



## IA Générative : de l'analyse de documents à la mise en place de collaborateurs virtuels

**Yapei LI: Senior Manager Nexialog**



**Kezhan SHI : Deputy Director of Analytics - Actuarial and Data Science**

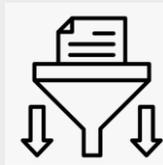


**Reda JARIR : CEO et co-fondateur d'URLab**



Notre atelier explore l'application de l'IA générative dans l'assurance, en se concentrant sur l'extraction, l'analyse et l'exploitation des informations. Il mettra également en lumière les perspectives offertes par l'IA agentique et l'orchestration IT pour introduire le concept de collaborateur virtuel.

## Extraction, Analyse, Structuration et Comparaison d'informations



1

**NexiaXtract** : un outil d'extraction et d'analyse des indicateurs financiers des rapports IFRS 17 et SFCR, utilisant la méthode Llama Index.



2

**DocMatrix**: Un outil d'analyse et de comparaison des conditions générales des contrats MRH avec la méthode RAG(Retrieval-Augmented Generation).

## Approche agentique x orchestration de l'infrastructure IT



3

### Création de collaborateurs virtuels

- Exploration de données et traitement de fichiers volumineux
- L'implémentation d'assistants IA

THINK SMART  ACT DIFFERENT

01

# Présentation de NexiaXtract

# Les rapports IFRS 17 et SFCR



La construction d'un benchmark d'indicateurs financiers est essentielle pour comparer la performance financière d'une entreprise à celle de ses pairs. L'extraction de ces données est l'un des sujets majeurs dans la construction du rapport.

## Les rapports IFRS 17

- ▶ Norme comptable internationale pour les contrats d'assurance, remplaçant IFRS 4.
- ▶ Mesure des engagements d'assurance sur la base des flux de trésorerie actualisés.
- ▶ Indicateurs: revenus, BEL, CSM, ratio des dépenses, ROE, ratio combiné etc

## Extrait d'un rapport IFRS 17



## Les rapports SFCR

- ▶ Document obligatoire dans le cadre de Solvabilité II
- ▶ Analyse de la situation financière et de solvabilité.
- ▶ Informations qualitatives sur la gouvernance et la gestion des risques.
- ▶ **Indicateurs** : Ratio de solvabilité, Fonds propres éligibles etc.

## Sommaire Rapport SFCR

SYNTHÈSE  
A. ACTIVITÉ ET RÉSULTATS  
B. SYSTÈME DE GOUVERNANCE  
C. PROFIL DE RISQUE  
D. VALORISATION À DES FINS DE SOLVABILITÉ  
E. GESTION DU CAPITAL  
ANNEXE  
Annexe I Modèles de déclaration quantitative

# Difficultés rencontrées et solutions trouvées

Pour extraire et structurer des informations à partir de ces différents rapports, nous avons rencontré des problèmes et trouvé des solutions, notamment grâce aux méthodes d'IA générative.

## Problématiques Rencontrées

- Présentation des informations sous forme de graphiques
- Identification des terminologies similaires
- Détection de la présence des différentes devises
- Détection des unités : millions, milliards
- Production d'analyses comparatives similaires

## Solutions Trouvées

- ✓ Utilisation de LlamaParse pour extraire des informations synthétiques à partir de rapports
- ✓ Construction d'un dictionnaire de vocabulaires similaires
- ✓ Identification des informations sur les devises et les unités avec LLM
- ✓ Génération d'analyses automatiques avec LLM

