

L'IA au Cœur des processus métiers

Cas d'usage en gestion des sinistres



Mansour SOW
Leader Manager - CEGC



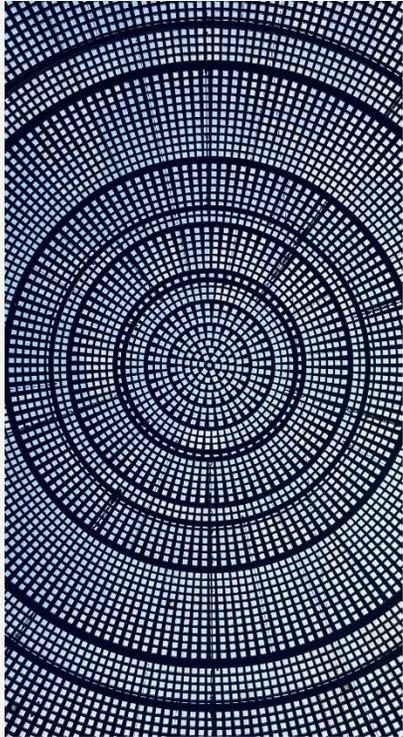
Loup ORTIZ
Senior consultant - MILLIMAN



Adel CHERCHALI
Senior consultant - MILLIMAN

Le développement de l'IA au service des processus d'assurance

Des solutions analytiques et innovantes sur différentes étapes de la chaîne de valeur



Marketing et Communication

- Efficacité commerciale de la distribution et développement de produits
- Ciblage et Valeur client, analyse comportementale, analyse de la performance des distributeurs
- Accélération du processus de souscription : images, emails, données de centres d'appels

Tarification et Souscription

- Sophistication de la tarification
- Rationalisation du processus de calcul de la prime pure
- Simplification de la souscription, sélection et scoring du risque
- Exploitation de sources de données externes

Modélisation des risques

- Développement de méthodologies analytiques et mathématiques avancées : pandémie, cyber, climatiques
- Hypothèses et risques actuariels et financiers (risques biométriques & comportementaux, sinistralité, scénarios économiques...)

Analyse des sinistres

- Optimisation du processus de gestion des sinistres (traitements des emails, reconnaissance d'images)
- Détection de fraude
- Analyse des bases sinistres et modèles prédictifs (yc données textuelles)

Smart Reporting et Data Viz

- Production automatisée de rapports narratifs des processus actuariels
- Agents conversationnels, suivis de KPI
- Qualité des données (détection d'anomalies et mesures correctrices)
- Smart reporting (visualisation)

Sommaire

- **Problématique** : l'indemnisation en assurance crédit, un processus complexe
- **Technologies** : CNN, Vision Transformer et LLM
- **Applications & Résultats**

Problématique : l'indemnisation, un processus complexe

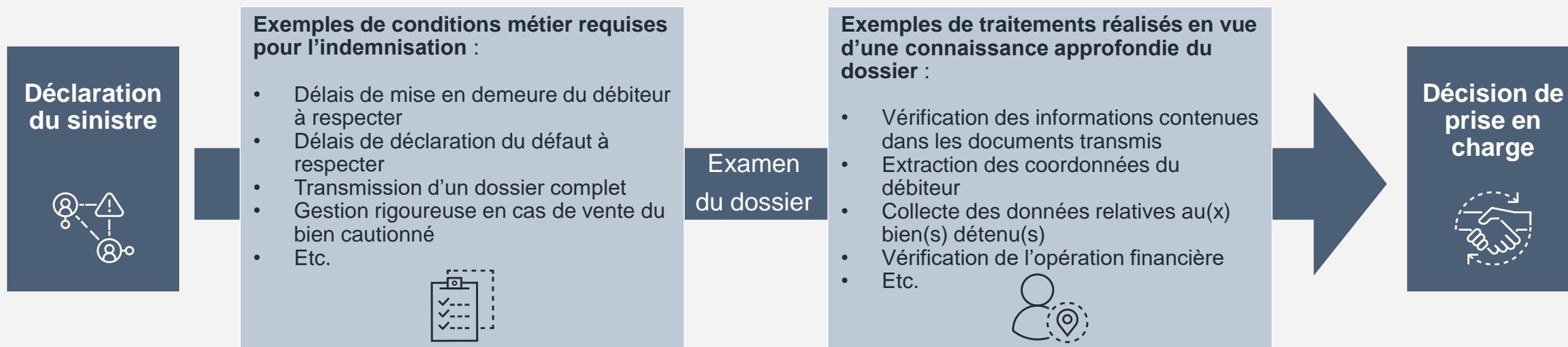
L'indemnisation en Assurance Crédit Caution : un processus complexe

L'indemnisation est un processus central dans la relation établie entre un assureur et ses clients ou partenaires

La **gestion des sinistres** en assurance crédit caution est un **processus complexe**, de la déclaration de la déchéance du terme jusqu'à l'indemnisation proprement dite. Un **contrat cadre** établi entre les filiales bancaires et l'organisme de caution en fixe les termes.

