The `SegFormer` model inference on CPU must execute in less than 10 seconds. The total processing time, combining all steps (preprocessing, analysis by different models, and API response), must not exceed two minutes.

— **Application security :** Application security must be based on fundamental principles. This implies robust authentication and password hashing (PBKDF2-SHA256). Data management requires multi-level security : strict validation of uploaded files (UUID names, controlled types), restrictive CORS policy, and input validation on critical APIs. Finally, the system must handle errors in a non-revealing manner.

2.2 Regulatory constraints

— **Personal data confidentiality :** Handling client documents subjects us to strict regulations. The project must be in full compliance with Moroccan law n°09-08 relating to the protection of natural persons with regard to the processing of personal data.

— **Auditability and traceability :** Decisions made by the system must be traceable. To achieve this, each analysis (result, document identifier, date) as well as the version of models used are systematically logged. In accordance with the target architecture, these logs are stored in the project's MongoDB database to guarantee a reliable audit trail.

— **Data sovereignty :** Unless explicitly authorized and supervised, sensitive data must not transit or be stored on servers located outside national territory.

2.3 Organizational constraints

— **User support :** To facilitate future tool adoption, particular attention must be paid to its ease of use. It is constrained by the need to provide a clear interface and sufficient documentation to demonstrate its added value compared to a standard process.

— **Inability to access production data :** This is the most structuring constraint of the project. Due to the high sensitivity of client information and very strict application of regulatory constraints, the bank was unable to provide an internal dataset.

— **Scarcity of alternative public data :** Consequently, the project is constrained to rely exclusively on public datasets. However, public datasets dealing specifically with identity or banking document falsification are extremely rare and often of limited size.

2.4 Target KPI / SLA

For the project to be considered successful and to consider its industrialization, the following qualitative indicators and performance objectives are targeted :

— **Detection recall :** Prioritize very high recall to identify virtually all falsifications, even if it means accepting more false positives.

— **Detection precision :** Maintain sufficient precision to avoid overloading with unnecessary alerts, without compromising the primary objective of maximum fraud coverage.

— **Efficiency gain :** Demonstrate a significant reduction in the overall time needed to analyze a document.

Service Level Agreement (SLA) objectives :

— **System responsiveness :** Maintain fast response times for document analysis, in accordance with thresholds defined in technical constraints.

— **Service availability :** Ensure high application availability during testing and demonstration phases.

3 Planning and project management

3.1 Adopted project management methodology

For this project, an **Agile** methodology, inspired by the **Scrum** framework, was adopted. This method was favored for its flexibility and ability to adapt to the exploratory nature of artificial intelligence development, where experimentation and continuous adjustment are essential to success.

The project was broken down into a series of **sprints**, short iterative development cycles, each aimed at producing a functional increment and validating technical hypotheses.

The project flow was organized as follows :

1. **Sprint 0 - Scoping and Initialization :** This first phase consisted of conducting context analysis (Chapter 1) and state of the art (Chapter 2). The objective was to define the product vision and constitute an **initial backlog** of functionalities to develop, based on defined requirements (Chapter 3).
2. **Iterative Development Sprints :** The project core (Chapters 4 and 5) was realized through several successive sprints. Each iteration focused on specific objectives :
   - Setting up basic architecture (Flask server, MongoDB database, React application skeleton).
   - Development of data processing pipeline and first implementation of basic **SegFormer** model.
   - Analysis of first results and visual model improvement (implementation of custom loss function and enhanced detection head).
   - Integration of semantic analysis pipeline (OCR with LLM Whisperer and analysis by **LLaMA-3**).
   - Development of result visualization interfaces and prototype finalization.
3. **Continuous validation and feedback loop :** It was precisely this iterative approach that first allowed us to identify *SegFormer* as the most promising architecture, then refine it sprint after sprint. At the end of each iteration, we evaluated model performance and new functionalities, which fed the backlog and reoriented priorities for the following sprint.

This Agile management allowed navigating the technical uncertainties inherent to the project

while guaranteeing constant progression towards the target solution, validated at each stage of its development.

3.2 Project flow : Gantt chart

The project took place over a five-month period, from early February to early July 2025. Planning was followed using a Gantt chart to ensure rigorous management of time and deliverables, as illustrated in Figure 14. This schedule was broken down into several main phases, each with clear objectives and deliverables. Report writing was not an isolated final phase, but a continuous process conducted in parallel with technical work, with an intensification phase at the end of the project.

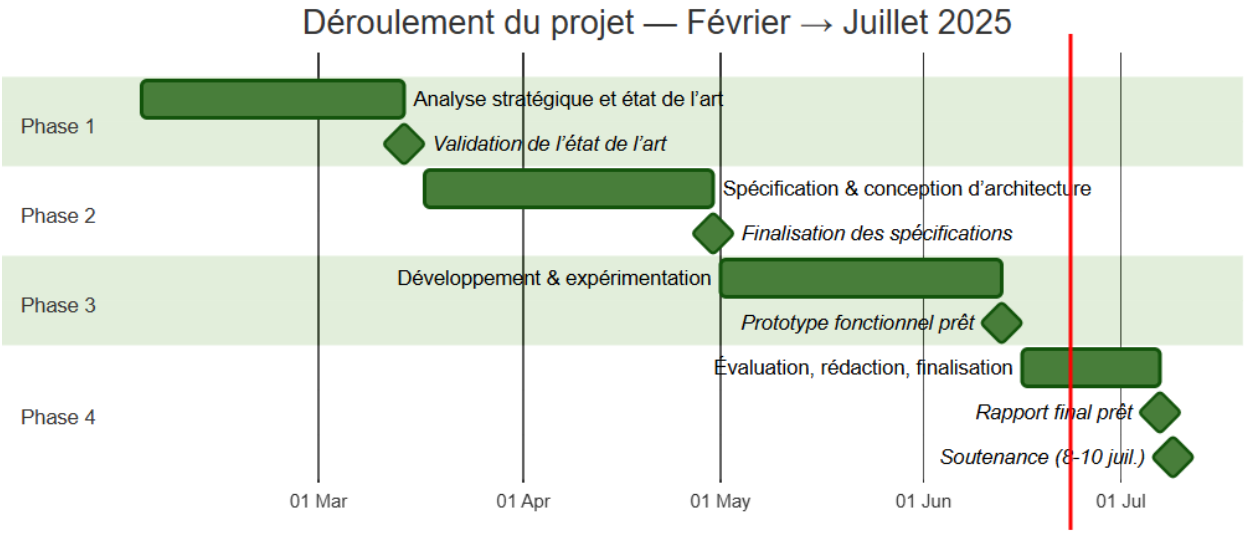


FIGURE 14 – Gantt chart illustrating project flow.

3.3 Risk management

Proactive risk management was conducted throughout the project to identify potential obstacles and implement mitigation strategies. This approach was essential to secure the schedule and guarantee deliverable quality. The major risks identified and their mitigation plans are presented below.

Risk 1 : Production data unavailability *Description and Impact :*

The most critical risk was the inability to access real client documents for confidentiality and regulatory reasons. Without data, AI model training was impossible, threatening the very feasibility of the project.

Mitigation Strategy :

The strategy consisted of building a composite and heterogeneous dataset by aggregating several public and academic sources (RVL-CDIP, SUPATLANTIQUE, Roboflow) and adding manual fraud simulations. This approach allowed creating a rich and diversified training corpus while respecting confidentiality constraints.

Risk 2 : Insufficient AI model performance *Description and Impact :*

There was a risk that standard models would not be performant enough to detect subtle falsifications or to understand the complex context of documents, which would have made the solution ineffective.

Mitigation Strategy :

For the visual component, a **SegFormer** architecture was significantly improved with a custom detection head and loss function. For the semantic component, the choice was made for a state-of-the-art model (**LLaMA-3**) driven by rigorous prompt engineering to guarantee factual and structured analyses.

Risk 3 : Computing resource limitations *Description and Impact :*

Deep learning model training is very resource-intensive on GPUs. Insufficient computing power could have limited the scope of our experiments and considerably slowed down the project.

Mitigation Strategy :

Using the **Kaggle** cloud platform gave us access to powerful GPU accelerators (NVIDIA T4), with a free quota of approximately **30 hours of GPU computing per week**. Additionally, optimization techniques, such as mixed precision training, were applied to reduce memory footprint and accelerate training cycles.

Risk 4 : Bias and result reliability *Description and Impact :*

An inherent risk in AI is that the model produces unreliable results, such as false alerts (low precision) or, more seriously, missed frauds (low recall).

Mitigation Strategy :

A deliberate strategy was implemented to optimize the visual model in favor of recall, to minimize the risk of missing fraud. This bias was assumed and documented, and the application prototype was designed to present results clearly to a human analyst, who keeps control of the final decision.

Chapter Conclusion

This chapter has formalized our project framework. By precisely defining functional and non-functional requirements, identifying structuring constraints — notably the absence of production data — and establishing clear planning, we now have a robust roadmap for the development phase. This specification work guarantees that the solution we will build will meet measurable and realistic objectives.

4
CHAPTER

SYSTEM ARCHITECTURE DESIGN

Chapter Introduction

After defining the specifications and planning in the previous chapter, we now enter the technical design phase. This chapter presents the detailed plan of our solution's software architecture. We will justify our strategic choices, from microservices adoption to technology stack selection. We will then model data flows, component interactions via UML diagrams, and finish with the precise design of the data model that will support the entire application.

1 Strategic Architectural Principles and Choices

1.1 Adoption of a Microservices Architecture

The system's global architecture is based on two guiding principles : (i) service decoupling and (ii) processing asynchronism. This philosophy responds to the business need to execute long and costly AI analyses without penalizing the user with interface blocking.

The client-server model therefore implements asynchronous processing that handles two use cases :

— Simple Mode : in-depth analysis of a single document ;

— Comparison Mode : detection of divergences between two documents.

1.2 Technology Stack Selection and Justification

The solution's architecture relies on a set of open-source technologies recognized for their flexibility and performance. The choice was made for the main technological pillars presented below.

React : a JavaScript library maintained by Meta, leader in creating interactive and reactive user interfaces. Its component-based approach allows building complex and dynamic web applications in an organized manner.

FIGURE 15 – React, the library for user interface (frontend).

Flask : a web micro-framework written in Python, renowned for its lightness and minimalist approach. It is ideal for building performant and custom RESTful APIs without imposing a rigid structure.

FIGURE 16 – Flask, the micro-framework for the application server (backend).

MongoDB : a document-oriented NoSQL database management system. It stores data in a flexible format similar to JSON, making it perfectly suited for storing unstructured or semi-structured data, such as complex results from AI models.

FIGURE 17 – MongoDB, the NoSQL database.

Groq : an inference platform specialized in ultra-fast execution of large language models (LLM). Its unique hardware architecture (LPU) enables quasi-instantaneous responses, which is crucial for our semantic analysis pipeline.

FIGURE 18 – Groq, the inference engine for AI.

2 Application Architecture and Data Flow

2.1 Architecture Overview

The general flow is identical (upload → file save → background thread), but the worker logic differs according to the chosen mode.

1. The operator selects a mode and submits their documents.
2. The Flask API validates the request, records the task (status `uploaded`) in MongoDB, then immediately launches processing in a separate thread.
3. Flask responds `202 Accepted`; the React interface is therefore freed without delay.
4. The worker updates the status (`processing`), executes the complete pipeline, then marks the task `completed`.
5. React periodically queries the API and, upon receiving `completed` status, retrieves the final report to display it.

Figure 19 illustrates this asynchronous mechanism common to both modes.

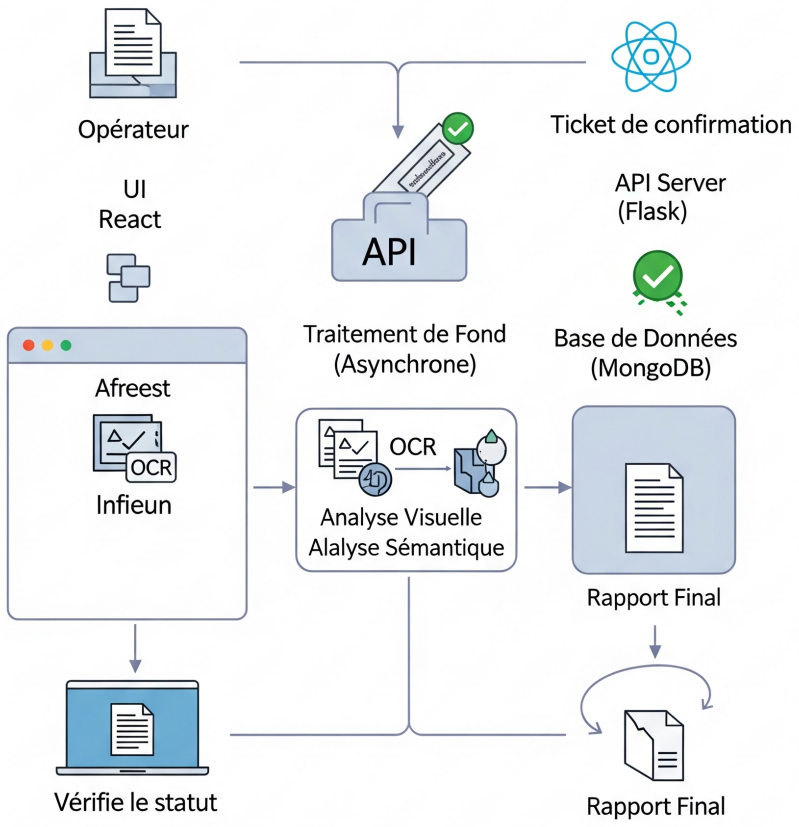


FIGURE 19 – System overview and data flow.

