

Retour d'expérience sur le déploiement d'IA sécurisées en local

Eric SanJuan

eric.sanjuan@univ-avignon.fr ; admin@adesis.fr



Laboratoire des Sciences
Juridiques, Politique, Économiques et de Gestion



MONTPELLIER INTERDISCIPLINARY CENTER
ON SUSTAINABLE AGRI-FOOD SYSTEMS
SOCIAL AND NUTRITIONAL SCIENCES



Sommaire

Quatre parties

- Présentation de l'installation : RAG pour ETL
- Application 1 : Recherche bibliographique: CLEF Simple Text
- Application 2 : Jurisprudence (cours d'appel)
- Application 3 : Traitement de tickets de caisse pour enquête

Sommaire

Quatre parties

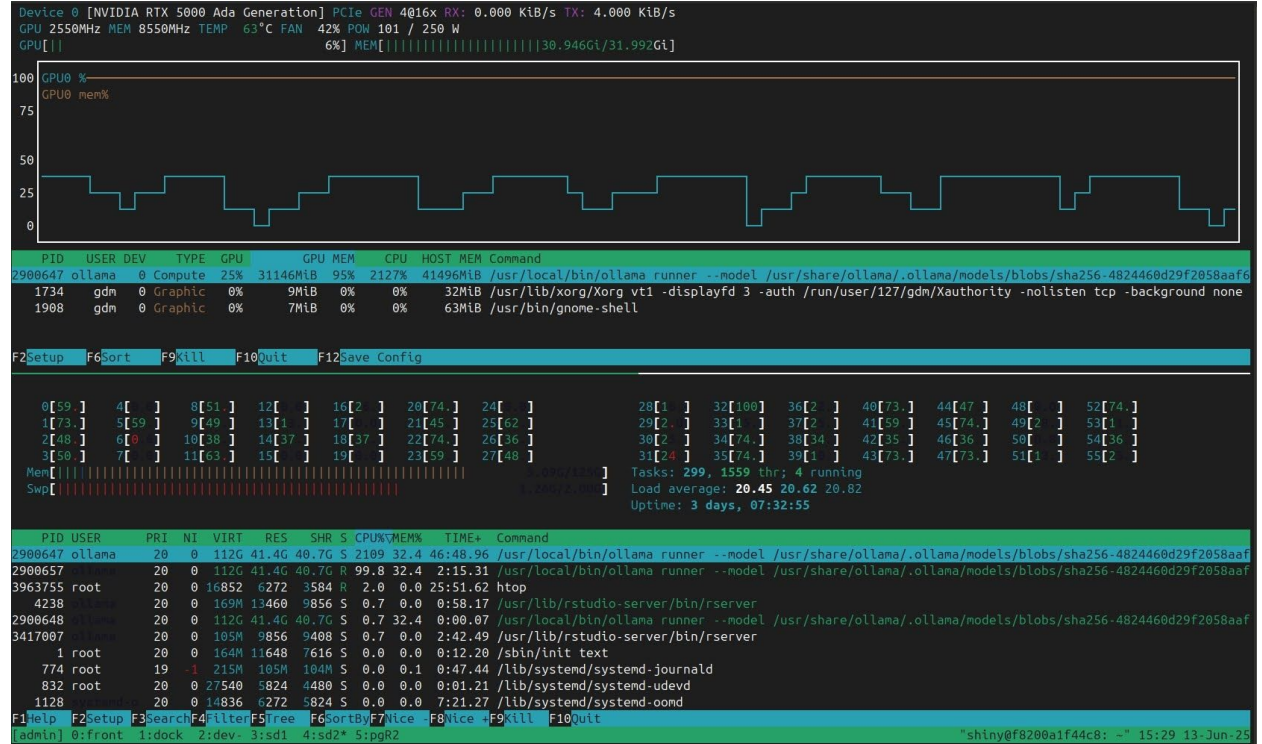
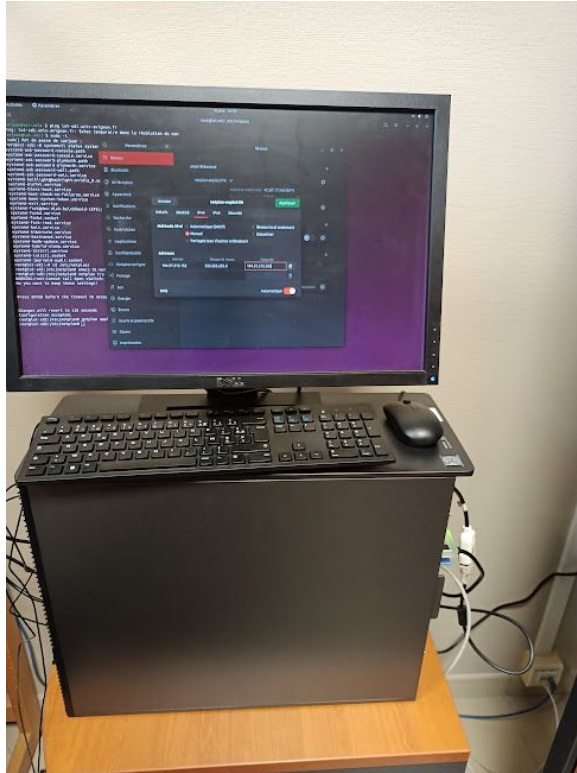
- **Présentation de l'installation : RAG pour ETL**
- Application 1 : Recherche bibliographique: CLEF Simple Text
- Application 2 : Jurisprudence (cours d'appel)
- Application 3 : Traitement de tickets de caisse pour enquête