L'équipe métier de gestionnaires sinistres est en charge de l'instruction des dossiers et est confrontée

- A des **enjeux de performance** : la bonne instruction des dossiers conditionne le recouvrement futur
- A une **forte contrainte temporelle** : les délais d'indemnisation et de prise en charge sont définis par l'accord cadre



Lecture automatique de documents PDF

La diversité du corpus documentaire

Courrier de mise en demeure

Date, nombre d'impayés, CRD...



Bulletin de salaire

Coordonnées, Revenus



Tableau d'amortissement

Taux d'intérêt, mensualités

Mois	Capital restant	Intérêt	Mensualité	Capital restant	Capital restant
1	11 700 €	0 520 €	3 214 €	0 €	141 400 €
2	11 700 €	0 520 €	3 214 €	0 €	132 740 €
3	11 700 €	0 519 €	3 214 €	0 €	123 921 €
4	11 700 €	0 517 €	3 214 €	0 €	114 703 €
5	11 700 €	0 520 €	3 214 €	0 €	105 084 €
6	11 700 €	0 527 €	3 214 €	0 €	95 067 €
7	11 700 €	0 530 €	3 214 €	0 €	84 710 €
8	11 700 €	0 535 €	3 214 €	0 €	74 040 €
9	11 700 €	0 538 €	3 214 €	0 €	63 087 €
10	11 700 €	0 540 €	3 214 €	0 €	51 884 €
11	11 700 €	0 544 €	3 214 €	0 €	40 461 €
12	11 700 €	0 547 €	3 214 €	0 €	28 861 €
13	11 700 €	0 549 €	3 214 €	0 €	17 090 €
14	11 700 €	0 550 €	3 214 €	0 €	5 191 €
15	11 700 €	0 551 €	3 214 €	0 €	0 €

Relevés de compte bancaire

Mensualité, transaction remarquable



Relevés d'imposition

Date, adresse, montants



Documents d'identité

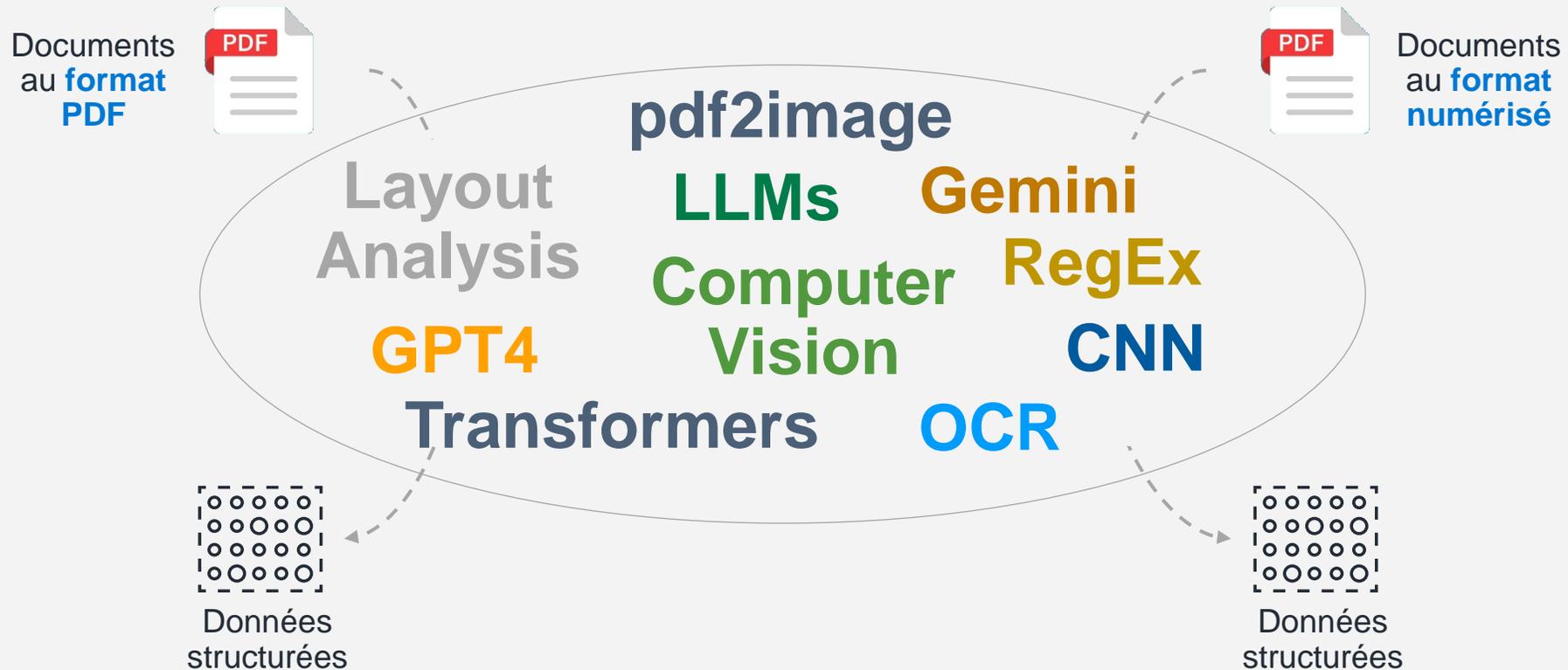
Nom, prénom, date d'établissement...