2.2 Detailed Microservices Decomposition

Within this project framework, these technologies are orchestrated so that each block fulfills a precise strategic role :

User interface (React) – Asynchronous control center Real-time dashboard ; each document is a component managing its own status (*In progress*, *Completed*, *Error*). Responsible for : secure capture, score/heatmap display, post-analysis chat.

Application server (Flask) – Lightweight orchestrator Exposes a REST API ; manages JWT, validation, task recording (`uploaded`) and delegation to worker. Frees itself immediately to accept other requests.

AI analysis engine – Specialized toolkit Suite of services called in sequence : LLM Whisperer (OCR), SegFormer (segmentation), LLaMA (reasoning), RAG for specific business rules.

Database (MongoDB) – Flexible memory Natural storage of semi-structured JSON reports : extracted texts, fraud scores, visualizations, logs. Collections for : users, document metadata, analysis reports.

2.3 Processing Pipeline Design

The processing pipeline executed by the worker adapts to the selected mode.

2.3.1 "Simple Analysis" Mode Process

1. **OCR :** complete text extraction.
2. **Visual detection :** SegFormer identifies manipulated zones.
3. **Document typing :** classification (statement, contract, ...).
4. **Content analysis :** business rules (RAG) or LLaMA for inconsistencies.

The result is a detailed legitimacy report, saved in database.

2.3.2 "Comparison" Mode Process

1. **OCR** on both documents.
2. **Semantic comparison :** LLaMA (Groq API) identifies additions/deletions.

We obtain a textual differences report, also stored.

3 Detailed System Modeling (UML)

To complete the textual description of the architecture, this section presents a series of UML sequence diagrams. These diagrams illustrate the dynamic interactions between different system components to realize the main use cases. They allow precise visualization of call flows, data exchange and operation chronology.

We have grouped them into two categories : flows related to user management and flows at the heart of the application's business logic.

3.1 Authentication and User Management Flows

Secure access management is a fundamental building block of the application. The following diagrams detail the registration, login and logout processes.

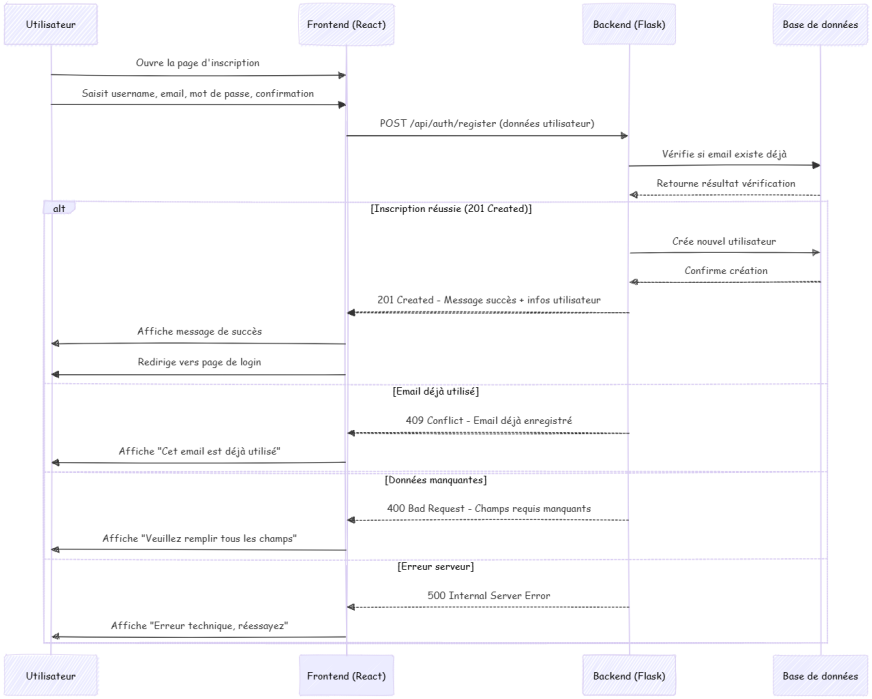


FIGURE 20 – Sequence diagram for new user registration.

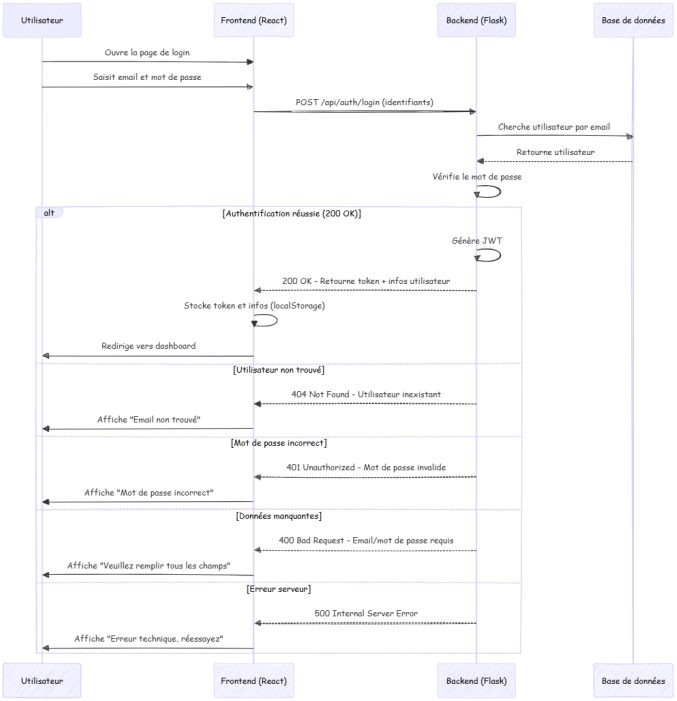


FIGURE 21 – Sequence diagram for existing user login.

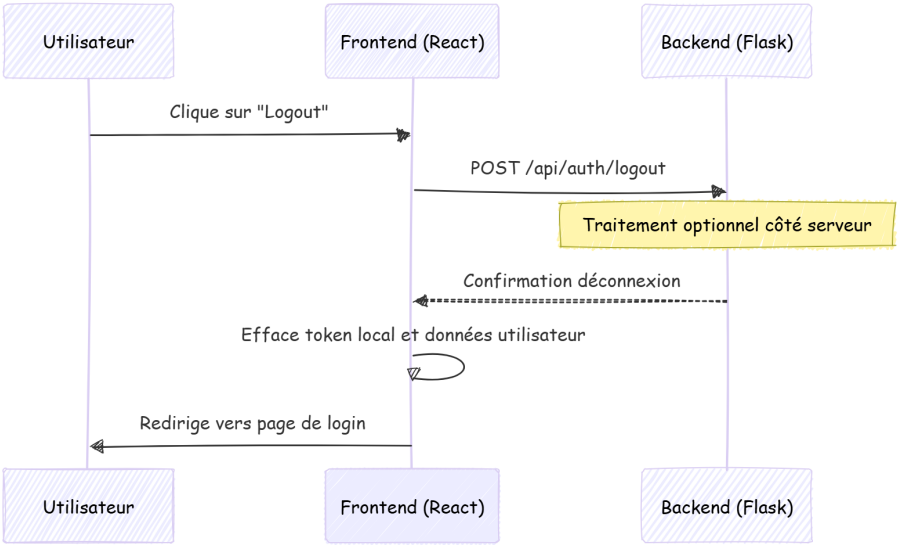


FIGURE 22 – Sequence diagram for user logout.

3.2 Main Business Functionality Flows

The application's core lies in its ability to analyze documents. The diagrams below illustrate in detail the two proposed analysis modes, highlighting their asynchronous nature.

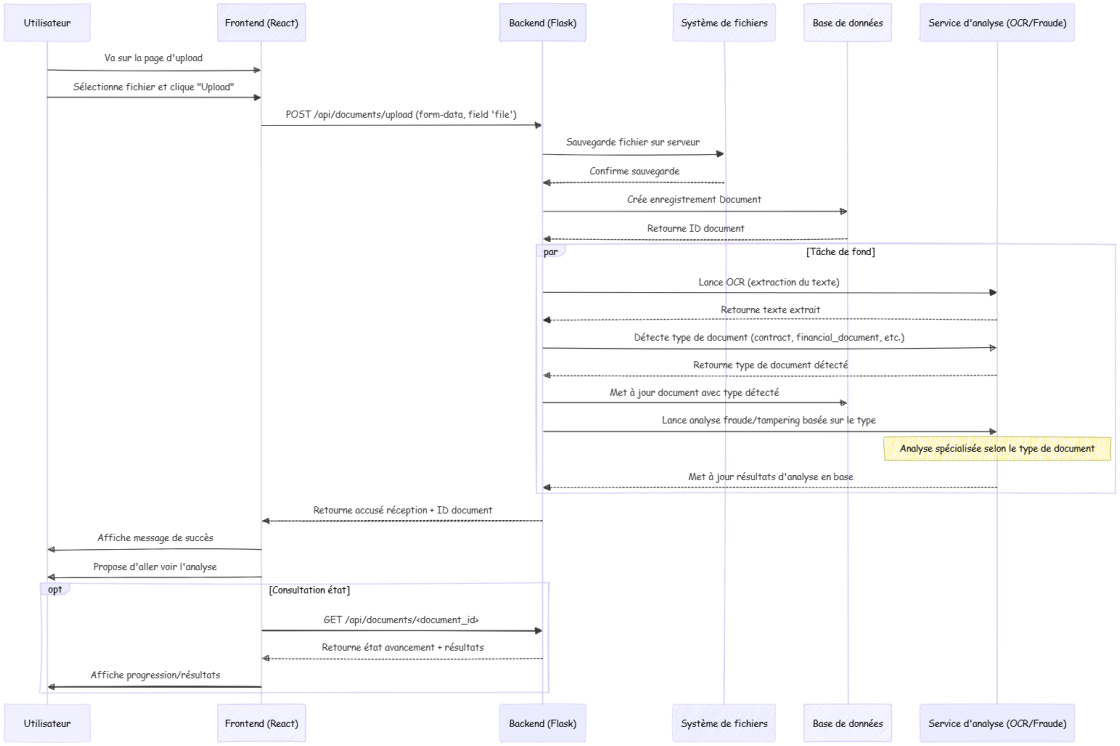


FIGURE 23 – Sequence diagram for the simple mode analysis process.

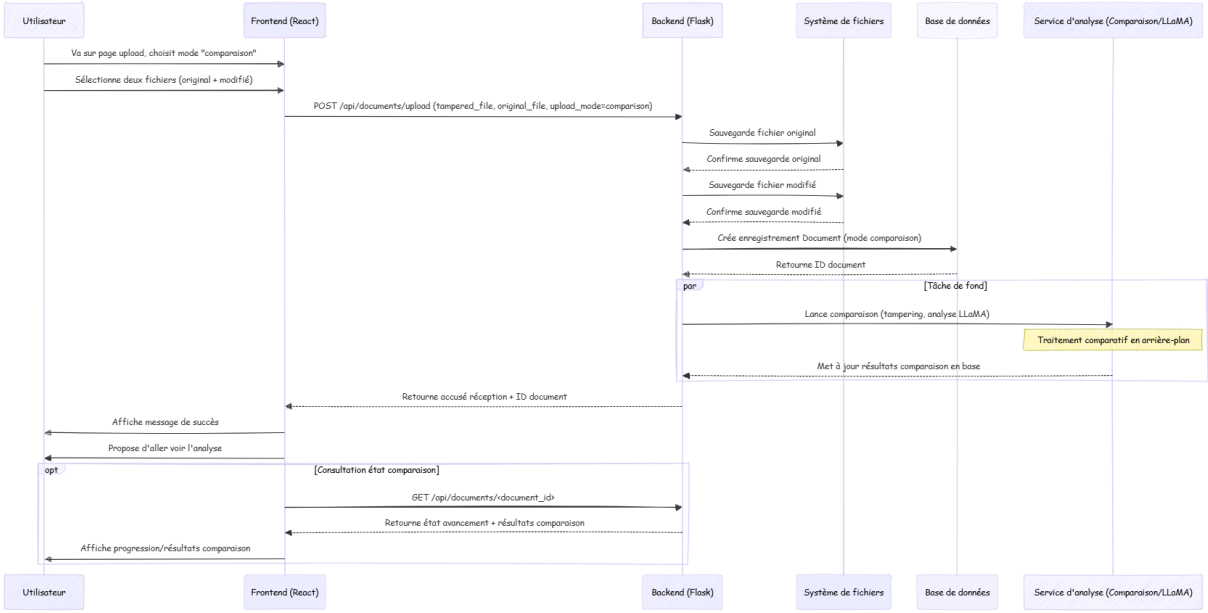


FIGURE 24 – Sequence diagram for the comparison mode analysis process.