Installation : RAG interne pour ETL

Fournir un contexte aussi précis que possible au LLM qui va produire un résumé simplifié

- contexte exhaustif de taille aussi réduite que possible destinés à être soumis à des LLMs locaux non connectés à un moteur de recherche.
- contrôle des hallucinations pour ETL: les résultats ne doivent pas introduire d'élément nouveau mais Extraire/Transformer/"Loader" les données

Installation : stations de travail vs serveurs



Installation: Service Ollama avec WebUI

Open WebUI - PostgreSQL Admin Permission... | Open WebUI

gemma3:27b v +

Rechercher un modèle

- llama3.2:3b-instruct-q8_0 3.2B
- llama3.2:latest 3.2B
- llama3.3:latest 70.6B
- llama4:scout 108.6B
- llava:latest 7B
- minicpm-v:latest 7.6B
- mistral-small3.1:latest 24.0B

Chat éphémère

In postgresql how do I grant admin role to an existing user?

ly admin) to an existing user in PostgreSQL using the **ALTER ROLE** or **GRANT**
SER is a powerful privilege and should be done with extreme caution. It allows
ider if a less privileged role is sufficient for the user's needs.

s and important considerations:

mplicity)

od.

ALTER ROLE username SUPERUSER;

Replace **username** with the actual username you want to grant admin privileges to.

Example:

```
sql
1 ALTER ROLE myuser SUPERUSER;
```

2. Using **GRANT ROLE**

This method is more explicit, and a good practice if you're managing roles formally. You need to be a superuser or the role

+ Envoyer un message

Les LLM peuvent faire des erreurs. Vérifiez les informations importantes.

Installation: Rstudio avec PostGreSQL

The screenshot displays the RStudio IDE interface. The main editor window contains R code for reading a TSV file, setting up a Rollama server, and writing the results. The Environment pane on the right shows the database connection and loaded data. The Console pane at the bottom shows the session restoration message and the database connection command.

```
1 library(readr)
2 DWP_pool <- read_delim("DWP_pool.tsv", delim = "\t", escape_double = FALSE, trim_ws = TRUE)
3 DWP_pool_s=DWP_pool[301:600,]
4 library(rollama)
5 options(rollama_server="http://iut-sd2.univ-avignon.fr:11434")
6 options(rollama_model="mistral:latest")
7 q_os <- make_query(
8   text = DWP_pool_s$title,
9   template = "{prompt}\n{text}",
10  prompt = "Ecrit le résumé d'une note encyclopédique ayant comme entrée :",
11  system = "Tu es un contributeur du chapitre en français du WikiPedia."
12 )
13 resp_mistral=query(q_os,screen=FALSE,model="mistral:latest",output = "text")
14
15 folder="DWP02_pool/"
16
17 DWP_M=DWP_pool_s
18 DWP_M$text=resp_mistral
19 f="DWP_pool_M.tsv"
20 write.table(DWP_M,file = paste(folder,f,sep = ""),sep = "\t",row.names = FALSE)
21
```