Lecture automatique de documents PDF

La multiplicité des technologies

La prise de décision d'indemnisation résulte de l'examen rigoureux et approfondi du dossier de **demande de prise** en charge transmis par l'établissement bancaire constitué d'une **multitude de documents PDF aux caractéristiques spécifiques**.



Technologies : CNN, Vision Transformer et LLM

Vision Large Language Model (VLLM)

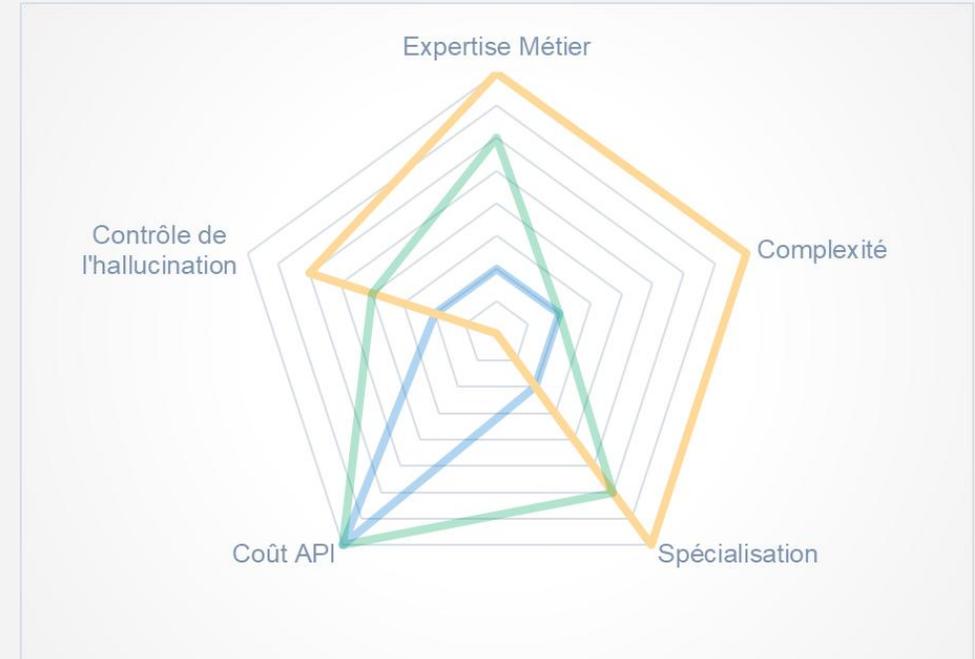
Les stratégies d'adaptation des LLM aux tâches spécifiques

La plupart des **modèles pré-entraînés** sont calibrés sur un ensemble d'informations qui ne sont pas spécifiques à un domaine en particulier :

- Obtention de réponses « générales » reflétant une incapacité à identifier un ou plusieurs documents sources qui fondent la réponse
- Performance perçue comme limitée par un utilisateur expert de son domaine

Pour répondre à ces problématiques, on identifie deux stratégies principales :

- **Approche 1 : Fine-tuning d'un modèle de vision spécialisé sur une tâche spécifique** → Continuer l'entraînement du modèle sous-jacent pour aboutir à un modèle très spécialisé sur une tâche spécifique (exemple: détection d'objets) **Offre plus de contrôle et de personnalisation, mais nécessite plus de ressources et de temps de développement.**
- **Approche 2 : Prompt Engineering & utilisation d'un modèle pré-entraîné :** GPT4-vision, Google Cloud Vision API et autres alternatives open source offrent une compréhension générale du contenu visuel et linguistique. **Nécessite moins de ressources initiales, mais peut entraîner des coûts continus et moins de flexibilité sur des tâches spécifiques.**



Reconnaissance d'image

Algorithme CNN

Prédiction d'objets via Convolutional Neural Network (CNN)

- L'architecture des **réseaux neuronaux convolutifs profonds** est utilisée pour extraire des caractéristiques d'une image, qu'il transmet à une couche de détection afin de prédire si un objet appartient à l'image
- L'algorithme prédit des boxes autour de **l'objet avec un score de confiance**.

L'algorithme procède comme suit :

- **Réduction du problème** : l'image est divisée en une grille de $(S \times S)$ cellules. Si le centre d'un objet se trouve dans une cellule de la grille, cette cellule est responsable de détecter cet objet.
- **Bounding Boxes** : chaque cellule de la grille prédit des bounding boxes et des scores de confiance pour ces boxes comme suit

$$\text{Confidence} = \mathbb{P}(\text{Object}) \times \text{IOU}$$

avec *IOU* la métrique « intersection over union » entre la boxe prédite par l'algorithme et la vraie boxe annotée par un humain :

$$\text{IOU} = \frac{\text{area of overlap}}{\text{area of union}}$$

- **Prédiction des classes** : chaque cellule prédit la probabilité de classe de l'objet (un vecteur de scores dont la longueur est égale au nombre de classes).