### Document original

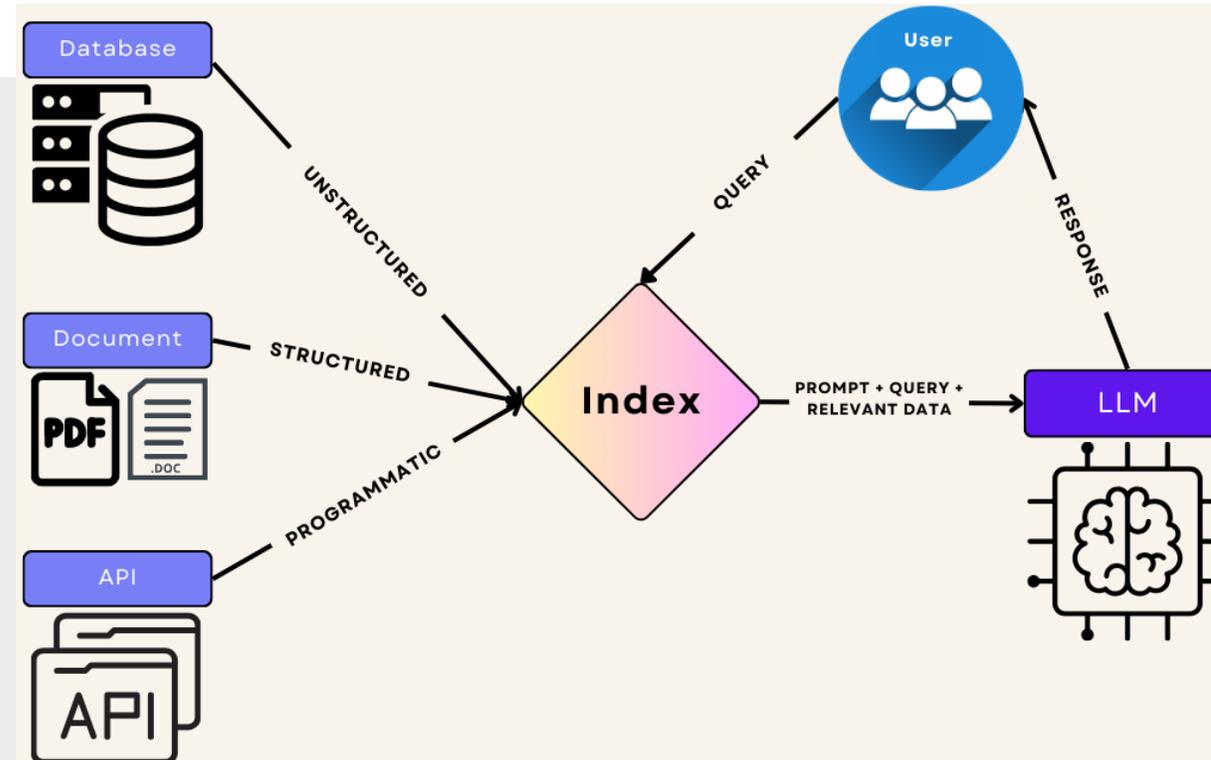


LlamaParse

### Résultats de parsing

```
AXA_Résultats_Annuels_2023_Presse.pdf
Job id: 6393bf06-034b-426d-9966-81fa08111921
Markdown Text JSON Images Layout XLSX Structured
# Résultats annuels 2023
# Croissance du chiffre d'affaires
|VIE|GESTION D'ACTIFS|DOMMAGES & PRÉVOYANCE SANTÉ| | | |
|---|---|---|---|---|---|
|53,0 Md€|32,0 Md€|16,1 Md€|1,6 Md€| |
| |+7%|+1%|-7%|-2%|
|+7% hors contrats Groupe1 non renouvelés|+7% hors contrats Groupe1 non renouvelés|+7% hors contrats Groupe1 non renouvelés| | | |
# Chiffre d'affaires 2023 vs. 2022
```

LlamaIndex est un cadre puissant permettant de connecter des données personnelles ou d'entreprise avec des modèles de LLMs. L'objectif est de faciliter l'intégration de données structurées et non structurées pour obtenir des réponses précises et pertinentes.



## 1. Récupération des sources de données

- **Base de données** : structurées (bases SQL, NoSQL)
- **Documents**: non structurées (PDF, des documents Word, des pages web)
- **API** : Intégration de données externes programmatique..

## 2. Indexation

- LlamaIndex **crée des structures d'index** pour faciliter la **recherche rapide et pertinente**.
- Permet de **gérer de grands volumes de données**.
- Rend les **requêtes plus efficaces**.

## 3. Interaction avec l'utilisateur

- **L'utilisateur envoie une requête** vers l'index.
- **L'index récupère les données pertinentes** associées à la requête.
- La requête, accompagnée des données récupérées, est combinée pour **former un prompt optimisé**.

## 4. Traitement par le LLM

- Le LLM (Modèle de Langage de Grande Taille) reçoit le prompt enrichi.
- Il **génère une réponse précise** basée sur le contexte et les données récupérées.
- La réponse est renvoyée à l'utilisateur.

## NexiaXtract

NexiaXtract est une solution innovante conçue pour simplifier l'extraction, l'analyse et le traitement des informations financières contenues dans les rapports IFRS 17 et SFCR des assureurs grâce aux Large Language Models. En automatisant ces processus complexes, il permet aux professionnels de la finance, aux actuaires et aux régulateurs d'accéder rapidement aux indicateurs clés, de faciliter leur analyse et de générer des commentaires.

IFRS 17 SFCR

### IFRS 17 – Extraction et analyse des indicateurs financiers

NexiaXtract automatise l'extraction des indicateurs clés des rapports IFRS 17, permettant aux actuaires et aux équipes financières de structurer, analyser et intégrer ces informations dans leurs analyses et modèles de projection.