4 Data Structure Design

4.1 NoSQL Data Model (MongoDB)

The choice of MongoDB as the database management system was imposed by its flexibility. Being a document-oriented NoSQL database, it is particularly suited for storing heterogeneous and semi-structured (JSON) results produced by different AI models. The data architecture is structured around three main collections :

Collection `users` This collection manages identity and security information for each user.

TABLE 10 – Structure of the `users` collection for account management.

Field	Type	Description
`_id`	ObjectId	Unique user identifier (primary key).
`username`	String	Unique username.
`email`	String	Unique email address for login.
`password`	String	Hashed and salted password for security.
`created_at`	DateTime	Account creation timestamp.

Collection `documents` This is the application's central collection. Each document represents an analysis task and its lifecycle.

TABLE 11 – Structure of the `documents` collection, core of the analysis workflow.

Field	Type	Description
`_id`	ObjectId	Unique document identifier.
`user_id`	ObjectId	Reference to the document owner user.
`doc_type`	String	Document type (ex : "contract", "financial_document").
`filename`	String	File name stored on server (with UUID).
`original_filename`	String	Original filename uploaded by user.
`file_path`	String	File access path on server.
`original_file_path`	String	Path to original file (in comparison mode).
`upload_mode`	String	Chosen analysis mode (ex : "single", "comparison").
`status`	String	Tracks asynchronous processing state ("uploaded", "analyzed", "complete").
`analysis`	Object	Stores semantic AI JSON report.
`fraud_detection`	Object	Stores visual AI JSON report (coordinates, scores).
`extracted_text`	String	Text extracted by OCR for main document.
`original_text`	String	Text extracted by OCR for original doc. (in comparison mode).
`error_message`	String	Stores error message in case of failure (optional).
`created_at`	DateTime	Upload timestamp.
`updated_at`	DateTime	Last update timestamp (ex : analysis completion).

Design Principles and Summary The design of this schema is based on several key principles :

— **Flexibility :** MongoDB's document-oriented format allows storing AI reports and extracted texts without rigid schema.

— **Relationality by Reference :** Links between entities are properly maintained via `ObjectId` references.

— **Asynchronous State Management :** The `status` field in the `documents` collection is essential for decoupling.

— **Security and Traceability :** Password hashing and operation timestamping ensure security and audit trail.

Chapter Conclusion

This chapter has established the complete technical plan for our application. Starting from modern architectural principles like microservices and asynchronism, we have defined a flexible technology stack (React, Flask, MongoDB) and modeled in detail the interactions and data structures via UML diagrams and collection schemas. We now have a clear and robust master plan.

Armed with this detailed design, the next chapter will focus on the concrete implementation phase. We will describe the implementation of different services, the training process of our AI models, and present the experimental results obtained.

5
CHAPTER

TECHNICAL IMPLEMENTATION : DATA, MODELS & INTEGRATION

Chapter Introduction

After designing our system architecture, this chapter focuses on its concrete implementation. We detail the process of building our dataset, a critical step made complex by the inability to access real documents. We will then present in depth the implementation of our two analysis pillars : the visual detection model, an improved version of SegFormer, and the semantic analysis model, which relies on advanced prompt engineering to drive LLaMA-3.

1 Data construction & processing

The performance and reliability of an artificial intelligence model are intrinsically linked to the quality and relevance of the data on which it is trained. Document fraud detection in banking environments faces a major challenge : the absolute confidentiality of client data. Due to strict regulatory and ethical constraints (such as GDPR), it was impossible to use real documents for training.

Faced with this constraint, our strategy was to build a high-quality composite dataset, by aggregating and creating data from public and academic sources. The objective was to create a dataset sufficiently large and diversified to simulate a wide range of fraud scenarios, thus ensuring that our model can generalize its learning to unknown real cases.

1.1 Constitution of a Heterogeneous Corpus

To establish a robust base, we merged three complementary corpora, each playing a strategic role :

— **The Modified "RVL-CDIP" Corpus (The Controlled Fraud Scenario) :** To simulate the most critical type of fraud, we selected 54 documents from the public `RVL-CDIP` corpus. On each one, we manually simulated text substitution by copy-paste with `GIMP`. For each altered document, a binary ground truth mask was meticulously created with the `LabelMe` annotation tool, providing us with a base of 54 perfectly controlled « original/falsified » pairs.

— **The "SUPATLANTIQUE" Dataset (Diversity of Attack Types) :** To diversify scenarios, we integrated this academic dataset that focuses on three families of tampering : *copy-move*, *retouching* and *splicing*. This corpus enriched our base with 102 falsified images and their masks.

— **The "Roboflow" Corpus (Volume Augmentation) :** To ensure good generalization, we added 496 new image/mask couples from the public Roboflow workspace « Document Forgery Detection ».

This fusion resulted in a base repository of 652 unique falsified images.

1.2 Data Preprocessing and Augmentation

Once the sources were aggregated, a technical preparation phase was conducted to guarantee dataset consistency and increase its richness.

Harmonization and Normalization :

— **Format Standardization :** All images were converted to grayscale and their resolution was standardized to 300 dpi.

— **Mask Normalization :** Binary masks were standardized : pixel value 255 was systematically assigned to falsified zones (*tampered*), and 0 to authentic zones.

— **Image Normalization :** To fully leverage our model's pre-training, image pixel values were normalized using standard ImageNet dataset mean and standard deviation.

"Offline" Augmentation for Dataset Doubling : To increase volume, an augmentation strategy was implemented to double the size of our dataset. For each of the 652 original images, a single augmented version was generated by randomly choosing a transformation from a panel of operations (light rotations, horizontal flips, brightness variations). This method allowed us to go from 652 to 1,334 image-mask pairs, creating an intrinsically richer extended dataset.

"On-the-fly" (Online) Augmentation : In addition to "offline" augmentation, random transformations (geometric and photometric) are applied in real time during training. Each image presented to the model is therefore slightly different at each epoch, which forces the model to learn intrinsic fraud characteristics and considerably improves its robustness.

1.3 Dataset Division and Balancing

The last preparation step consists of organizing this final corpus of 1,334 images for efficient training and fair evaluation. This preparation takes place in three stages : stratified image division, patch cutting for technical reasons, and finally filtering these patches for class balancing.

Stratified Division (Split) : First, the dataset was divided into two sets, ensuring that the proportion of different fraud types was preserved in each :

— **Training set :** Composed of 1,231 images.

— **Validation set :** Composed of 103 images.

— A third **test set** was set aside for final and impartial performance evaluation.

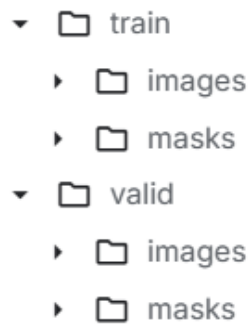


FIGURE 25 – Illustration of stratified dataset division into training and validation sets.

2 Visual & semantic modeling

2.1 Visual modeling

A Comprehensive Technical Approach

The core of our visual detection system relies on a fully customized modeling pipeline, designed to address the specific challenges of document fraud detection. Faced with the fineness of manipulations, high document resolution and extreme class imbalance, we developed a holistic approach built around five fundamental pillars : strategic data preparation, specialized model architecture, multi-component loss function, adaptive two-phase training strategy, and rigorous validation methodology.

2.1.1 Preprocessing and Data Strategy

Our preprocessing pipeline transforms high-resolution documents into an optimal format for deep learning, while preserving critical details necessary for micro-alteration detection.

Adaptive Patch Cutting

Justification for Patch Cutting : The strategy of cutting images into patches was adopted to respond to a major technical constraint while maximizing model performance. The justification breaks down as follows :

— **Problem — Memory Limitation (OOM)** : Very high-resolution document images (e.g. :

2000×3000 pixels) are too large for GPU memory. Their direct processing systematically leads to *Out of Memory (OOM)* type errors, which blocks all training.

— **Discarded Alternative — Image Resizing** : Reducing the global image resolution was an option considered but quickly discarded. This method causes irreversible loss of critical information, as fraud zones, often very small, become blurred or disappear completely.

— **Adopted Solution — 512×512 Patch Grid** : The retained solution is to cut each image into a grid of tiles (patches) of 512×512 pixels with a 10% overlap (51 pixels). This strategic overlap guarantees that no alteration located at patch boundaries is truncated, thus preserving the contextual integrity necessary for detection. This technique offers four fundamental advantages :

  — **Complete Detail Preservation** : It solves the memory constraint while guaranteeing that 100 % of original image pixels are preserved without alteration in patches.

  — **Data Augmentation** : It multiplies the number of samples available for training (each image generating dozens of patches), which favors better model generalization capacity.

  — **Class Imbalance Mitigation** : At pixel scale, the dataset presents extreme imbalance. Out of more than 5.1 billion total pixels, only **0.74 %** are fraud pixels (value 255), versus **99.26 %** for background (value 0). Patch cutting is a crucial step that subsequently allows massively discarding entirely "healthy" patches (containing only background pixels), thus preparing the ground for efficient dataset rebalancing.

Intelligent Filtering by Thresholding As established previously, class imbalance at pixel level is extreme. The cutting strategy now offers us a powerful lever to address this problem. Without targeted intervention, the model would be exposed to an overwhelming majority of "healthy" examples, which would encourage it to adopt a lazy strategy : systematically predict « no fraud » to achieve high precision without acquiring real detection skill.

To force the model to focus on rare but important cases, an aggressive filtering strategy is implemented. It consists of :

— **Systematically keeping** all patches containing a minimum of **75 fraudulent pixels**.

— **Drastically sub-sampling** entirely healthy patches, keeping only a ratio of **5%** of them.

The results of this selection are as follows :

— Out of a total of 12,652 patches generated for training, the 5,159 patches containing an alteration were kept, but only 257 (i.e. 3.4 %) of the 7,493 healthy patches were retained. **The**

proportion of patches of interest thus went from 41 % to more than 95 %.

— For validation, the same logic was applied to the 864 patches generated, keeping the 417 patches with fraud against only 20 healthy patches.

2.1.2 Enhanced SegFormer Architecture

Our architecture is based on SegFormer, a modern Transformer model. However, its performance for microscopic fraud detection is intrinsically limited. The problem comes from its standard decoder which, to be effective on normal-sized objects, merges feature maps at various scales, thus diluting fine details (a modified character, an erased signature) that are crucial for our task. These weak signals then risk being ignored during final prediction.

Foundation : Mix Transformer Encoder (MiT) To maintain a robust base, we use the `mit-b1` encoder pre-trained on ImageNet. It provides us with a powerful hierarchical representation of visual features, enabling efficient knowledge transfer from natural scenes to the document domain.

Innovation : Augmented Decoding Head To overcome the decoder limitation, our main innovation does not replace it but augments it via an intelligent overlay : our `EnhancedSegFormerDecodeHead` module. It intercepts feature maps from the encoder, applies specialized mechanisms to act as a software "magnifying glass", then transmits these enriched maps to the original SegFormer decoder for final segmentation.

This enhanced decoding head integrates three key mechanisms :

— **Micro-Object Attention** : A dedicated 3×3 convolutional layer applied on the high-resolution feature map (`x[0]`) to amplify subtle alteration signals.

— **Spatial Attention** : A spatial attention mechanism (7×7 convolution + sigmoid) that generates adaptive weights for each image region.

— **Residual Connections** : Preservation of original features via residual connections, avoiding signal degradation during enhancement.

Regularization and Stabilization To ensure stable training, we implemented :

— Strong negative bias initialization (-6.0) in the decoding head, encouraging initial caution

— Dropout regularization (rate 0.1) to prevent overfitting

— Gradient clipping (maximum norm 1.0) for numerical stability

2.1.3 Multi-Component Loss Function

Our `SizeAwareWeightedFocalTverskyLoss` function translates complex business objectives into a coherent optimization signal, combining several complementary strategies.

Weighted Tversky Component The Tversky index offers fine control of the precision-recall trade-off via parameters α and β :

$$\text{Tversky} = \frac{TP + \epsilon}{TP + \alpha \cdot FP + \beta \cdot FN + \epsilon} \tag{9}$$

where TP, FP, FN represent respectively true positives, false positives and false negatives, and ϵ a smoothing term.

Adaptive Focalization The focal component, controlled by parameter γ, amplifies loss on difficult examples :

$$\text{Loss}_{\text{focal}} = (1 - \text{Tversky})^{\gamma} \tag{10}$$

Size-Aware Weighting The most innovative mechanism dynamically analyzes each altered region. Objects with surface area less than 80 pixels receive a multiplier weight of 7.0, guaranteeing that even micro-alterations contribute significantly to the loss.

Contour Enhancement An additional weight (factor 2.5) is applied to pixels located on altered zone contours, encouraging precise delimitations via morphological operations (dilation-erosion).

The complete architecture, integrating all these components, is illustrated by the figure below.