Figure : IOU score

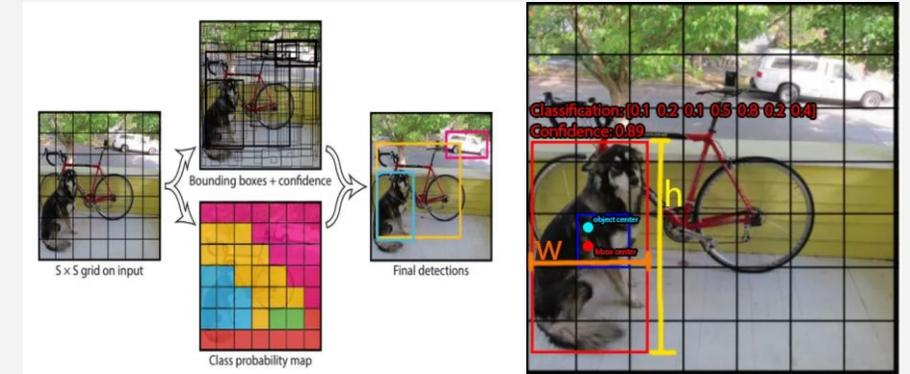


Figure : Processus d'entraînement
(source: Redmon et al. 2016)

Extraction de texte

Algorithme de Vision Transformer (1/2)

Algorithme Vision Transformer

1. Conversion du fichier PDF en **image**
2. Tokenization de l'image (**découpage de l'image en blocs**)
3. Conservation de la trace de la position des blocs pour maintenir la structure hiérarchique logique de l'image
4. Exploitation du **mécanisme d'attention** pour récupérer la partie la plus importante de la structure de l'image et utiliser les poids du réseau neuronal comme embeddings. Les embeddings résument la structure de l'image (**image encodée**)
5. Passage au **décodeur** qui génère séquentiellement le texte basé sur l'image encodée

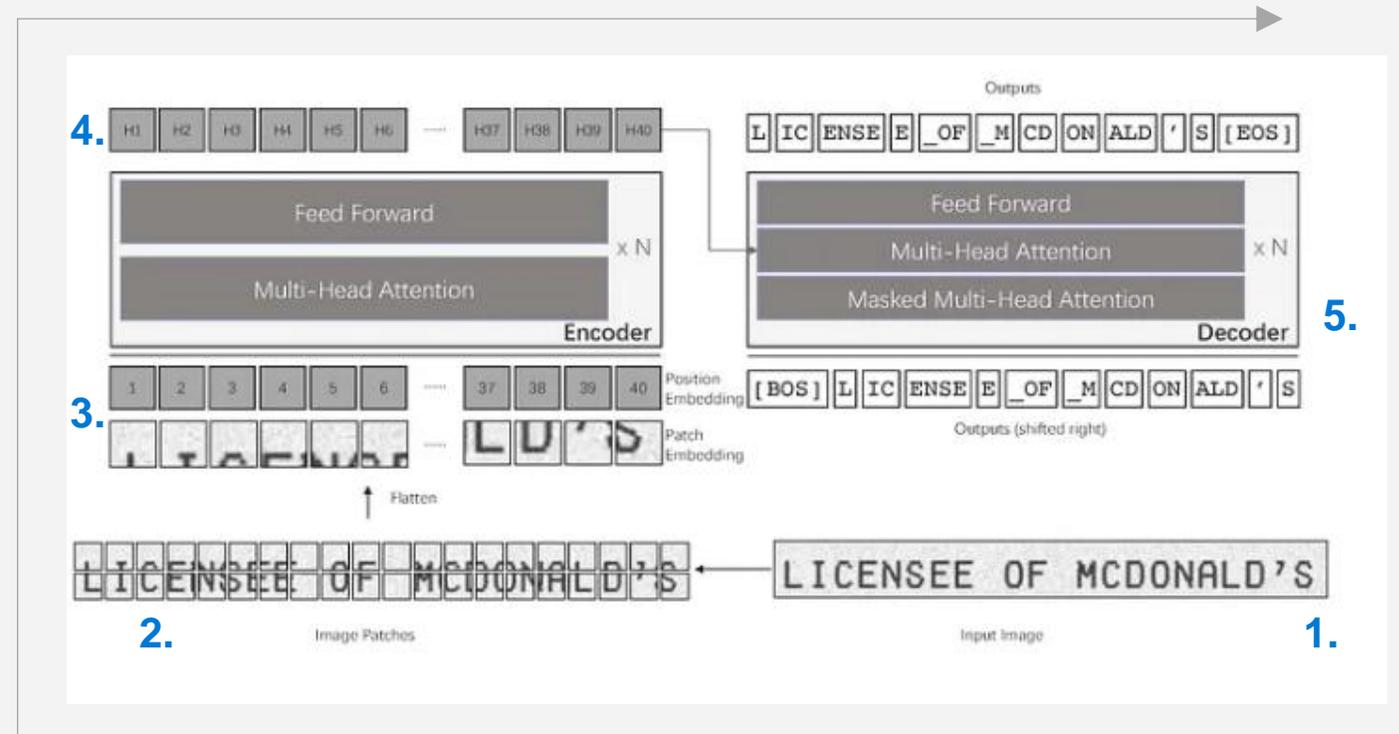


Figure : Processus d'entraînement
(source: Li et al. 2022)