Glissez et déposez jusqu'à 40 fichiers PDF



Drag and drop files here

Limit 200MB per file • PDF

Browse files

Lancer le processus

THINK SMART  ACT DIFFERENT

  
DiotSiaci

02

# Présentation de DocMatrix

# DocMatrix : tableau comparatif des documents

## Exemples d'application

- ✓ **Contrats d'assurance** : Comparez les exclusions, garanties et montants de primes
- ✓ **Documents financiers** : Analysez et mettez en parallèle des chiffres-clés
- ✓ **CV et profils candidats** : Identifiez le meilleur profil en fonction de vos critères

## Fonctionnement

- 1 **Upload des documents** : Chargez vos fichiers à analyser
- 2 **Posez vos questions** : Rédigez des questions spécifiques ou importez un fichier de questions préétabli
- 3 **Génération du tableau comparatif** : Les réponses sont extraites et organisées automatiquement
- 4 **Justification des résultats** : Chaque réponse est sourcée et annotée directement dans un PDF

The screenshot displays the DocMatrix application interface. On the left is a dashboard with 'Outils Populaires' including DocMatrix, Comparaison de deux versions, CR de Réunion, ChatDoc, Création de présentations, and Traduction. The main area shows a comparison table with three rows of questions and answers. A green arrow points from the dashboard to the table, and another points from the table to a source document on the right.

Question	Document 1	Document 2	Document 3
36 Quelle est la franchise applicable pour les dommages causés par des tempêtes, grêle ou neige ?	montant indexé indiqué sur vos Conditions Particulières	230 euros	Indiqué dans vos Conditions Particulières 30 %
37 Quel est le plafond d'indemnisation pour les dommages causés par des tempêtes, grêle ou neige ?	25 fois la valeur en euros de l'indice contractuel	Montants de garantie identiques et événements assimilés	À concurrence des plafonds de capitaux particuliers dans vos Conditions Particulières
38 Quelles sont les exclusions pour les dommages causés par des tempêtes, grêle ou neige ?	les cercueils et les urnes funéraires	Les dommages causés par la pluie, la neige, la grêle, le vent, le gel ou toute autre manifestation atmosphérique aux objets assés	Les événements cycloniques avec vents ≥ 145 km/h ou 215 km/h. Les dommages de jardins non assurés

The source document on the right contains the following text:

**2.3 Dégâts des eaux**

1. Les événements causés par « évènements climatiques - Gel ».

2. Les dommages causés par l'eau entrée par vos portes, portes-fenêtres, fenêtres, soupiraux, lucarnes, les conduits...

# Malgré les techniques courantes, des difficultés persistent pour une application efficace ...



Il existe de nombreuses ressources dans la communauté NLP pour concevoir des systèmes RAG, mais certaines limites subsistent.

## 1 Adaptation des chunks selon la question

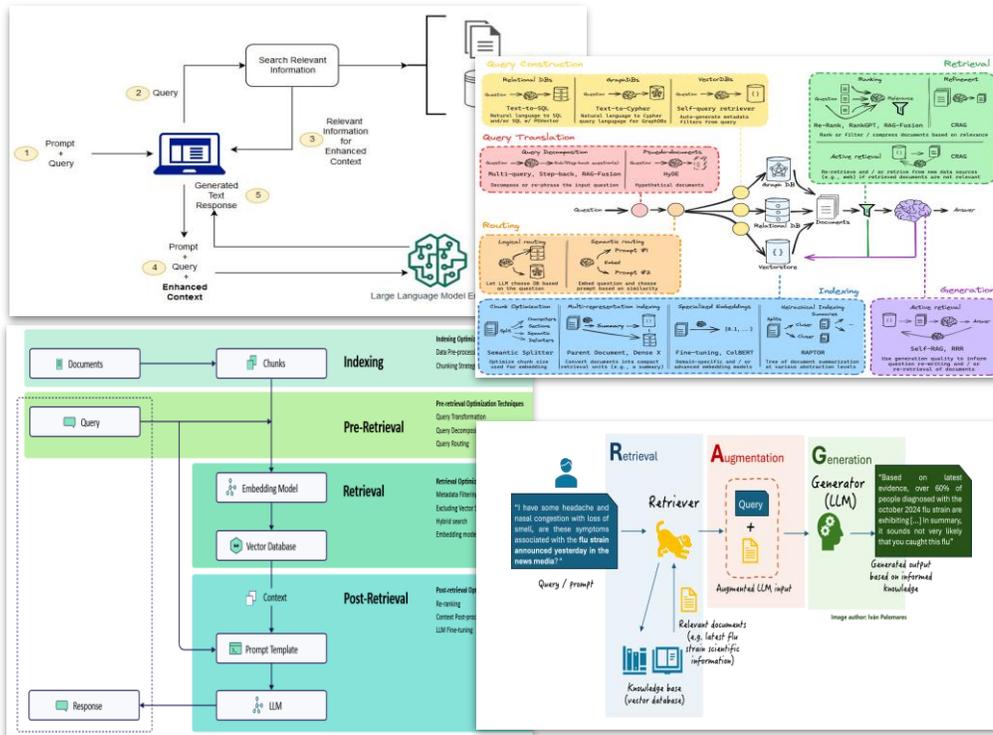
- La taille du texte pertinent varie considérablement selon la question posée.
  - **Exemple** : « Quelle est la date d'effet ? » → Une ligne suffit.
  - **Exemple** : « Résumer les exclusions des contrats » → Plusieurs pages peuvent être nécessaires.
- Un découpage uniforme des documents ne permet pas toujours d'obtenir des réponses adaptées.