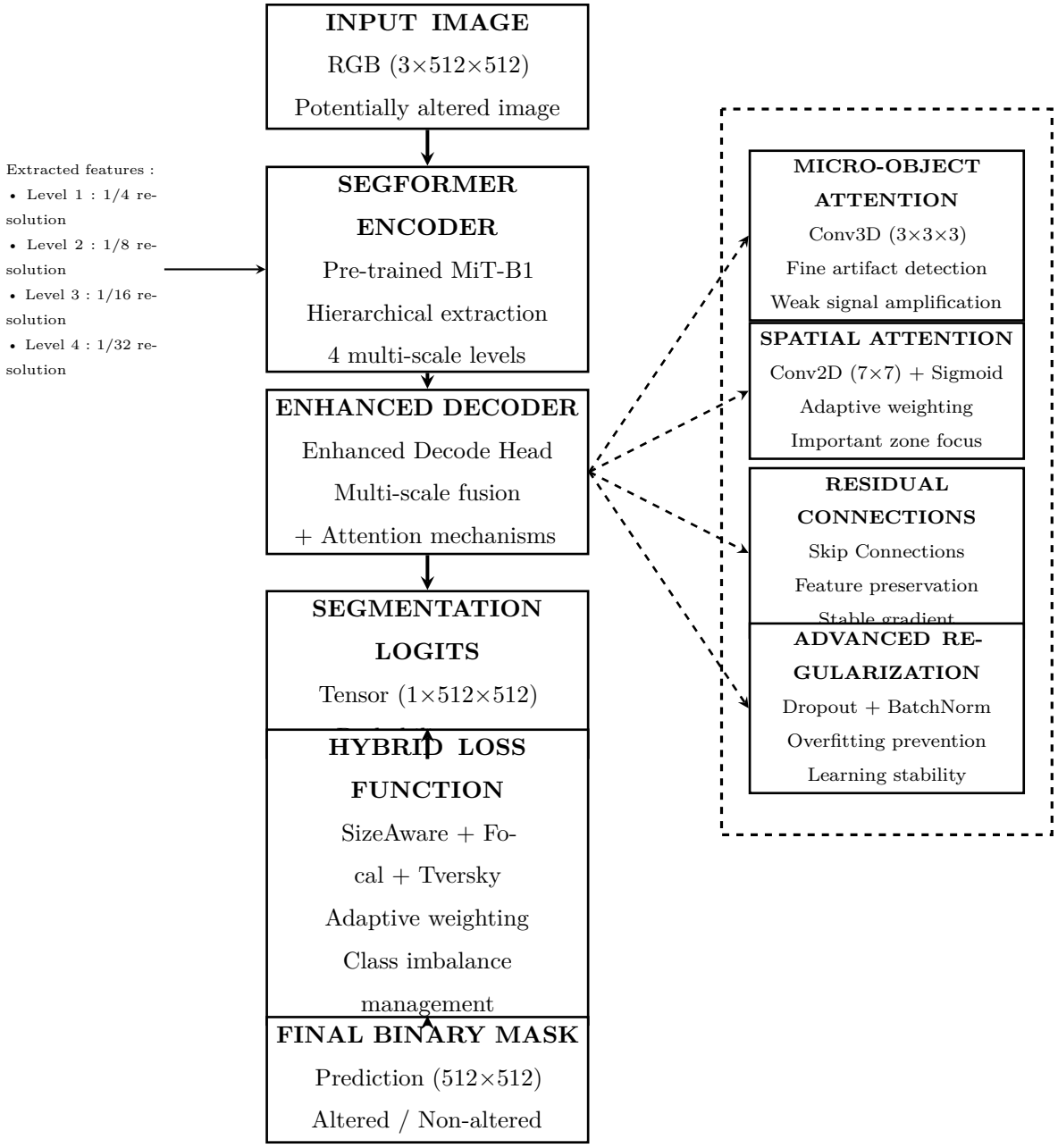


FIGURE 26 – Enhanced SegFormer Architecture - Fraud detection pipeline with innovation modules

2.1.4 Adaptive Training Strategy

Our training strategy follows a bi-phasic curriculum, with evolving optimization objectives.

Phase 1 : Hyper-Sensitive Discovery (0-30% of epochs) The initial objective is to develop maximum sensitivity to alterations :

— Loss parameters favoring recall : $\alpha = 0.35$, $\beta = 0.90$

— Linearly increasing tampered weight : $12.0 \rightarrow 18.0$

— Optimization centered on recall metric

Phase 2 : Balanced Refinement (30-100% of epochs) The second phase aims for precision-recall balance via F1-Score optimization :

— Parametric evolution : $\alpha : 0.35 \rightarrow 0.50$, $\beta : 0.90 \rightarrow 0.70$

— Tampered weight decrease : $18.0 \rightarrow 12.0$

— F1-Score optimization with dynamic threshold search

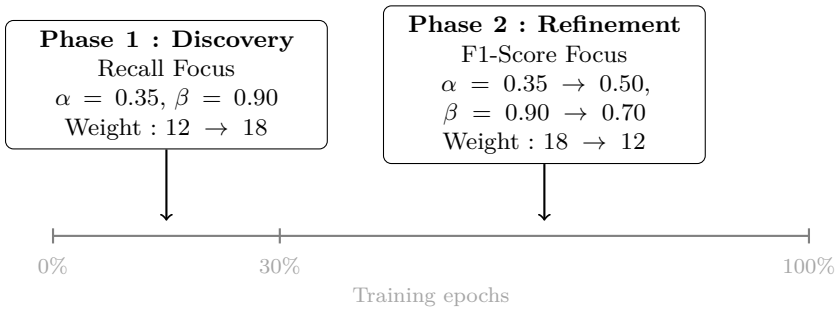


FIGURE 27 – Adaptive Two-Phase Training Strategy

Technical Optimizations Computational efficiency is ensured by :

— **Mixed precision (FP16)** : 50% reduction in memory usage

— **Gradient accumulation** (8 steps) : Simulation of effective batches of size 16

— **AdamW optimizer** with `OneCycleLR` scheduler

— **Proactive memory management** : Automatic GPU cache cleaning

2.1.5 Robust Validation and Inference

Comprehensive Metrics Suite Our evaluation goes beyond traditional global metrics by separately calculating performance for each class :

— **Global metrics** : Dice, IoU, Precision, Recall, F1-Score

Innovation	Mechanism	Impact
Micro-Object Attention	Conv 3×3 on HR features	+15% small zone detection
Spatial Attention	Conv 7×7 + Sigmoid	Adaptive weighting
Size-Aware Loss	Connected component analysis	Weight ×7 objects <80px
Edge Enhancement	Morphological operations	Contours ×2.5 more precise
2-Phase Training	Adaptive curriculum	Recall then F1 balance
Dynamic Threshold	Automatic search	Continuous optimization

TABLE 12 – Key Architecture Innovations

— **Per-class metrics** : Distinct analysis of performance on healthy and altered regions

— **Confusion metrics** : Detailed confusion matrix with error analysis

Dynamic Threshold Optimization At each epoch, an exhaustive search tests 30 candidate thresholds (0.1 to 0.99) on the validation set, selecting the one maximizing F1-Score. This adaptive approach ensures fair and reproducible evaluation.

Specialized Model Portfolio Our system maintains four specialized champion models :

— **Recall Champion** : Zero tolerance for missed frauds

— **Precision Champion** : Minimization of false alerts

— **F1 Champion** : Optimal balance

— **Dice Champion** : Global segmentation performance

This strategy offers crucial business flexibility, enabling deployment of the most suitable model according to the desired risk context.

2.2 Semantic modeling

Semantic modeling constitutes the second pillar of our system. It no longer focuses on form (pixels) but on **content** — text and its meaning. Its two main objectives are to extract textual content from any document with maximum reliability, then use artificial intelligence to forensically analyze differences between two versions of this document.

2.2.1 Text Extraction : A Hierarchical OCR Pipeline

The quality of any semantic analysis depends entirely on the fidelity of extracted text. This is why we implemented a robust, multi-level OCR (Optical Character Recognition) process.

— **Objective :** Transform any document (image, PDF, Word) into exploitable text, overcoming challenges posed by visual noise, varied fonts and complex layouts.

— **OCR Tool Choice :** After thorough evaluation, we selected `LLM Whisperer` as the main engine due to its superior ability to interpret visual context, which is crucial for financial documents. Table 13 details this comparison.

Criterion	`LLM Whisperer` **(Main Choice)**	`Tesseract` **(Fallback)**	`EasyOCR`	`PaddleOCR`
Technology	LLM (Transformer)	LSTM / HMM	CNN + LSTM	CNN + LSTM
Precision (Complex Docs)	**Very high** (tables, columns, noise)	Low–medium	Medium–high	High, especially for Asian languages
Contextual Understanding	**High**	Very low	Low	Medium
Deployment	Cloud API	Local library	Local library	Local library
Ideal Use Case	High-fidelity financial/official docs	Basic OCR	Balanced performance/simplicity project	Very fast multilingual applications

TABLE 13 – OCR tools comparison

— **Decision Architecture :** To guarantee resilience, our system follows a hierarchical flow :

1. An orchestrator detects document type to choose the most direct method.
2. First extraction attempt on images is always made via `LLM Whisperer`, with a retry mechanism (up to 3 attempts) to handle network randomness.
3. In case of definitive failure of the main service, the system automatically switches to `Tesseract` as local backup solution, ensuring high availability.

2.2.2 AI Semantic Analysis

Once text is faithfully extracted, the core analysis intervenes. This process is driven by a large language model, chosen and instructed to act not as an assistant, but as a true expert.

— **Objective :** Detect, categorize and evaluate the risk of *any* modification — whether flagrant like an altered amount, or subtle like a punctuation change.

— **Selected Model and its Architecture :** `LLaMA-3 70B`

Our choice fell on Meta's `LLaMA-3 70B` model. This decision is based on architectural and performance advantages specific to our mission :

— **"Decoder-Only" Transformer Architecture :** LLaMA-3 is, like GPT series models, a "decoder-only" type architecture. This means it is fundamentally designed for one task : predicting the next word in a sequence. This architecture makes it exceptionally good at understanding long context (our prompt and the two documents to compare) and generating textual continuation that is not only coherent but also logically structured, which is perfect for writing an analysis report.

— **Optimized Attention (Grouped Query Attention - GQA) :** The main bottleneck of large Transformer models is the attention mechanism, which is very costly in computation and memory. LLaMA-3's 70B model integrates a key optimization called *Grouped Query Attention*. This is a technique that simplifies the attention process without significantly degrading performance. Concretely, this allows the model to process information much faster, making its use via Groq API quasi-instantaneous and viable for an interactive application.

— **Training Data Quality :** An LLM's power doesn't just come from its size. LLaMA-3 was pre-trained on a colossal corpus of more than 15 trillion tokens, which underwent aggressive filtering to retain only the highest quality data. This vast and clean knowledge base gives it superior reasoning capabilities and greater factual fidelity, two qualities essential for a forensic task.

— **Specialization through "Instruction Fine-Tuning" :** Beyond its initial training, the model underwent a specific adjustment phase where it learned to follow complex instructions. It's this capability we exploit to transform it into an on-demand analyst.

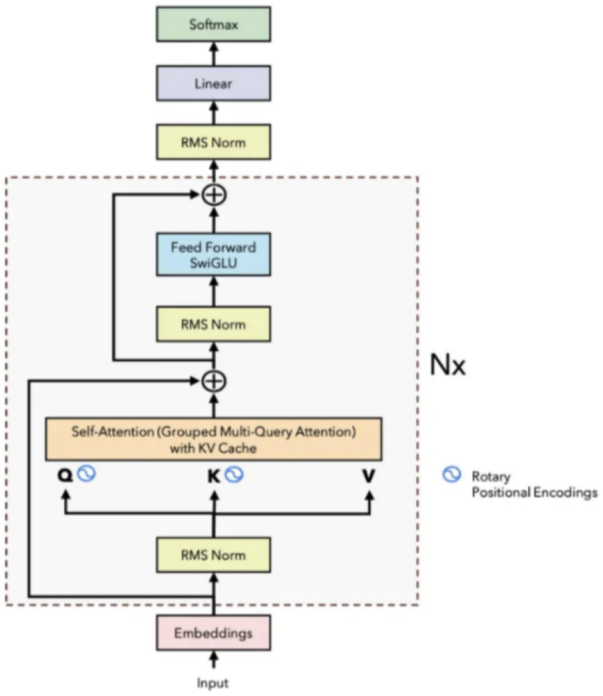


FIGURE 28 – Llama-3 model architecture.

— **Prompt Engineering : The Art of Guiding AI**

To obtain such precise results, we cannot simply ask a simple question. We implement an advanced "prompt engineering" strategy :

— **The "Role-Prompt" : Assigning Expertise :** The very first instruction, "You are an elite forensic analyst... ", is a powerful technique. It constrains the model to draw from its most relevant knowledge related to analysis, fraud and financial terminology. It immediately adopts a factual tone and meticulous approach.