Extraction de texte

Algorithme de Vision Transformer (2/2)

Entraînement

- Le modèle est entraîné sur des **millions de documents PDF** comportant des formules mathématiques, du texte, des graphiques et des tableaux
- Le jeu de données est augmenté en utilisant plusieurs techniques d'**augmentation de données** pour le rendre robuste aux PDF scannés par exemple.

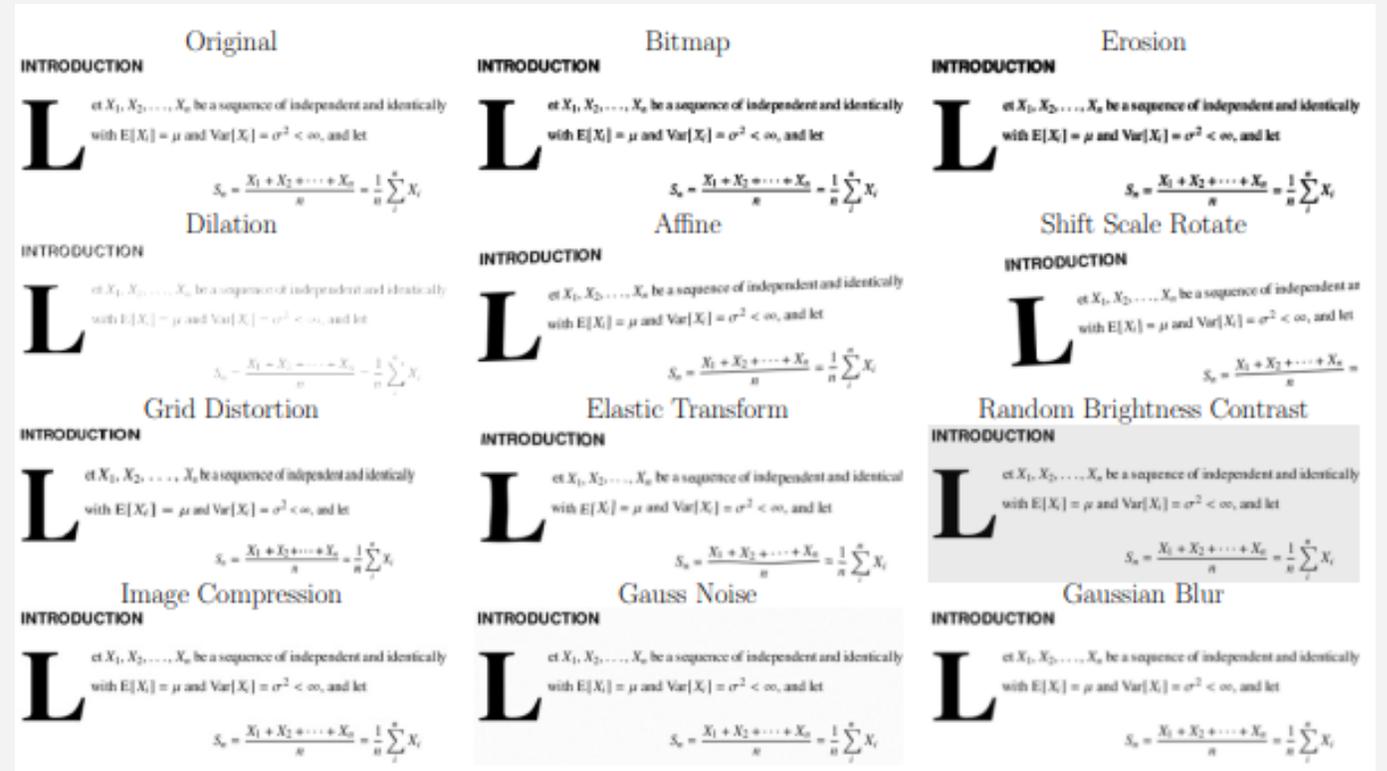


Figure : Processus d'entraînement
(source: Blecher et al. 2023)

Compréhension de la structure d'un document

Algorithme de Vision Transformer + CNN

Vision Transformer avec prédiction de la structure du document

- Utilisation d'un **Vision Transformer** avec une couche **CNN** pour prédire **la structure du PDF** et détecter les sections, textes, graphiques et tableaux au sein d'une image.
- Utilisation des sections détectées pour **analyser les sections pertinentes** (générer un résumé d'une image par exemple...).
- Identification de chaque **subdivision du document** et de la typologie correspondante (tableau, image, graphique, paragraphe etc.) afin d'utiliser la **technologie adéquate pour l'extraction des informations** sous-jacentes