Environment

aminer - sanjuan

PostgreSQL

- aminer
 - pg_toast_temp_13
 - public
 - citations
 - dblp
 - doc : json
 - id : text
 - dblp_v
 - id : text
 - title_v : vector
 - abstract_v : vector
 - fos

Files

Home > embeddings

Name	Size	Modified
..		
embeddings.Rproj	205 B	Oct 8, 2025, 8:14 AM
rol_gen_Q.R	731 B	Sep 28, 2025, 11:48 PM
rol_gen_L.R	739 B	Sep 28, 2025, 11:47 PM
rol_gen_M.R	745 B	Sep 28, 2025, 11:45 PM
rol_gen_G.R	733 B	Sep 28, 2025, 11:45 PM

Console

R 4.5.1 · ~/embeddings/

Session restored from your saved work on 2025-Sep-29 12:16:40 UTC (8 days ago)

```
> library(DBI)
> con <- dbConnect(odbc::odbc(), .connection_string = "Driver={PostgreSQL Unicode};database=aminer", time
out = 10)
>
```

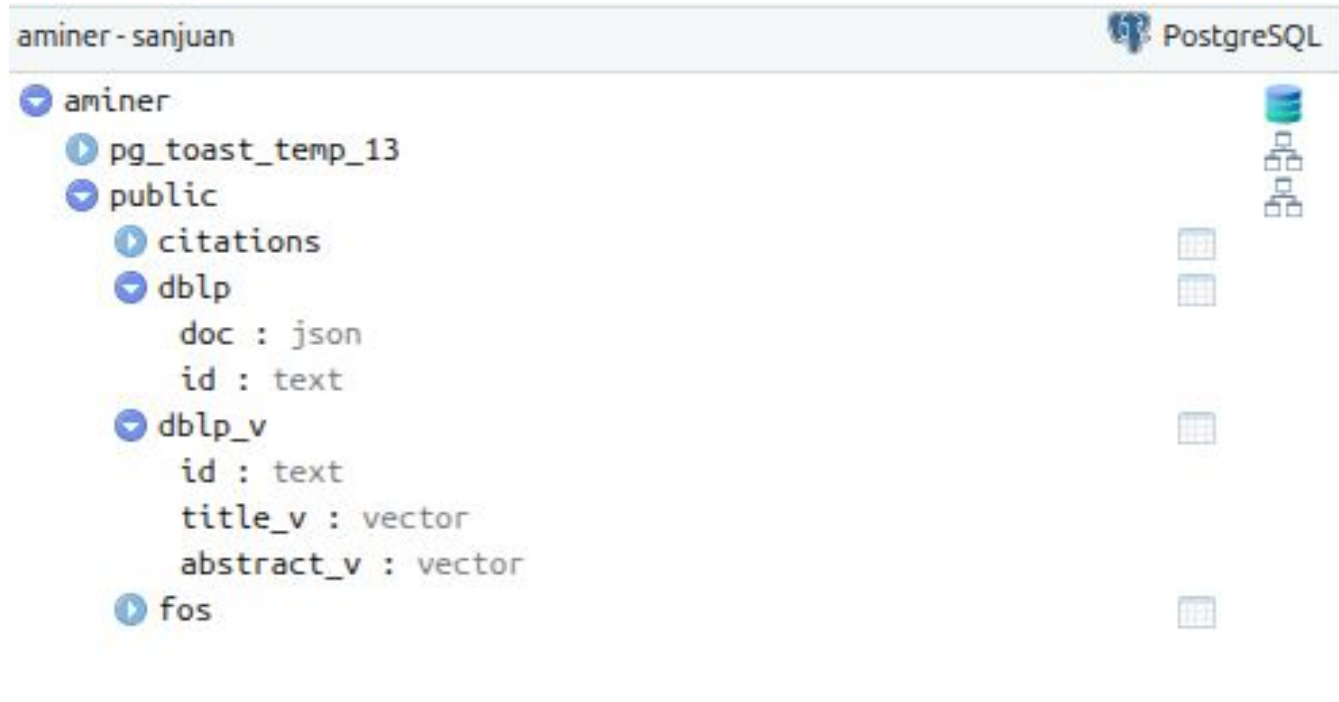
Installation : Rollama

```
library(readr)
DWP_pool <- read_delim("DWP_pool.tsv", delim = "\t", escape_double = FALSE, trim_ws = TRUE)
DWP_pool_s=DWP_pool[301:600,]
library(rollama)
options(rollama_server="http://iut-sd2.univ-avignon.fr:11434")
options(rollama_model="mistral:latest")
q_os <- make_query(
  text = DWP_pool_s$title,
  template = "{prompt}\n{text}",
  prompt = "Ecrit le résumé d'une note encyclopédique ayant comme entrée :",
  system = "Tu es un contributeur du chapitre en français du WikiPedia."
)
resp_mistral=query(q_os,screen=FALSE,model="mistral:latest",output = "text")

folder="DWP02_pool/"

DWP_M=DWP_pool_s
DWP_M$text=resp_mistral
f="DWP_pool_M.tsv"
write.table(DWP_M,file = paste(folder,f,sep = ""),sep = "\t",row.names = FALSE)
```