— **The "Task-Prompt" : Imposing a Strict Framework :** We don't ask for an "analysis", we demand a structured report with precise sections : `Summary`, `Detailed changes`, and `Risk`. This constraint transforms the model from a creative text generator into a predictable data processing tool. This guarantees that each analysis is complete and, most importantly, that its response can be read and automatically interpreted by our application to extract key information.

— **Parameter Control : Mastering Behavior :**

— `temperature = 0.05` : This setting, close to zero, virtually eliminates all creativity

and randomness. The model will always choose the most probable and logical word sequence, guaranteeing a factual and deterministic report, essential for reliable analysis.

— `max_tokens = 8000` : By allocating large response capacity, we give the model freedom to produce comprehensive reports, even for very long documents with numerous modifications, without ever risking being cut off mid-analysis.

— **Complete Analysis Flow**

The end-to-end semantic process is therefore as follows :

1. OCR-extracted texts are first **cleaned and normalized** to eliminate artifacts.
2. They are then inserted into custom-designed prompts and sent to `LLaMA-3` via Groq API.
3. The response, already structured by the prompt, is received.
4. **Post-processing** modules extract critical fields (like risk score) using regular expressions.
5. The system produces a **final JSON object** containing the complete report, summary, risk level and source texts, ready to be presented to the user or used by other application modules.

Chapter Conclusion

This chapter has detailed the core of our technical contribution. Starting from a strong constraint of absence of real data, we built a robust and balanced composite dataset. We then implemented our two analysis engines : a SegFormer vision model significantly improved by a custom detection head and loss function, as well as a semantic analysis pipeline exploiting LLaMA-3's power via rigorous prompt engineering. All these building blocks constitute a functional and innovative solution.

Now that the construction phase is complete, the next chapter will be entirely devoted to evaluating its performance. We will present quantitative results from our models, illustrate their capabilities on concrete examples and critically analyze their strengths and limitations.

6
CHAPTER

EVALUATION, ASSESSMENT AND PERSPECTIVES

Chapter Introduction

The previous chapters have led us from problem definition to the design of a complete technical solution. This final chapter aims to close the loop : we will present and analyze the experimental results of our models, illustrate the functioning of the application prototype, then draw a critical assessment of the project, evaluating its strengths, limitations and business value. Finally, we will trace a roadmap for future developments.

1 Experimental Evaluation & Results Analysis

This section presents the empirical validation of our solution. We begin by detailing the test protocol, then analyze the quantitative and qualitative performance of our main model. In accordance with the business objective of minimizing risk, the model presented here is the one that was optimized to maximize the **Recall** metric during training. This strategic choice aims to address the critical challenge of not missing any potential fraud.

1.1 Validation protocol and test sets

— **Test Set :** All performances presented below are calculated on an independent test set, composed of 10 unique images, which was not used at any stage of training or model selection. This separation guarantees an objective and impartial evaluation of our model's generalization capacity.

— **Evaluated Model :** The detailed results come from the checkpoint of our model that achieved the best **Recall** score on the validation set during training.

— **Per-Class Metrics :** For fine analysis, metrics (F1-Score, Recall, Precision, Dice Score) are presented separately for the « Clean » class (healthy zones) and the « Tampered » class (altered zones), to evaluate both model reliability and detection efficiency.

1.2 Quantitative results analysis

Table 14 synthesizes the average performance of our main model on the entire test set.

Evaluated Class	F1-Score	Recall	Precision	Dice Score
'Clean' Class (Healthy)	99.88 %	99.77 %	99.98 %	99.88 %
'Tampered' Class (Altered)	62.61 %	**93.95 %**	47.48 %	62.61 %

TABLE 14 – Average performance of the Recall-optimized model on the test set.

Results Interpretation : The quantitative results analysis highlights our model's dual performance. First, it excels in identifying non-altered zones, with Precision and F1-Score scores for the 'Clean' class that approach 100 %. This demonstrates its high reliability and low propensity to generate false alerts on legitimate document parts.

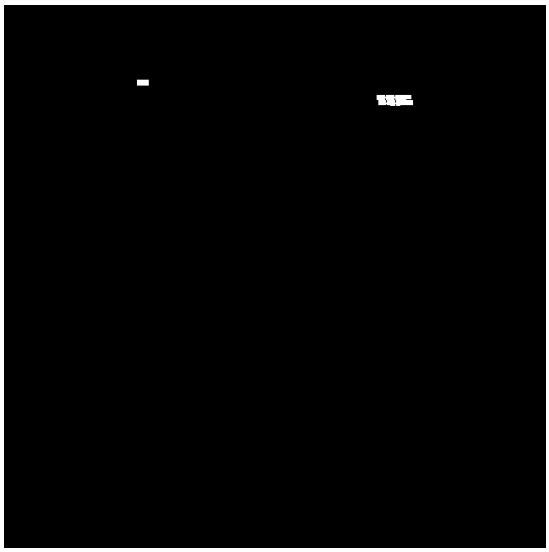
Second, the performance on the 'Tampered' class confirms the effectiveness of our strategy. The **average Recall of 93.95 %** is particularly high, proving that the model achieves its main objective : identifying the vast majority of alterations. This high sensitivity is the result of the training strategy (cf. Figure 16) and the custom loss function (cf. section 4.2.1.3) that prioritize exhaustive detection. This approach is accompanied by an average Precision of 47.48 % for this same class, reflecting the deliberate trade-off in favor of detection, a strategy consistent with maximum security stakes.

1.3 Qualitative prediction analysis

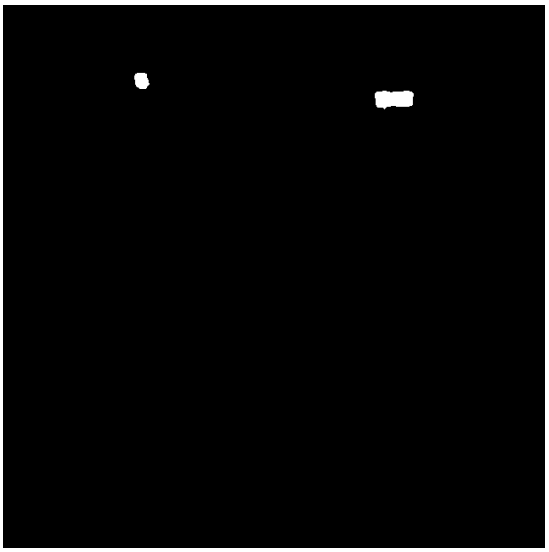
Beyond numbers, visual analysis of specific cases allows understanding the behavior and nuances of our model.

Success Case 1 : High Sensitivity and Reliability Figure 29 illustrates a success case particularly representative of our model's dual performance. Facing subtle text alterations, it achieves a perfect **Recall** of 1.0000 for the 'Tampered' class, successfully identifying the entirety of modified zones. The prediction obtains excellent geometric overlap with ground truth, as evidenced by the **Dice Score** of 0.6968.

Simultaneously, the model demonstrates near-perfect reliability on healthy zones, achieving a **Precision** of 1.0000 and an **F1-Score** of 0.9994 for the 'Clean' class, meaning no false alerts were generated on non-modified parts of this document. This ability to be both very sensitive to the weakest fraud signals while remaining extremely precise on legitimate zones validates the robustness of our architecture (cf. Figure 26).



Ground Truth Mask (Case 1)

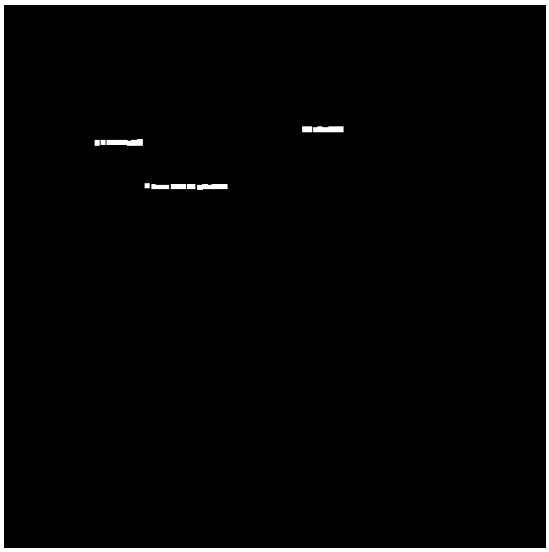


Model-Predicted Mask (Case 1)

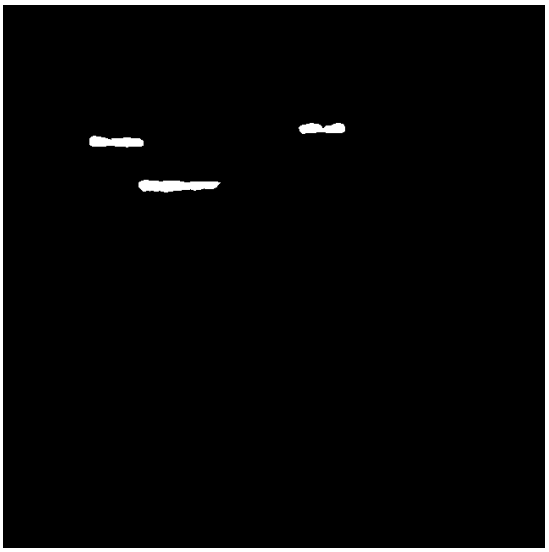
FIGURE 29 – Visual comparison between ground truth mask and model prediction for a success case on subtle textual alteration.

Success Case 2 : Robust and Precise Detection Figure 30 demonstrates the model's ability to generalize its performance on another type of alteration. For this case, the model achieves an excellent **Recall** of 0.9388 for the 'Tampered' class, managing to identify almost all of the fraudulent zone. The **Dice Score** of 0.6616 attests to good geometric correspondence between prediction and ground truth.

In parallel, the model displays remarkable reliability on healthy zones, with a **Precision** of 0.9998 and an **F1-Score** of 0.9986 for the 'Clean' class. These near-perfect scores indicate almost total absence of false alerts on legitimate document parts. This balance between high sensitivity to manipulations and great precision on authentic zones validates the effectiveness of our approach, notably the custom loss function and adaptive training strategy designed for this complex task.



Ground Truth Mask (Case 2)

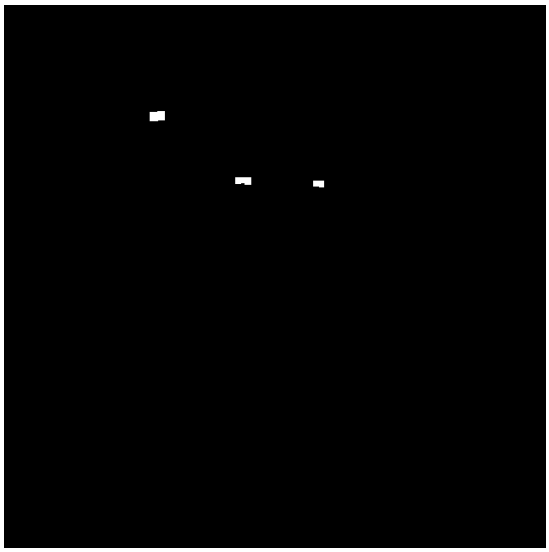


Model-Predicted Mask (Case 2)

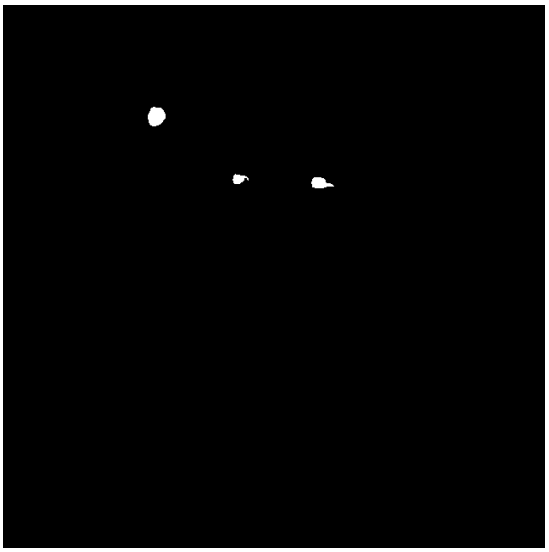
FIGURE 30 – Second success case : precise detection of 'copy-move' type falsification.

Success Case 3 : Robust and Precise Detection Figure 31 demonstrates the model's ability to generalize its performance on another type of alteration. For this case, the model achieves a very good **Recall** of 0.8584 for the 'Tampered' class, managing to identify a large majority of the fraudulent zone. The **Dice Score** of 0.6516 attests to good geometric correspondence between prediction and ground truth.

In parallel, the model displays remarkable reliability on healthy zones, with a **Precision** of 0.9998 and an **F1-Score** of 0.9995 for the 'Clean' class. These near-perfect scores indicate almost total absence of false alerts on legitimate document parts. This balance between high sensitivity to manipulations and great precision on authentic zones validates the effectiveness of our approach, notably the custom loss function and adaptive training strategy designed for this complex task.