SEMANAL MACRO
CLAVES DE LA SEMANA PASADA

EEUU
El PIB de los EE.UU. se incrementó con la recuperación del primer trimestre y desde del mes de julio que muestra una aceleración de la cifra mensual de una décima hasta 0,3% y el segundo de dos hasta 0,7%. Por lo tanto se situó en 0,4% en términos anuales, lo que se sitúa en un nivel de crecimiento elevado, ya que el nivel de actividad de un momento se cancela en el del otro del primer al segundo trimestre por la misma materia de negocio. En este sentido, se trata de un punto positivo en la medida en que si se ha mantenido estable en el 2020, en su consecuencia, desde la semana pasada con 0,3%, la base de ingresos anualizados supone un registro bastante único con el crecimiento del 2% y por otro lado, el crecimiento de esta semana es de 0,3%.

EUROZONA
El crecimiento de la zona euro se recuperó tras haberse estancado la cual ya alcanzó cuatro meses consecutivos en este primer trimestre, al avanzar en 0,1%. El PIB de la zona euro pasó un trimestre sin cambios que a primera instancia, a un trimestre de crecimiento de 0,1%.

EUROZONA
El crecimiento de la zona euro se recuperó tras haberse estancado la cual ya alcanzó cuatro meses consecutivos en este primer trimestre, al avanzar en 0,1%. El PIB de la zona euro pasó un trimestre sin cambios que a primera instancia, a un trimestre de crecimiento de 0,1%.

Claves para esta semana

País	Indicador	Valor	Unidad	Fecha	Unidad	Fecha
EEUU	PIB	21.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Consumo privado	18.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Exportaciones	1.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Importaciones	1.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	PIB Core	18.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Consumo privado Core	15.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Exportaciones Core	1.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Importaciones Core	1.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	PIB Core	18.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Consumo privado Core	15.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Exportaciones Core	1.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Importaciones Core	1.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01

SEMANAL MACRO
CLAVES DE LA SEMANA PASADA

EEUU

País	Indicador	Valor	Unidad	Fecha	Unidad	Fecha
EEUU	PIB	21.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Consumo privado	18.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Exportaciones	1.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Importaciones	1.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	PIB Core	18.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Consumo privado Core	15.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Exportaciones Core	1.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Importaciones Core	1.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01

Zona Euro

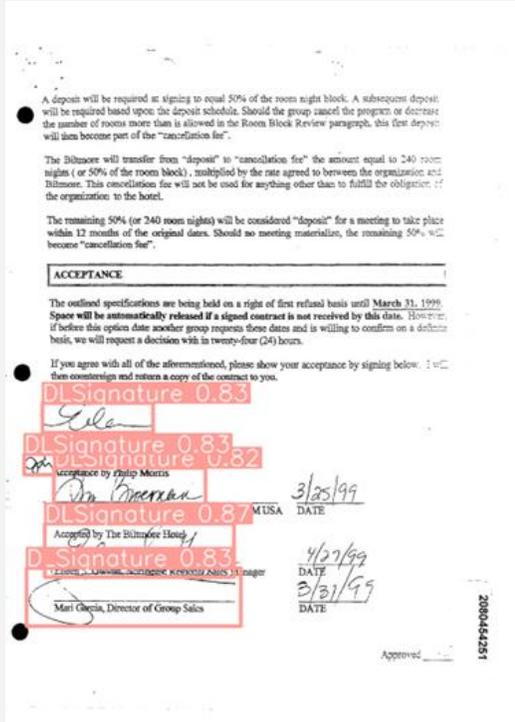
País	Indicador	Valor	Unidad	Fecha	Unidad	Fecha
EEUU	PIB	21.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Consumo privado	18.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Exportaciones	1.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Importaciones	1.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	PIB Core	18.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Consumo privado Core	15.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Exportaciones Core	1.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01
EEUU	Importaciones Core	1.000	Millones de millones de dólares	2020-07-01	2020-07-01	2020-07-01

Applications & Résultats

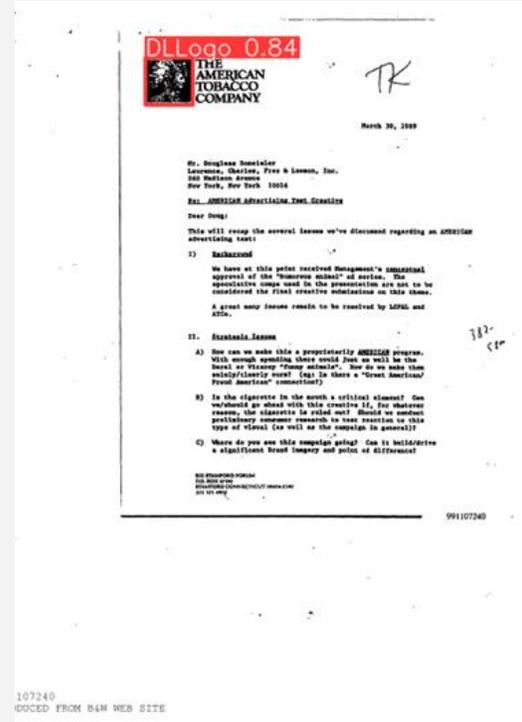
Cas d'usage n°1 : détection de signatures