## 2 Sous-performance due à l'utilisation de la similarité d'embedding

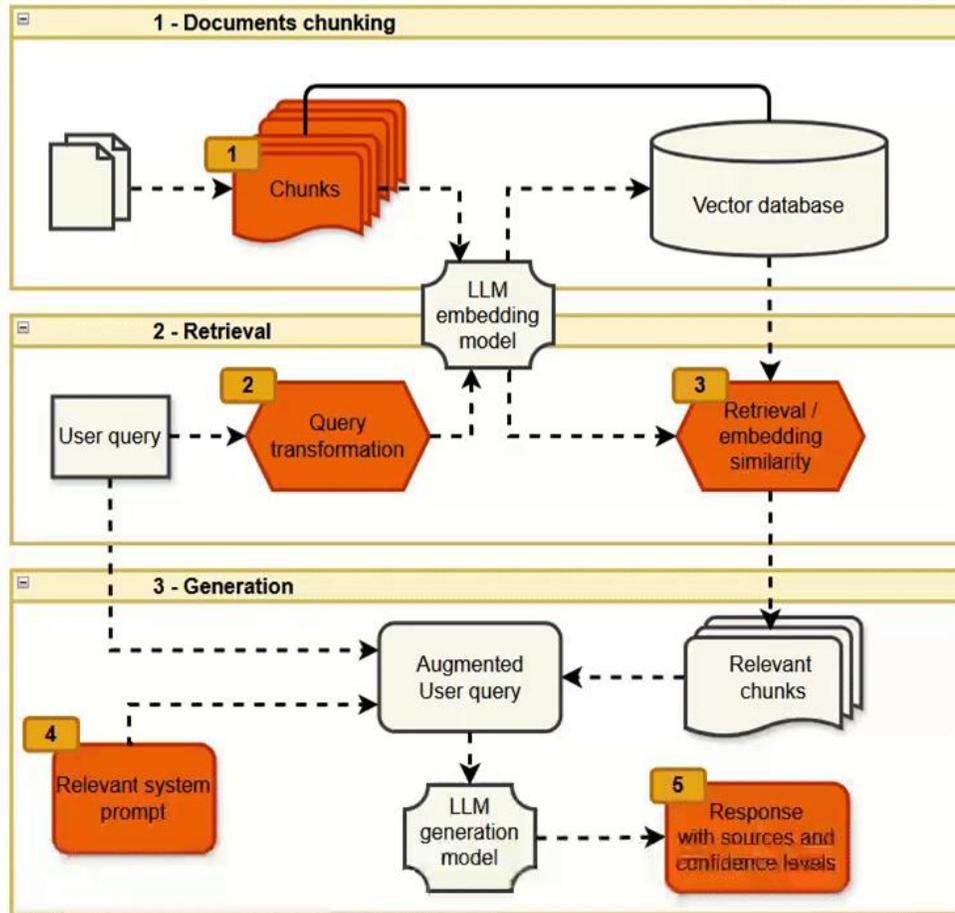
- Identifier des passages similaires via les embeddings **ne garantit pas** une réponse correcte.
- Une approche basée uniquement sur la similarité peut entraîner des résultats hors contexte.

## 3 Absence de justification des réponses

- Sans indication des passages précis utilisés, il est difficile de vérifier la fiabilité des réponses.
- Fournir des sources claires est essentiel pour assurer **transparence et traçabilité**.



# Des solutions innovantes et adaptées ont été mises en place pour surmonter ces défis ...



1

## Plusieurs types de parsing combinés

- Parsing pour la recherche de chunks (retrieval)
- Parsing pour la génération de réponse
- Parsing pour la justification des réponses