Ground Truth Mask (Case 3)

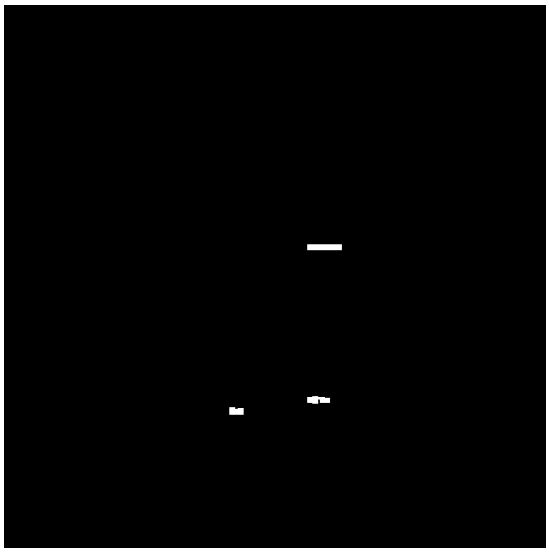


Model-Predicted Mask (Case 3)

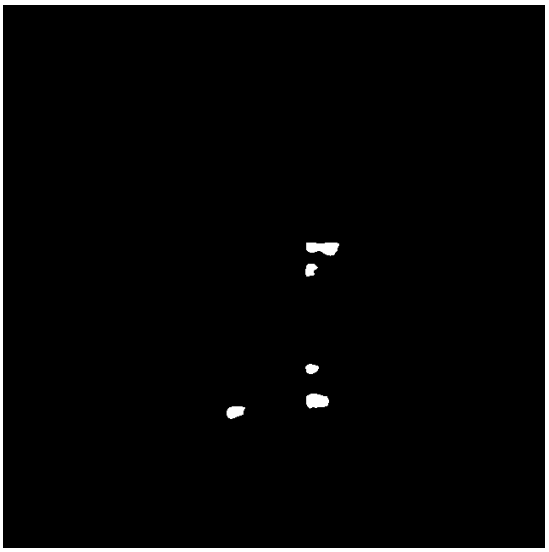
FIGURE 31 – Third success case : precise detection of 'copy-move' type falsification.

Analysis of Recall/Precision Trade-off in Action : Figure 32 perfectly illustrates the strategic trade-off of our Recall-optimized model. On one hand, the model successfully identifies almost all of the altered zone, achieving a **Recall** of 0.9594 for the `Tampered` class. This very high score confirms that the main objective of not missing fraud is achieved.

On the other hand, this high sensitivity is accompanied by a **Precision** of 0.4273 for this same class. This means that while the model did find the fraud, the detection zone it predicted is larger than the actual alteration, including a significant number of false positives. This behavior is the expected counterpart of our strategy : by setting the model to be extremely sensitive, we accept that it generates some prediction « noise ». This approach is consistent with the objective of prioritizing exhaustive detection, considering that human verification to eliminate the few false positives is an acceptable operational cost compared to the risk of undetected fraud.



Ground Truth Mask



Model-Predicted Mask

FIGURE 32 – Illustration of Recall/Precision trade-off : correct detection but background noise.

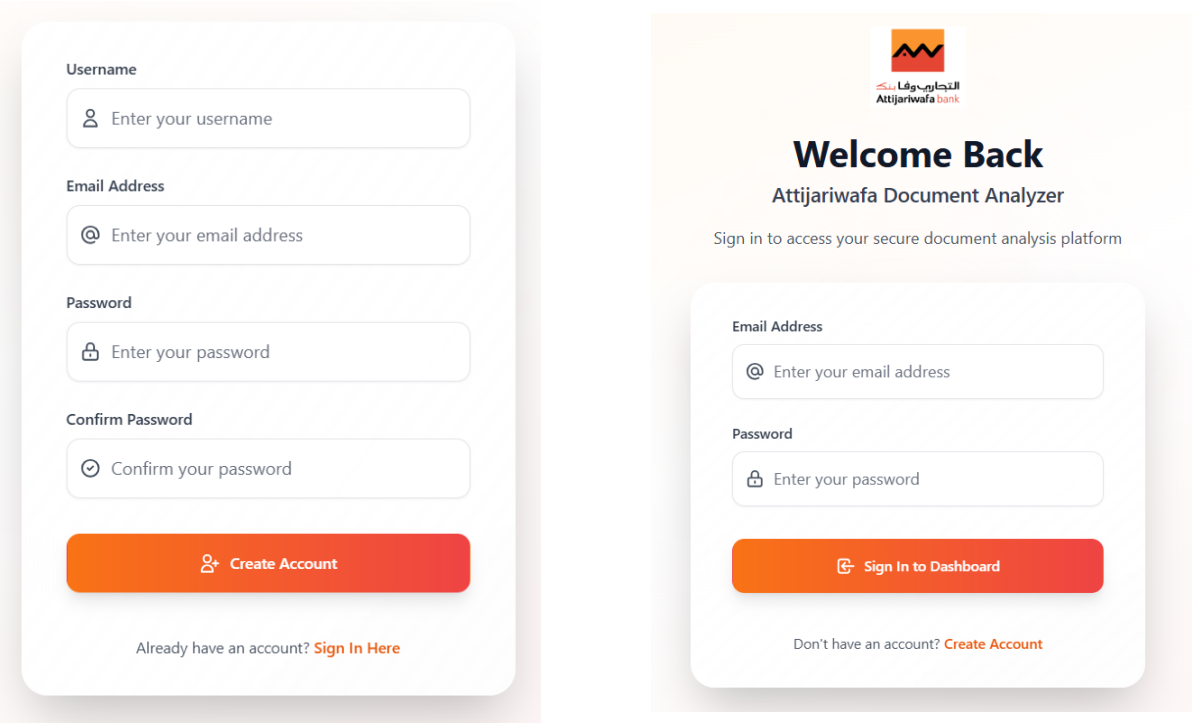
2 Industrialization : From Model to Application Prototype

After validating our detection technology's performance in the previous section, it is essential to demonstrate how this algorithmic power was transformed into a concrete and exploitable work tool. This section presents the application prototype that was developed to orchestrate our AI models and make their results accessible and interpretable for a business analyst. We illustrate here the complete application workflow, from user login to analysis results consultation.

2.1 The Analyst's Journey : An Intuitive Workflow

The interface was designed to offer a simple and intuitive user experience (UX), guiding the analyst through a logical process in several key steps, available via two distinct analysis modes.

Step 1 : Authentication and Platform Access The analyst's journey begins with a secure authentication portal. The user can create an account via a standard registration form or log into their existing account (Figure 33). Each session is protected, ensuring that only authorized users can access the analysis platform.



Registration page

Login page

FIGURE 33 – User authentication flow.

Step 2 : Launching Analysis via Dashboard Once logged in, the analyst accesses a central dashboard (Figure 34). From this screen, they can initiate a new analysis by choosing between « Simple Analysis » mode and « Document Comparison » mode.

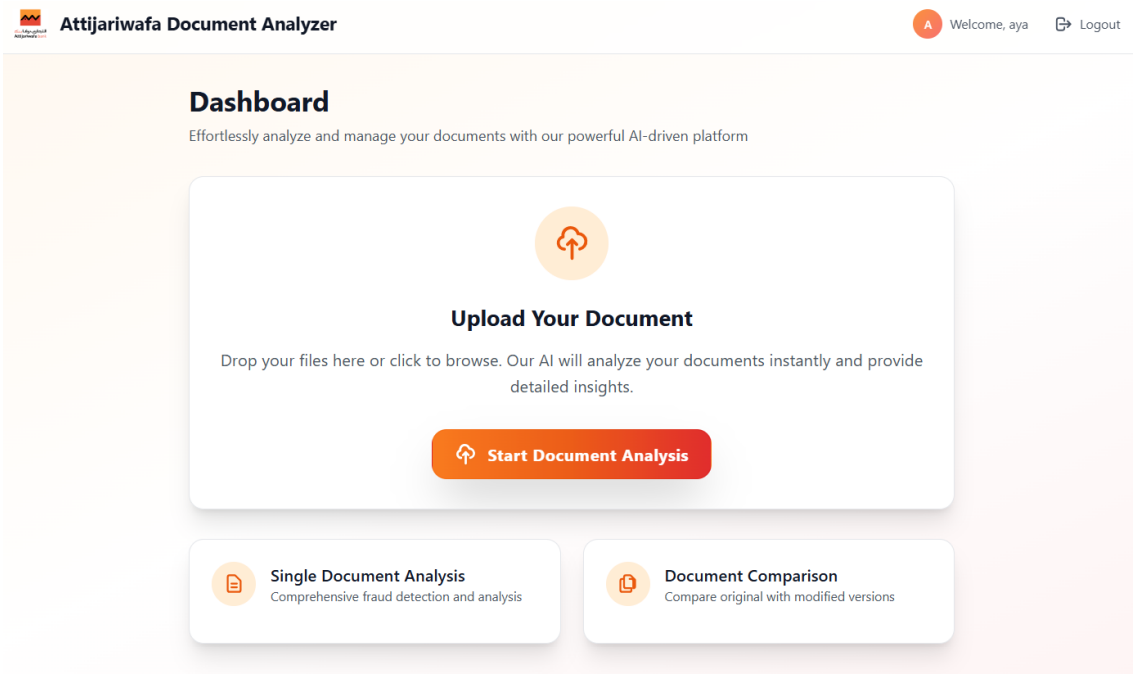
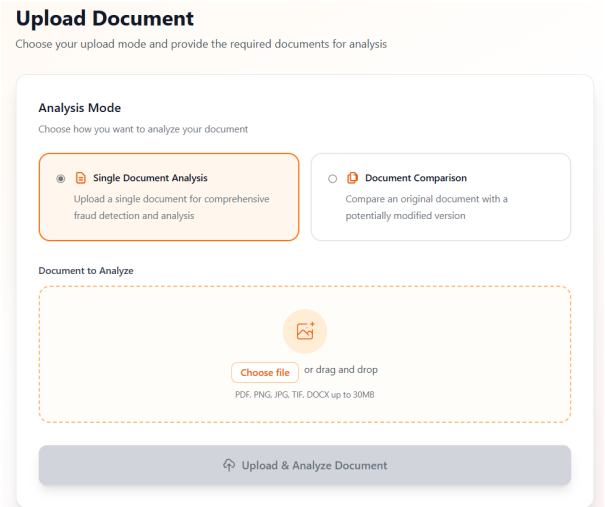


FIGURE 34 – Main dashboard offering choice of analysis mode.

The submission interface then adapts to the selected mode (Figure 35), requesting one or two documents via simple drag-and-drop. Processing is then launched as a background task asynchronously, as illustrated by the loading screen (Figure 36).



Simple Mode submission.

Comparison Mode submission.

FIGURE 35 – Submission interfaces adapted to chosen analysis mode.

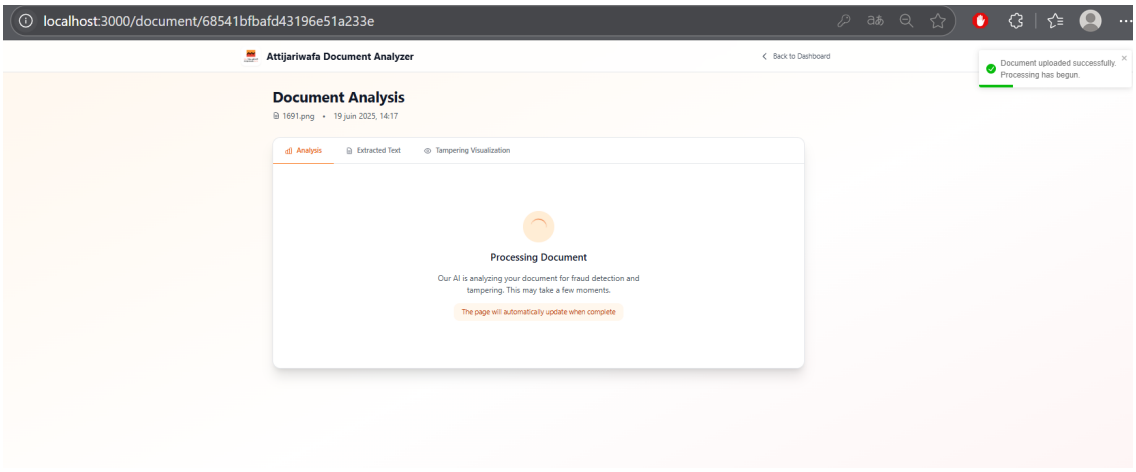
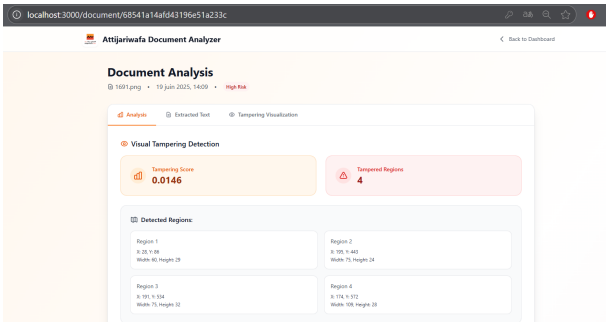


FIGURE 36 – Asynchronous processing tracking screen after submission.