Approche 1 : Réseaux de neurones convolutionnels pour la détection d'objet

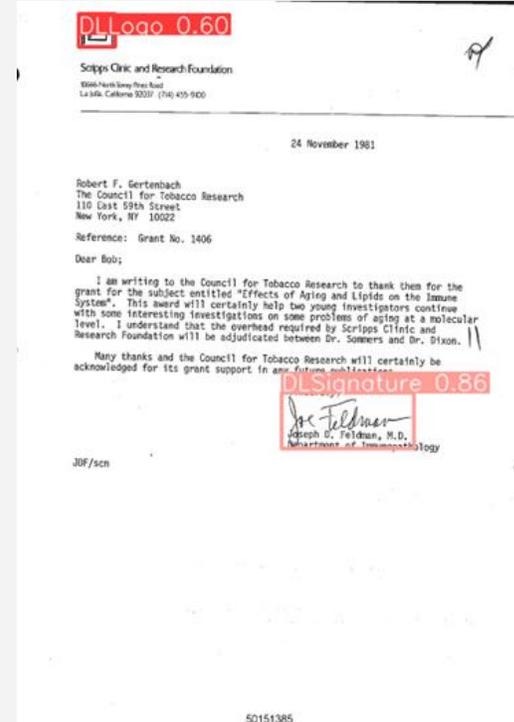
Calibrage des hyperparamètres d'une version pré-entraînée d'algorithmes de deep learning afin de spécialiser le modèle pour la détection de signatures



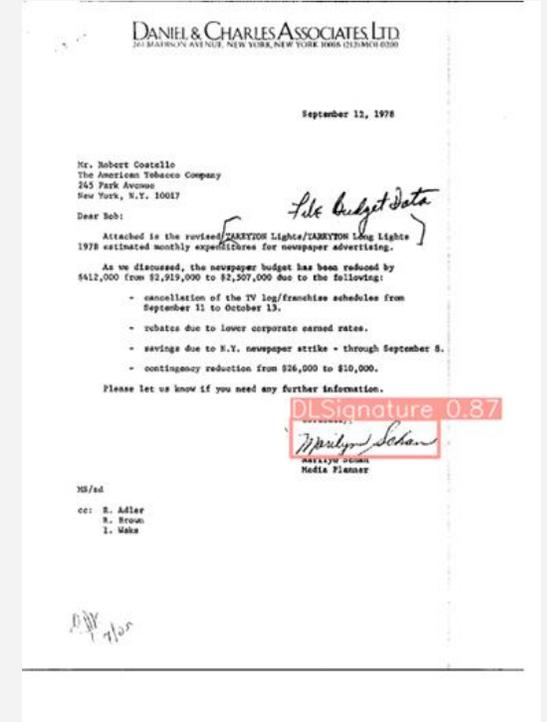
Détection de plusieurs signatures sur une même page



Détection du logo de l'entreprise ; absence de signature



Détection d'une signature et du logo de l'entreprise



Détection d'une signature et annotation manuscrite qui n'est pas une signature

Cas d'usage n°1 : détection de signatures

Approche 2 : LLM multimodal

Exploitation des **modèles de LLM multimodaux** tels que GPT4 – Vision capables de fournir une **description de documents numérisés**.

Les images sont traitées automatiquement par le biais du **prompt** ci-dessous à l'aide d'une API. La réponse est formatée en fichier json avec des champs tels que la description de l'image et le numéro de signature pour **automatiser le processus de vérification**.

Requête utilisateur



«

Étant donnée l'image jointe, fournissez les informations suivantes :

- Une description de l'image : en particulier, indiquez si l'image contient une signature
- Le numéro de la signature dans l'image

»

Réponse de GPT4-Vision

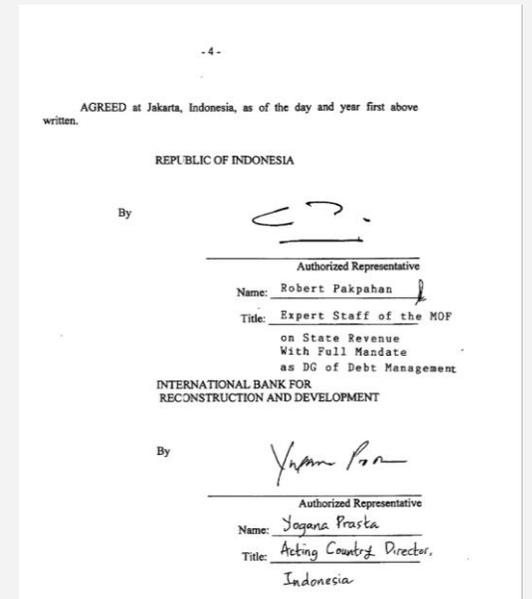


«

L'image montre un document avec du texte et **deux signatures**.