3

## Sélection de chunks

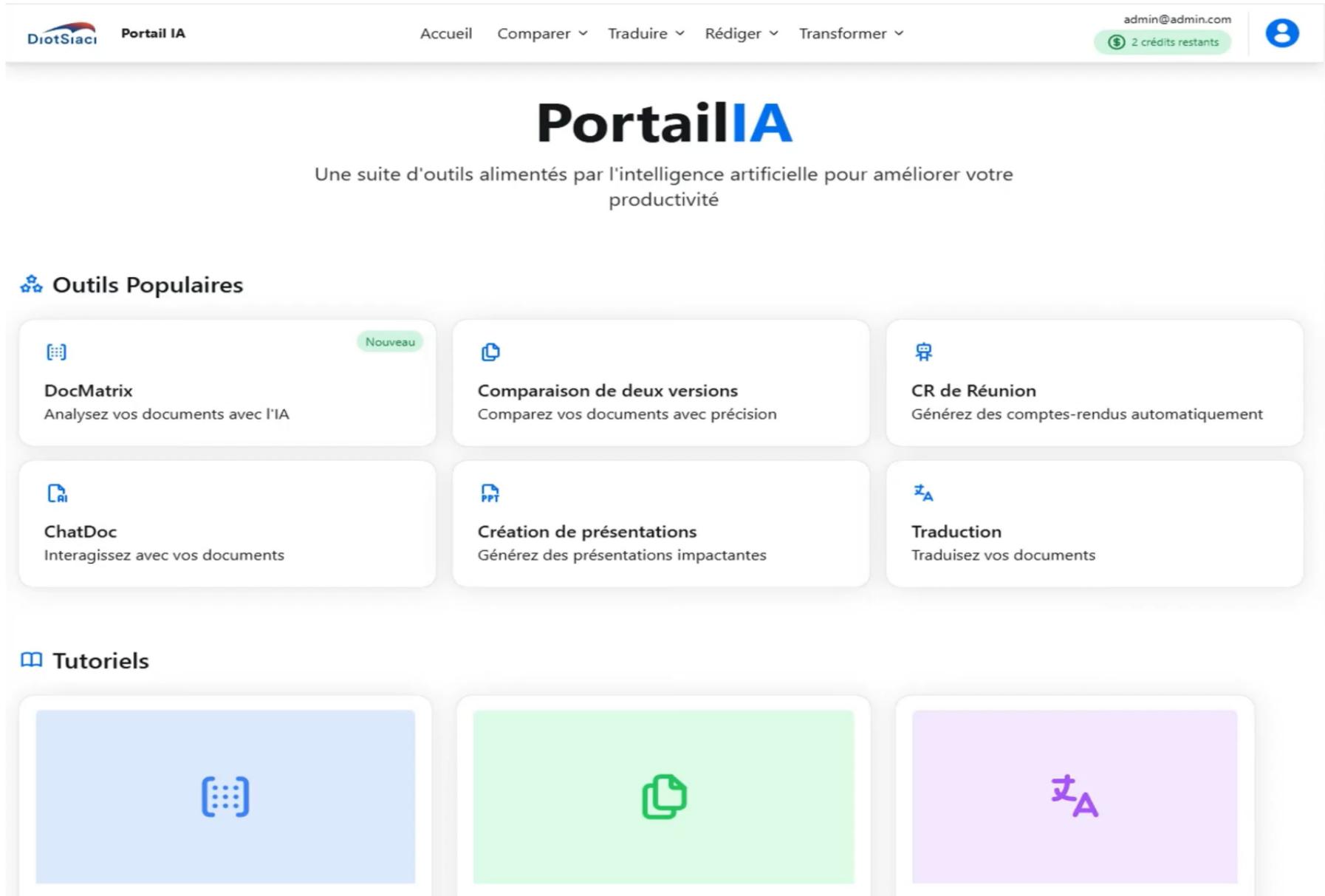
- Combinaison détection exacte et floue (fuzzy matching)
- Utilisation de la similarité embeddings
- LLM pour affiner la sélectionner les chunks

5

## Génération de réponses

- Réponse justifiée compréhensible et précise
- Agent pour générer une réponse complète
- Annotation des passages sources dans les pdf originaux

# Pour aboutir à une application finale offrant des réponses rapides, pertinentes et justifiées



The screenshot shows the 'Portail IA' dashboard. At the top, there is a navigation bar with the DiotSiaci logo, the title 'Portail IA', and menu items: 'Accueil', 'Comparer', 'Traduire', 'Rédiger', and 'Transformer'. On the right side of the navigation bar, the user is logged in as 'admin@admin.com' and has '2 crédits restants'.

## Portail IA

Une suite d'outils alimentés par l'intelligence artificielle pour améliorer votre productivité