Installation : PostGreSQL + JSON + pgVector



Installation : pgVector

				(Displaying up to 1,000
	id	title_v	abstract_v	
1	1892892095	[0.06941398,0.03600585,0.083829604,-0.0063420...	[0.019542724,0.0021982114,0.07585799,0.028652...	
2	1893117846	[-0.05766046,-0.040311746,-0.00029231416,-0.04...	[-0.0112954285,-0.06377826,-0.04502894,-0.0625...	
3	18933200981	[0.0033271536,0.07457372,-0.057642035,0.03408...	[-0.036597297,0.046115108,-0.03933748,0.004331...	
4	1893302682	[-0.049472854,-0.049458172,-0.06946542,-0.0687...	[-0.02156585,-0.036530618,-0.05595649,-0.03007...	
5	1893200836	[-0.037849177,-0.005250847,0.0046259635,-0.031...	[-0.015196021,-0.03285868,-0.047562376,-0.1069...	
6	1893200780	[-0.03978418,0.012520332,0.020201985,-0.009428...	[0.008341246,0.05765245,-0.01494705,-0.0093828...	
7	1893034632	[-0.01719861,-0.03396671,0.012941862,0.0351529...	[0.040404078,-0.04188154,-0.027470078,0.029658...	
8	1893196298	[-0.0099693835,0.021934746,0.052379824,0.0494...	[0.009155645,-0.007277598,0.12302289,0.042672...	
9	1893120560	[-0.05873765,0.061107934,-0.013298873,0.070725...	[0.012462661,0.025829172,0.03779955,0.0717207...	
10	1892990328	[-0.08095625,0.087146394,0.010515917,0.044099...	[-0.038409013,-0.014408833,-0.088665776,-0.054...	
11	1893378478	[-0.044500217,-0.0066363774,-0.017724015,-0.06...	[-0.07075117,0.009923035,-0.015972475,-0.06325...	
12	1893174176	[0.103583544,-0.07011629,-0.00463379,-0.04536,...	[0.062597394,-0.04566114,-0.05087917,-0.015451...	
13	1892974185	[-0.015287041,-0.00869078,0.021901986,0.019920...	[-0.043091945,0.017510226,-0.13652222,-0.04692...	
14	1893686228	[0.039673354,-0.025408734,-0.068016596,-0.1254...	[0.044675928,0.031360257,-0.07810905,-0.108542...	
Showing 1 to 14 of 1,000 entries 3 total columns				