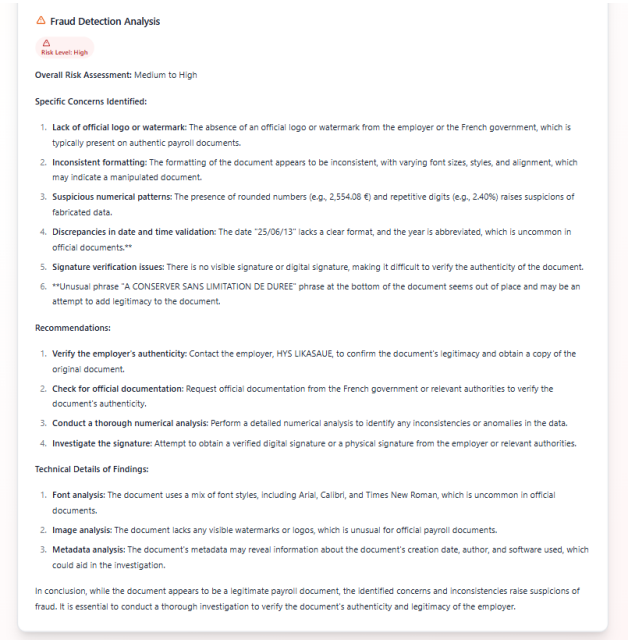
Step 3 : Results Consultation and Interpretation Once analysis is complete, the application presents results in a tabbed interface designed for clear and detailed exploration.

2.1.1 Results in « Simple Analysis » Mode

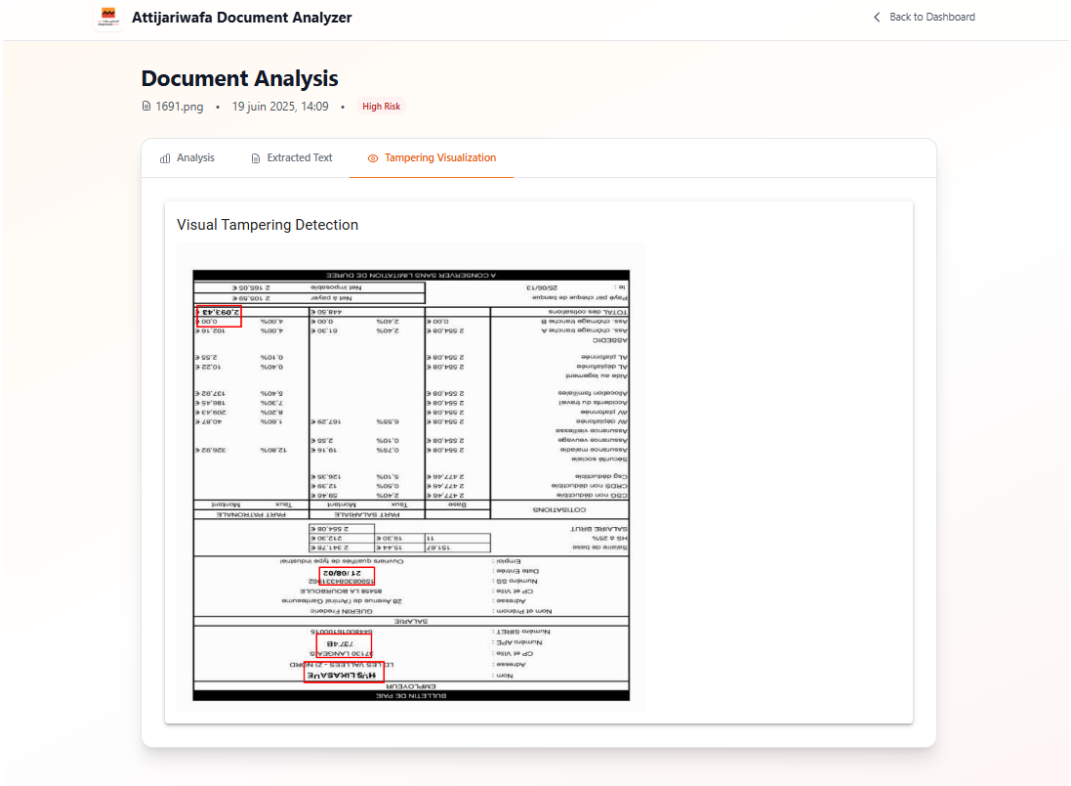
In simple mode, the interface merges outputs from different models into a synthetic view (Figure 37). The **Analysis** tab provides a falsification score and coordinates of detected regions. The **Tampering Visualization** tab materializes these detections by superimposing red frames on the document. Finally, the **Extracted Text** tab ensures complete traceability by displaying raw text from OCR.



Visual analysis synthesis.



Textual analysis result.

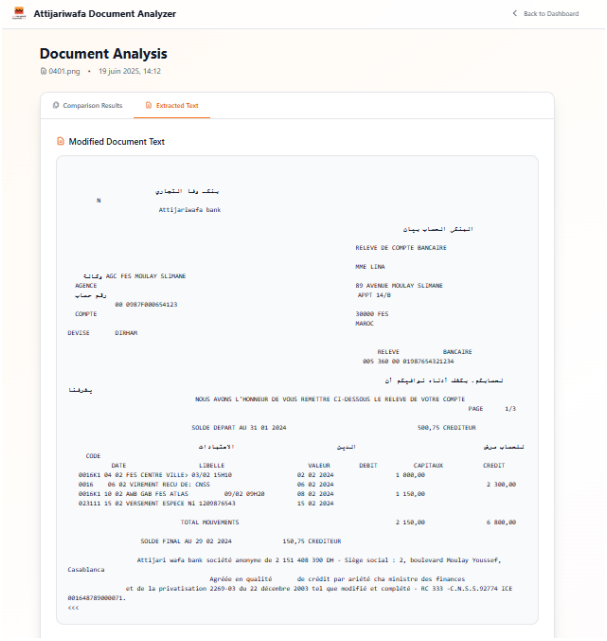


Altered zones visualization.

FIGURE 37 – Results interface for « Simple analysis » mode.

2.1.2 Results in « Comparison » Mode

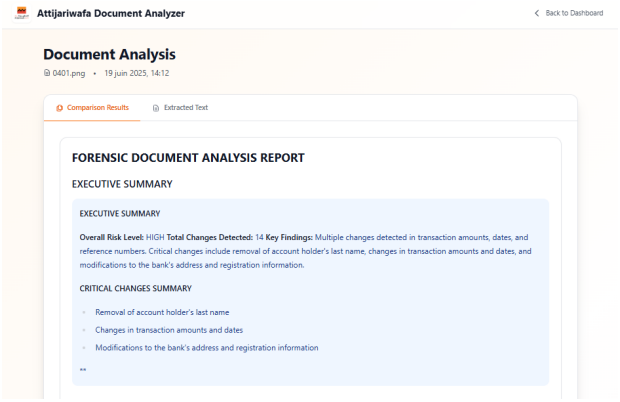
In comparison mode, the application generates a complete forensic report. For total transparency, extracted texts from both documents are first presented (Figure 38). The main report (Figure 39) then details the risk level, exhaustive list of line-by-line changes, and actionable recommendations for the analyst.



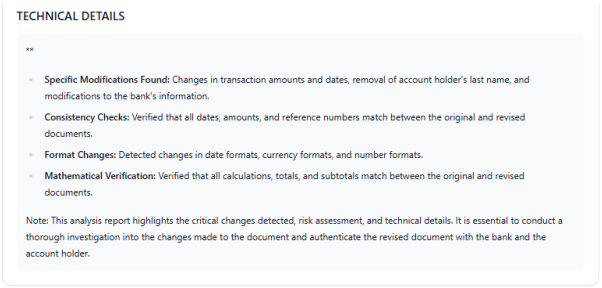
Modified document text.

Original document text.

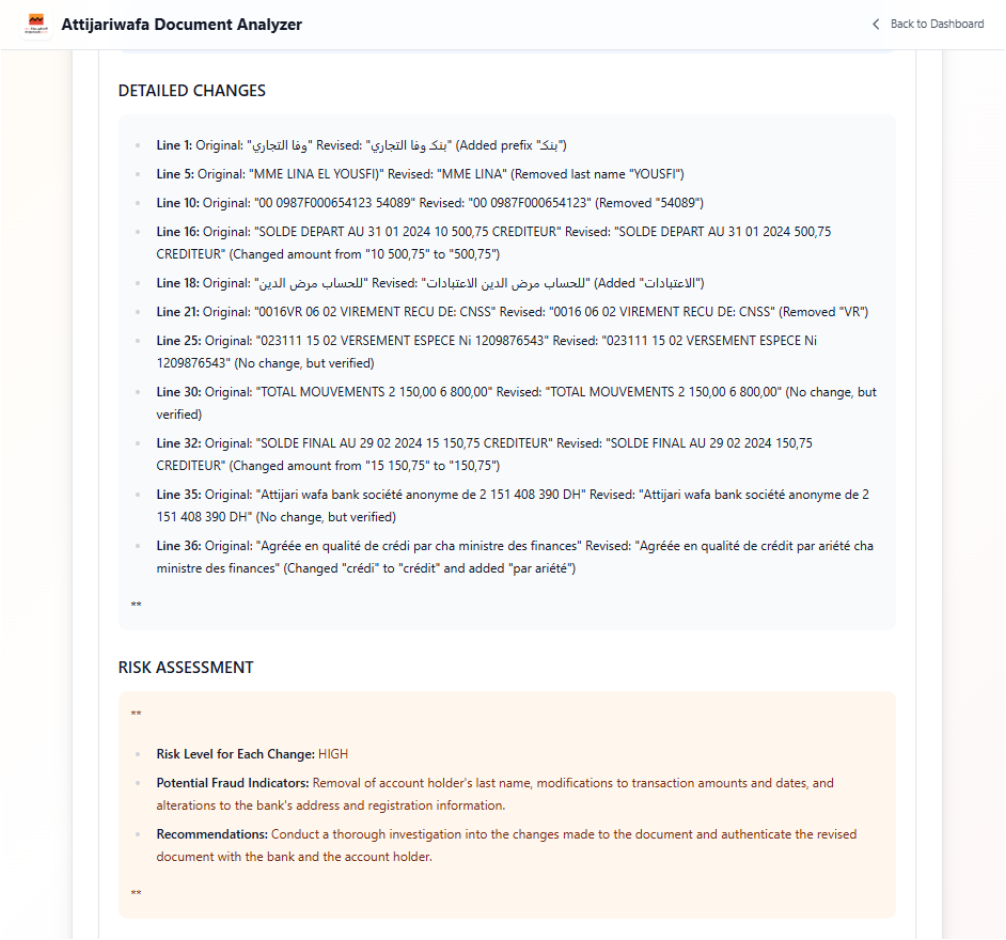
FIGURE 38 – OCR-extracted text consultation in comparison mode.



Part 1 : Executive Summary



Part 3 : Technical Details



Part 2 : Changes List

FIGURE 39 – The three parts of the forensic analysis report generated in comparison mode.

2.2 Application Pipeline Temporal Performance

An industrial solution's efficiency is measured not only by the relevance of its results, but also by its execution speed. To validate our prototype's responsiveness, we measured the complete document processing time, as illustrated in Figure 40.

The total time, from file submission to final report generation, is **24.99 seconds**. This delay breaks down between the different micro-services that orchestrate the analysis :

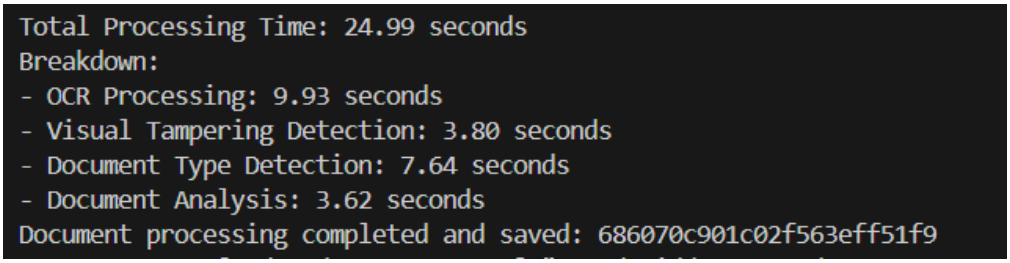


FIGURE 40 – Chronometric detail of document processing steps.

— **Optical Character Recognition (OCR Processing)** : 9.93 seconds

— **Document Type Detection** : 7.64 seconds

— **Visual Tampering Detection** : 3.80 seconds

— **Document Analysis (semantic)** : 3.62 seconds

Analysis of these times confirms several key points of our architecture. First, the visual detection model, which is at the heart of our solution, executes in only **3.80 seconds**, validating its suitability for industrial integration without creating bottlenecks. Second, other services' times remain within quite acceptable limits for asynchronous processing, guaranteeing a fluid user experience where the analyst is not blocked waiting for results.