### 🛠️ Outils Populaires

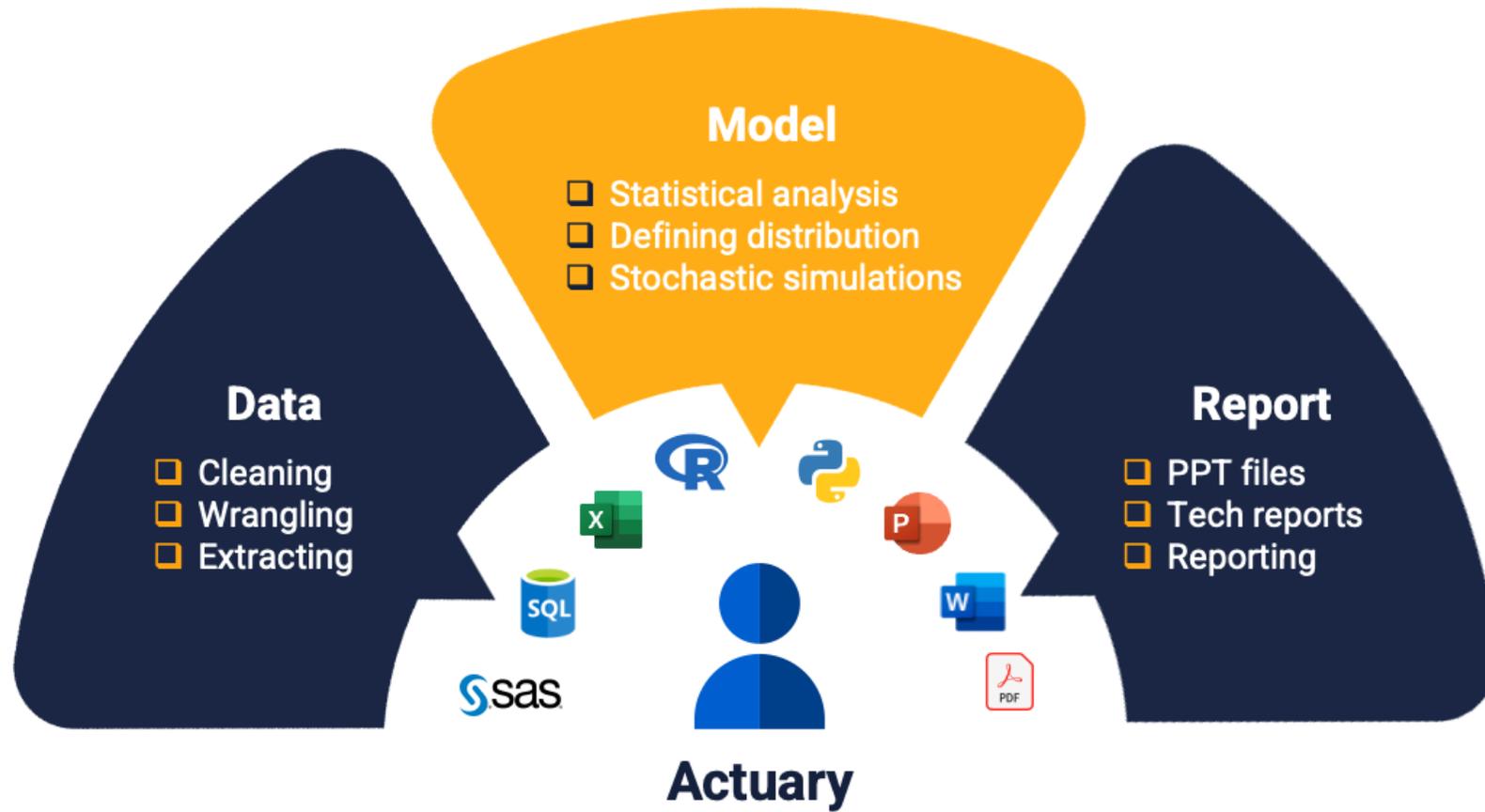
- DocMatrix** (Nouveau)  
Analysez vos documents avec l'IA
- Comparaison de deux versions**  
Comparez vos documents avec précision
- CR de Réunion**  
Générez des comptes-rendus automatiquement
- ChatDoc**  
Interagissez avec vos documents
- Création de présentations**  
Générez des présentations impactantes
- Traduction**  
Traduisez vos documents

### 📖 Tutoriels

The 'Tutoriels' section contains three large colored boxes with icons: a blue box with a DocMatrix icon, a green box with a document icon, and a purple box with a translation icon.