La première signature est apposée sous le titre "REPUBLIQUE D'INDONESIE" et **la seconde** sous le titre "BANQUE INTERNATIONALE POUR LA RECONSTRUCTION ET LE DEVELOPPEMENT".

»



Input (document numérisé)

Cas d'usage n°2 : traitement de la pièce d'identité

Pipeline sur mesure vs LLM multimodal

INPUT



n-page PDF files



Conversion
en images



Liste d'images

Algorithme de
détection d'objet



Liste d'images
recto/verso de CNI



Modèle de reconnaissance
optique de caractères (OCR)



Liste de chaînes
de caractères



Retraitement
RegEx

Nom : DURAND
Prénom : ANTOINE PAUL
Naissance : 01/01/1960
Délivrance: 01/01/2012
Expiration: 01/01/2022

Données extraites

OUTPUT



n-page PDF files

LLM
multimodal



- Nom : DURAND
- Prénom : ANTOINE PAUL
- Naissance : 01/01/1960
- Délivrance: 01/01/2012
- Expiration: 01/01/2022

Données extraites

Résultats et KPIs

Un choix technologique au service du processus métier

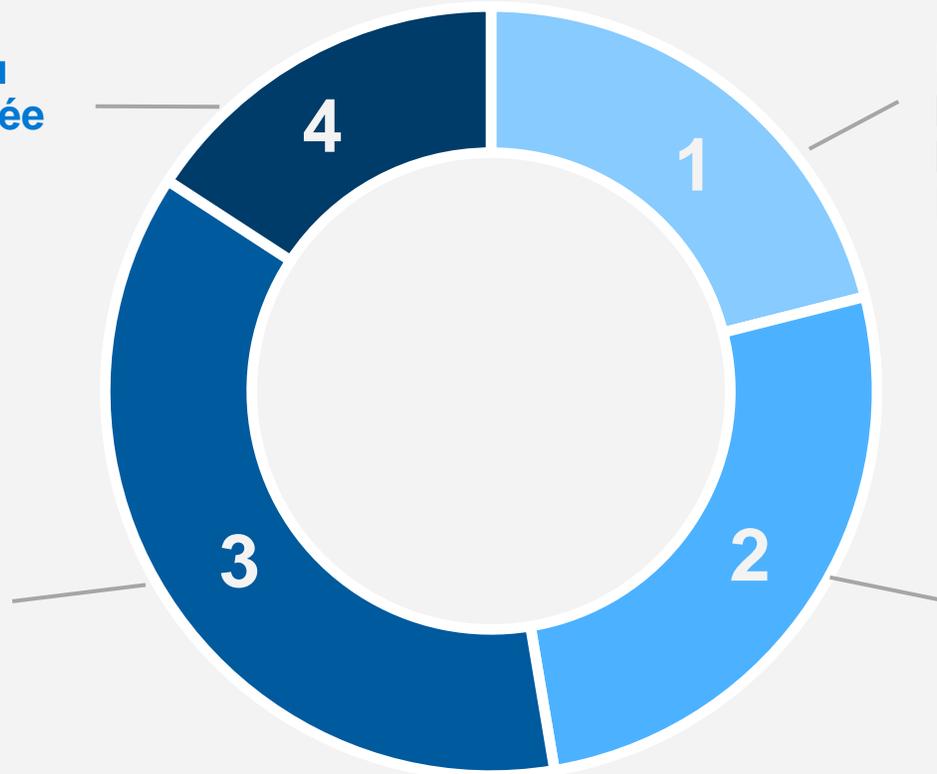
	Pipeline sur mesure – Fine Tuning		LLM multimodal	
			Usage général	Usage OCR spécifique
Précision	++	Précision et robustesse d'un modèle entraîné à la réalisation d'une tâche spécifique	+	++
Hallucination	+	Score de confiance associé à la prédiction du modèle	++	+
Ressources hardware	GPU/RAM		Appel API	
Coûts de développements	€€	Labellisation des données, conception, calibrage, suivi de la performance et monitoring	€	Prompt engineering
Coûts d'utilisation	-		€€	Coût proportionnel à l'utilisation avec incertitude future (inflation, nouveaux modèles...)

La réussite du projet : du diagnostic initial à l'adoption métier

Répartition des efforts pour la réussite du projet

L'adoption de la solution par les métiers repose sur une **conduite du changement progressive et adaptée**

La **conception de la solution** nécessite une **étroite collaboration** entre les gestionnaires de sinistres, les data scientists et la direction des systèmes informatiques pour un **déploiement optimal**



La **compréhension approfondie** du processus métier est indispensable pour poser le **bon diagnostic** et **prioriser les tâches à automatiser**

L'utilisation des **technologies adéquates** pour l'automatisation de chaque opération métier est nécessaire pour aboutir à la **réussite du PoC**



Merci

Mansour SOW

mansour.sow@c-garanties.com

Loup ORTIZ

loup.ortiz@milliman.com

Adel CHERCHALI

adel.cherchali@milliman.com