3 Project Assessment : Critical Analysis, Robustness and Business Value

3.1 Technical Limitations and Model Bias

A lucid analysis of this project requires recognizing its intrinsic limitations, which are mainly related to the nature of training data. The most structuring constraint was the inability to access

real production documents, forcing us to build a corpus from public sources (RVL-CDIP, SUPATLANTIQUE) and simulations. Although this composite dataset is heterogeneous, it cannot perfectly replicate the statistical distribution and specific nature of artifacts encountered in Attijariwafa Bank's operational flows, which constitutes a limitation to the model's guaranteed generalization in production.

The second limitation is an assumed bias in visual detection model design. To respond to the business challenge of not missing any potential fraud, the model was deliberately optimized to maximize Recall, reaching a score of 93.95

Finally, reliance on the LLAMA-3 language model, while very performant, carries a residual risk of "hallucination" or misinterpretation of very specific financial contexts, a challenge inherent to the current state of LLM technology.

3.2 Architecture Validation : Robustness, Scalability and Security

Faced with these limitations, system architecture design constitutes the project's main strength, validating its viability for industrial deployment.

Robustness and Scalability : The strategic choice of microservices architecture and asynchronous processing is the solution's pillar. This decoupling ensures that long processing or failure on an AI model does not impact either application availability or user experience, with the API immediately responding `202 Accepted` to free the interface. This architectural efficiency is complemented by visual analysis model performance. The choice of SegFormer was motivated by its excellent performance/computational cost trade-off, making it suitable for industrial integration. This efficiency allows respecting the constraint of inference time less than 10 seconds on CPU for this model, contributing decisively to overall system responsiveness. Resilience is also integrated, as evidenced by the automatic fallback mechanism from OCR to Tesseract in case of main cloud service failure.

Security : Application security has been integrated according to requirements. User authentication is secured by JWT tokens and passwords are protected by hashing (PBKDF2-SHA256). Strict measures for uploaded file validation (UUID names, type control) and restrictive CORS policy are in place to prevent common vulnerabilities. The entire design respects the regulatory framework of Moroccan law n°09-08, ensuring traceability and logging of all analyses in the MongoDB database.

3.3 Business Impact and Value Analysis (ROI)

The project's added value for Attijariwafa Bank is tangible and structured around three performance axes.

Fraud Risk Reduction : The most direct impact is financial risk mitigation. The model acts as a safety net much more sensitive than human control, whose effectiveness is limited by cognitive fatigue. Banks equipped with such automated systems significantly reduce the cost induced per euro of fraud.

Operational Efficiency Optimization (ROI) : Return on investment comes from automating a costly and time-consuming manual process. By processing a document in less than two minutes, the system can drastically reduce expert workload. This allows reallocating this qualified time to higher value-added tasks, such as analyzing complex cases flagged by AI, and increasing the overall volume of processed files without increasing staff.

Client Journey Improvement : Analysis speed has direct impact on client experience. Faster document processing accelerates key processes like account opening or credit granting, which today occur mainly via digital channels. This fluidity and responsiveness are major competitive advantages in a market where digital experience quality has become a selection criterion.

4 Roadmap and Evolution Perspectives

The functional prototype and promising results obtained constitute a solid foundation for the future. This section draws a realistic evolution trajectory, from concrete next steps of industrialization to more distant research axes that will maintain technological advantage against constantly evolving threats.

4.1 Industrialization Plan and Short-term Improvements

The transition from this prototype to a robust production tool follows a pragmatic three-phase action plan, aimed at bridging the gap between development environment and real-world requirements.

1. **Phase 1 : Pilot Project and Data Collection.** The absolute priority is overcoming data limitations. This phase will consist of deploying the solution in a controlled perimeter, with a

group of business analysts. The objective will be twofold : on one hand, gather their qualitative feedback on interface and alert relevance (notably false positive management) ; on the other hand, and especially, begin building an internal dataset, anonymized and representative of documents processed by the bank.

2. **Phase 2 : Model and UX Refinement.** Data and feedback collected during the pilot will feed a continuous improvement cycle. Models, particularly SegFormer, will be retrained on this new corpus to refine their precision and reduce "prediction noise". In parallel, the user interface will be adjusted to best meet analysts' needs and suggestions.
3. **Phase 3 : Technical Integration and Deployment.** Once the model is stabilized and validated on internal data, the last step will be its complete integration into AWB information systems. Building on already envisioned MLOps configurations (Docker/K8s), the system will be exposed as a secure internal API, ready to be consumed by different business processes (KYC, credit granting, etc.) for large-scale automation.

4.2 Strategic Evolutions : from Security to Explainability

Beyond industrialization, solution sustainability requires anticipating future evolutions of fraud and AI technologies.

4.2.1 Defense against Adversarial Attacks

Fraud sophistication will not stop at simple image retouching. The next major threat will come from adversarial attacks, where tiny modifications, imperceptible to humans, are applied to a document with the explicit goal of deceiving AI models. Future research must therefore focus on hardening our models, exploring "Adversarial Training" techniques to make them more robust against this type of targeted manipulation.

4.2.2 Towards Explainability through Generative AI

Explainability is the key to trust. A strategic evolution therefore consists of going beyond simple detection *heatmap* to provide justification in natural language. Inspired by approaches like *FakeShield* (ICLR 2025), we envision chaining *SegFormer* output with an LLM from the Llama family.

Concretely, we have already built a first prototype : we load the image (local or via URL), encode it in Base64, trace boxes detected as suspicious, then transmit the image and description of

each zone to the « **meta-llama/llama-4-maverick-17b-128e-instruct** » model hosted at Groq. This multimodal model (≈ 17 B parameters) merges visual representation and textual context to generate, at low temperature (0.1), a justification. It also specifies the nature of the altered field, the benefit sought by the fraudster and potential impact, thus transforming simple pixel highlighting into explanation usable by the analyst. We are still in testing phase to evaluate reliability, cost and latency of this chain, but first results show significant gain in interpretability for business users.

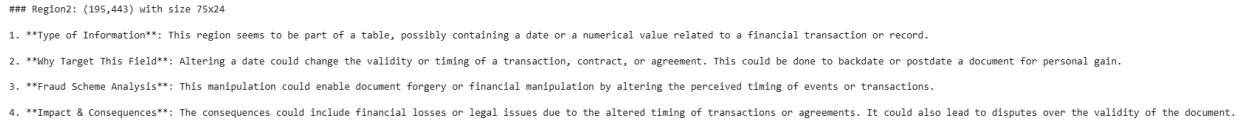


FIGURE 41 – Example produced by our prototype of natural language explanation (generated by *meta-llama/llama-4-maverick-17b-128e-instruct*).

4.3 Long-term Research Horizons : the Edge AI Era

In the longer term, a major technological breakthrough could come from Edge AI. For use cases requiring maximum confidentiality and instant response, such as verifying a document in a branch or directly from the client's mobile application, we can imagine deploying lightweight versions of our models directly on the device (smartphone, tablet). This "Edge AI" approach would allow performing analysis without the sensitive document ever leaving the user's device. This would offer unparalleled confidentiality and security guarantees, while eliminating network latency, paving the way for a new generation of decentralized trust tools.

Chapter Conclusion

This final chapter has completed our journey from problem to solution. Through experimental evaluation, we validated the effectiveness of our approach with a Recall of 93.95
The traced roadmap, from pilot deployment to Edge AI exploration, offers a realistic evolution path that will maintain the bank's technological edge while addressing emerging threats in an increasingly digitized financial landscape.

GENERAL CONCLUSION

Faced with sophisticated document fraud, a critical vulnerability generated by the digital transformation of the banking sector, and given the inefficiency of manual controls, this project aimed to design and deploy an artificial intelligence system for precise and rapid detection of alterations.

Our structured approach progressed from strategic analysis (Chapter 1) and state-of-the-art review (Chapter 2) to the design of a robust microservices architecture (Chapters 3 and 4). This approach culminated in a technical implementation phase (Chapter 5) that consisted of building a composite dataset and developing two technological pillars : a significantly improved **SegFormer** vision model and a semantic analysis pipeline driven by **LLaMA-3** via advanced prompt engineering.

The experimental results (Chapter 6) validated the effectiveness of our approach. The system combines excellent reliability on authentic documents with very high sensitivity to alterations, in line with the objective of detecting the vast majority of frauds. The resulting functional application prototype confirms the solution's viability and its potential to reduce financial risk, optimize operational efficiency and secure client journeys.

Our analysis acknowledges several limitations, notably a dependence on public data that cannot perfectly replicate the production context. The assumed bias in favor of exhaustive detection implies a false positive rate requiring human verification, while reliance on LLMs carries a residual risk of erroneous interpretation.

This project nevertheless constitutes a solid foundation. The defined roadmap opens clear perspectives : a pilot project to collect internal data and refine the model, followed by research on defense against adversarial attacks and improvement of AI explainability. Conducted within Attijariwafa Bank's Digital Center, this work provides a concrete contribution to the group's digital transformation strategy and demonstrates the potential of artificial intelligence to strengthen trust and security at the heart of tomorrow's banking.

REFERENCES

AI/ML Models and Architectures

[1] Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2020). *An Image is Worth 16x16 Words : Transformers for Image Recognition at Scale*. arXiv. https://arxiv.org/abs/2010.11929

[2] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., et al. (2014). *Generative Adversarial Networks*. arXiv. https://arxiv.org/abs/1406.2661

[3] Stanford University. *CS231n : Convolutional Neural Networks for Visual Recognition Course Notes*. https://cs231n.github.io/

[4] Stanford University. *CS224n : Natural Language Processing with Deep Learning Course Notes*. https://web.stanford.edu/class/cs224n/

[5] Distill.pub (2021). *A Gentle Introduction to Graph Neural Networks*. https://distill.pub/2021/gnn-intro/

Image Analysis and Segmentation Models

[6] Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net : Convolutional Networks for Biomedical Image Segmentation*. arXiv. https://arxiv.org/abs/1505.04597

[7] Google. *DeepLab in TensorFlow Model Garden*. https://github.com/tensorflow/models/tree/master/research/deeplab

[8] Xie, E., Wang, W., Yu, Z., et al. (2021). *SegFormer : Simple and Efficient Design for Semantic Segmentation with Transformers*. arXiv. https://arxiv.org/abs/2105.15203

Large Language & Speech Models

[9] Meta AI. *Meta Llama 3 Official Page*. https://ai.meta.com/blog/meta-llama-3/

[10] OpenAI. (2022). *Introducing Whisper*. Blog Post. https://openai.com/research/whisper

Platforms, Frameworks & Tools

[11] Groq. *Official Website*. https://groq.com/

[12] MongoDB, Inc. *MongoDB Official Website*. https://www.mongodb.com/

[13] Meta Platforms, Inc. *React Official Website*. https://react.dev/

[14] Pallets Projects. *Flask Official Website*. https://flask.palletsprojects.com/

[15] The GIMP Team. *GIMP (GNU Image Manipulation Program)*. https://www.gimp.org/

[16] Google. *Tesseract OCR on GitHub*. https://github.com/tesseract-ocr/tesseract

[17] Jaided AI. *EasyOCR on GitHub*. https://github.com/JaidedAI/EasyOCR

[18] Baidu. *PaddleOCR on GitHub*. https://github.com/PaddlePaddle/PaddleOCR

[19] Roboflow. *Official Website*. https://roboflow.com/

[20] MIT CSAIL. *LabelMe Official Website*. http://labelme.csail.mit.edu/Release3.0/

Datasets

[21] University of Maryland. *RVL-CDIP Dataset*. https://www.cs.cmu.edu/~aharley/rvl-cdip/

[22] Maan, K., Gupta, A., et al. (2023). *DocTamper : An Extensive Benchmark for Document Tampering Detection*. arXiv. https://arxiv.org/abs/2308.13673

[23] Chinese Academy of Sciences. *CASIA Image Tampering Detection Evaluation Database*. http://forensics.idealtest.org/

[24] Wen, B., Zhu, Y., et al. (2016). *COVERAGE – A Novel Database for Copy-Move Forgery Detection*. IEEE Xplore. https://ieeexplore.ieee.org/document/7784337

Fraud and Crime Reports

[25] Entrust. (2025). *Identity Fraud Report 2025*. https://www.entrust.com/resources/report/identity-fraud-report-2025

[26] Sumsub. (2025). *State of Identity Verification – iGaming & Finance 2025*. https://sumsub.com/resources/reports/state-of-identity-verification-2025

[27] LexisNexis Risk Solutions. (2024). *True Cost of Fraud™ Study – EMEA 2024*. https://risk.lexisnexis.com/insights-resources/research/true-cost-of-fraud-emea-study

[28] INTERPOL. (2024). *Africa Crime Trend Report 2024*. https://www.interpol.int/en/News-and-Events/News/2024/Africa-Crime-Trend-Report-2024

[29] African Union. (2024). *Cross-Border Crime Report*. https://au.int/en/documents/2024-cross-border-crime-report

[30] World Bank. (2024). *Documentary Fraud Risk Note – West Africa*. https://documents.worldbank.org/en/publication/documents-reports/documentary-fraud-risk-west-africa-2024

[31] CTMS. (January 28, 2025). *Documentary and Identity Fraud in Francophone Africa*. https://www.ctms.fr/blog/fraude-documentaire-identitaire-afrique-francophone