# From LLM to virtual coworker

# Actuary's atomic tasks



## Gen AI Vs Classical solutions

### Gen AI models

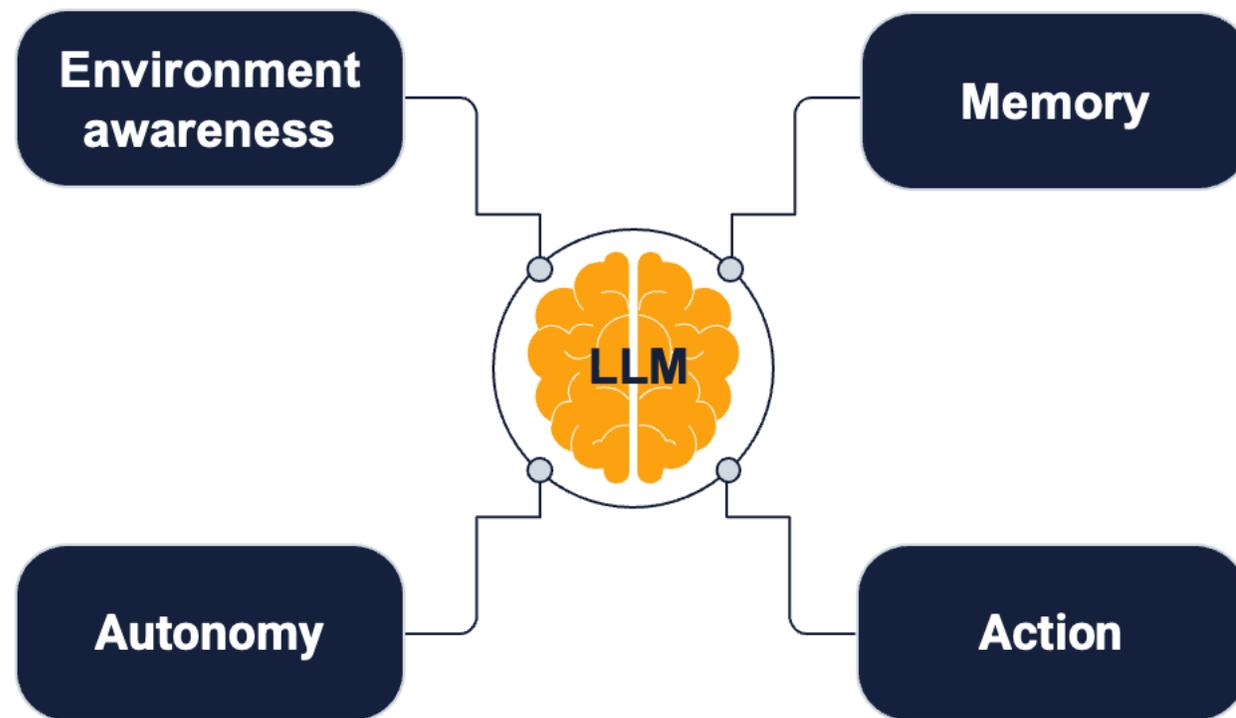
- ↑ Good at understanding the context
- ↑ Can use tools and can reason
- ↓ Can't scale and can't handle large data

### Specialized IT tools

- ↑ Tailor made for predefined tasks
- ↑ Perfect for scaling and handles large data.
- ↓ no notion of context awareness



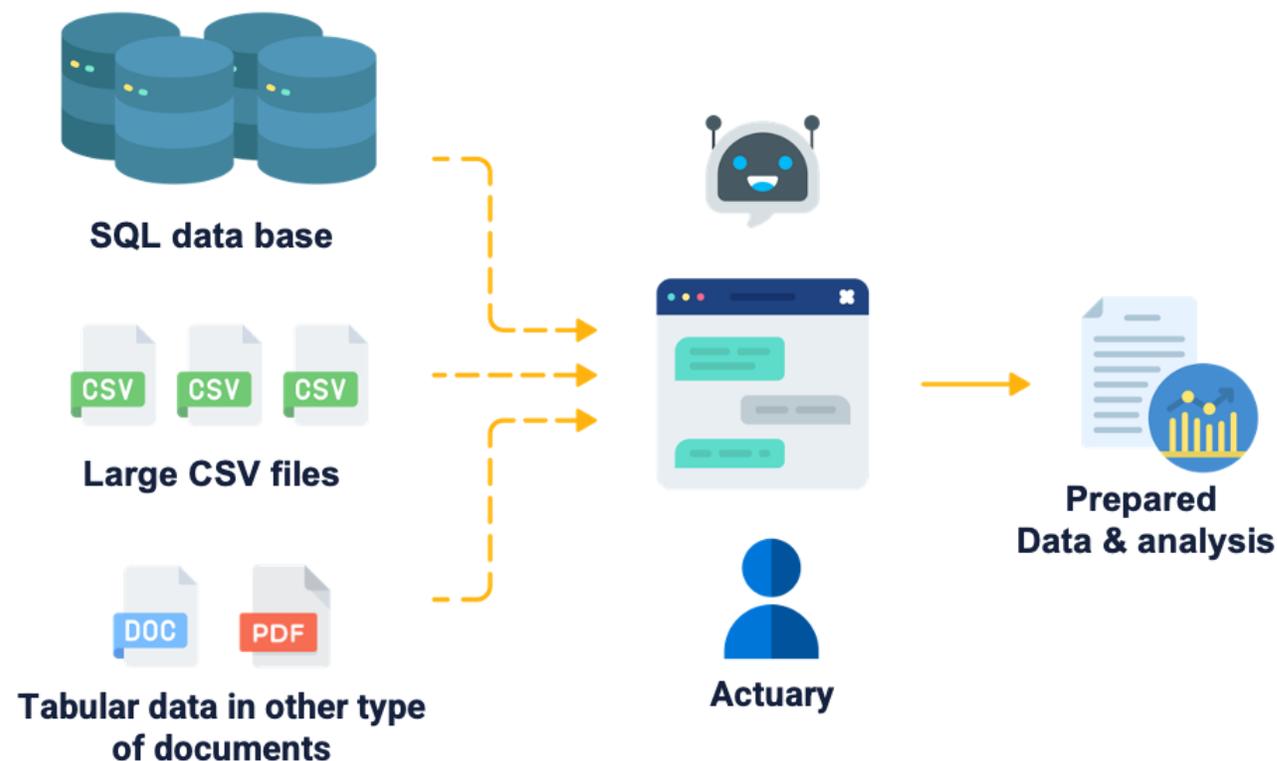
# What 's an AI Agent?



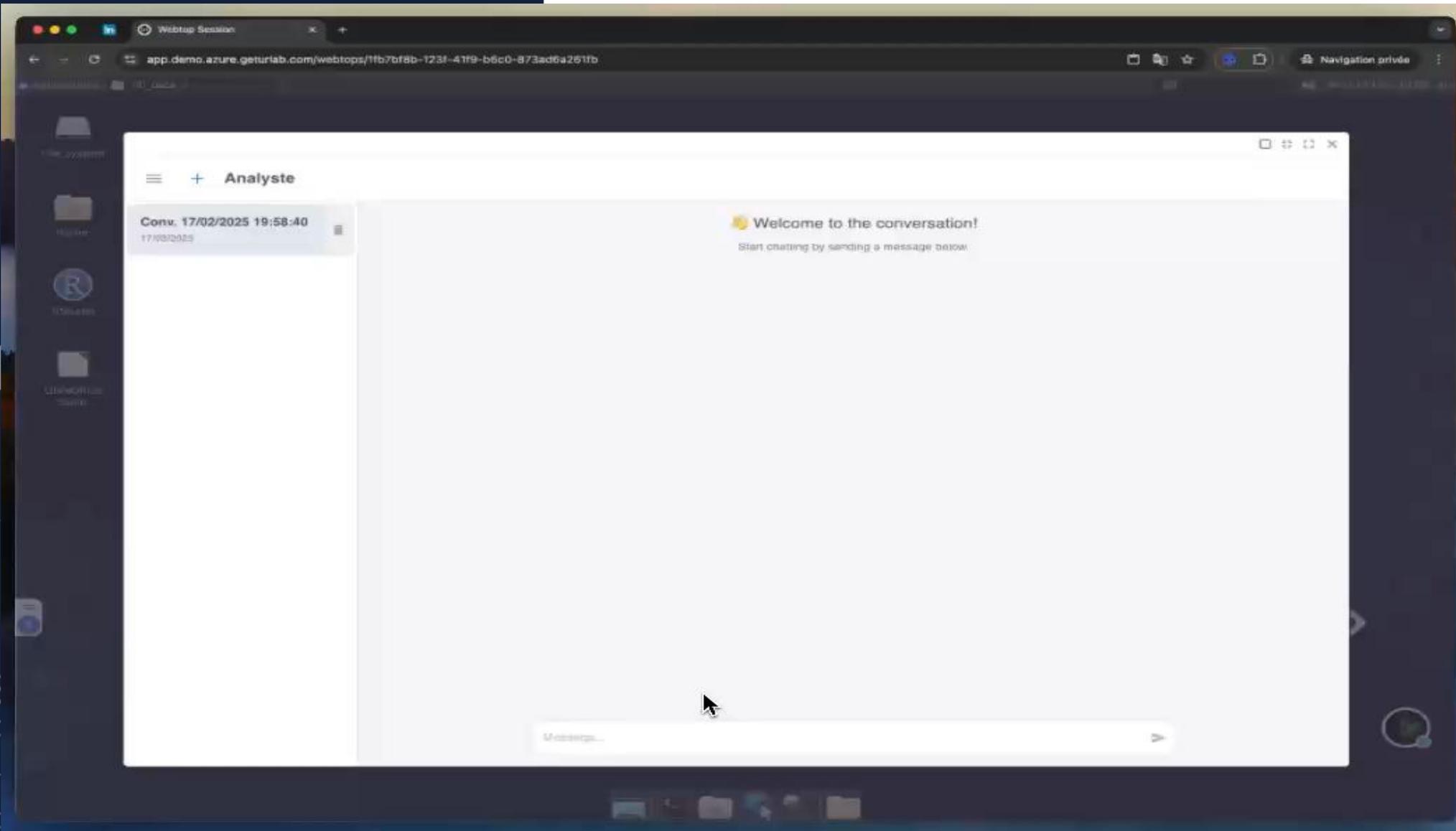


# IO Agent Intelligent Data Processing

## Analyze data in Seconds



**Query complex databases, large CSV files and extract data from PDF simply by chatting**





**Kezhan SHI**  
Deputy Director of  
Analytics - Actuarial  
and Data Science



**Yapei LI**  
Senior Manager  
Nexialog Consulting



**Reda JARIR**  
CEO et co-fondateur  
d'URLab





**Donnez votre avis sur les Journées IARD 2025**



**Et à bientôt au sein des groupes de travail de la commission IARD !**