Installation : PostGreSQL avec type JSON

(Displaying up to 1,000 records)

	doc	id
1	<pre>{"_index":"dblp1","_type":"_doc","_id":"2985886803","_score":1,"_source":{"id":2985886803,"authors":[{"name":"Cifuentes Cristina","id":2985886803,"org":"Polytechnic University of Valencia"}]}}</pre>	
2	<pre>{"_index":"dblp1","_type":"_doc","_id":"2985896746","_score":1,"_source":{"id":2985896746,"authors":[{"name":"Anirudh Paranjothi","org":"Sri Lanka Institute of Advanced Technological Education"}]}}</pre>	
3	<pre>{"_index":"dblp1","_type":"_doc","_id":"2985911758","_score":1,"_source":{"id":2985911758,"authors":[{"name":"Marko Luoma","id":2806897,"org":"University of Jyväskylä"}]}}</pre>	
4	<pre>{"_index":"dblp1","_type":"_doc","_id":"2985922145","_score":1,"_source":{"id":2985922145,"authors":[{"name":"Vicente Casares","org":"Polytechnic University of Valencia"}]}}</pre>	
5	<pre>{"_index":"dblp1","_type":"_doc","_id":"2985926846","_score":1,"_source":{"id":2985926846,"authors":[{"name":"David Beech","id":2082641,"org":"University of York"}]}}</pre>	
6	<pre>{"_index":"dblp1","_type":"_doc","_id":"2985926990","_score":1,"_source":{"id":2985926990,"authors":[{"name":"Mihai Bujanca","org":"Advar University of Medicine and Pharmacy"}]}}</pre>	
7	<pre>{"_index":"dblp1","_type":"_doc","_id":"2990255558","_score":1,"_source":{"id":2990255558,"authors":[{"name":"Mohammed Hazim Alkawaz","id":2990255558,"org":"University of Al-Qadisiyah"}]}}</pre>	
8	<pre>{"_index":"dblp1","_type":"_doc","_id":"2985948189","_score":1,"_source":{"id":2985948189,"authors":[{"name":"Zhuotao Liu","org":"Universitat de València"}]}}</pre>	
9	<pre>{"_index":"dblp1","_type":"_doc","_id":"2985949526","_score":1,"_source":{"id":2985949526,"authors":[{"name":"Nelson Diaz","org":"Universitat de València"}]}}</pre>	
10	<pre>{"_index":"dblp1","_type":"_doc","_id":"2985950007","_score":1,"_source":{"id":2985950007,"authors":[{"name":"Diandian Gu","org":"Peking University"}]}}</pre>	
11	<pre>{"_index":"dblp1","_type":"_doc","_id":"2985951359","_score":1,"_source":{"id":2985951359,"authors":[{"name":"Nikolaos Sarafianos","id":2082641,"org":"University of York"}]}}</pre>	
12	<pre>{"_index":"dblp1","_type":"_doc","_id":"2985955412","_score":1,"_source":{"id":2985955412,"authors":[{"name":"Xiangfu Zhong","org":"Depa University of Technology"}]}}</pre>	
13	<pre>{"_index":"dblp1","_type":"_doc","_id":"2990148539","_score":1,"_source":{"id":2990148539,"authors":[{"name":"Guilherme Kano","org":"Universidade Federal do Rio de Janeiro"}]}}</pre>	

Showing 1 to 13 of 1,000 entries | 2 total columns

Sommaire

Quatre parties

- Présentation de l'installation : RAG pour ETL
- **Application 1 : Recherche bibliographique: CLEF Simple Text**
- Application 2 : Jurisprudence (cours d'appel)
- Application 3 : Traitement de tickets de caisse pour enquête

App1: Leçons apprises des épisodes précédents

CLEF SimpleText task 1 2022 - 2023

- Documents : DBLP < 2019 titres et résumés + Citations ACM
- Requêtes : Articles de presse grand public sur des questions sociétales ou des technologies liées à l'informatique

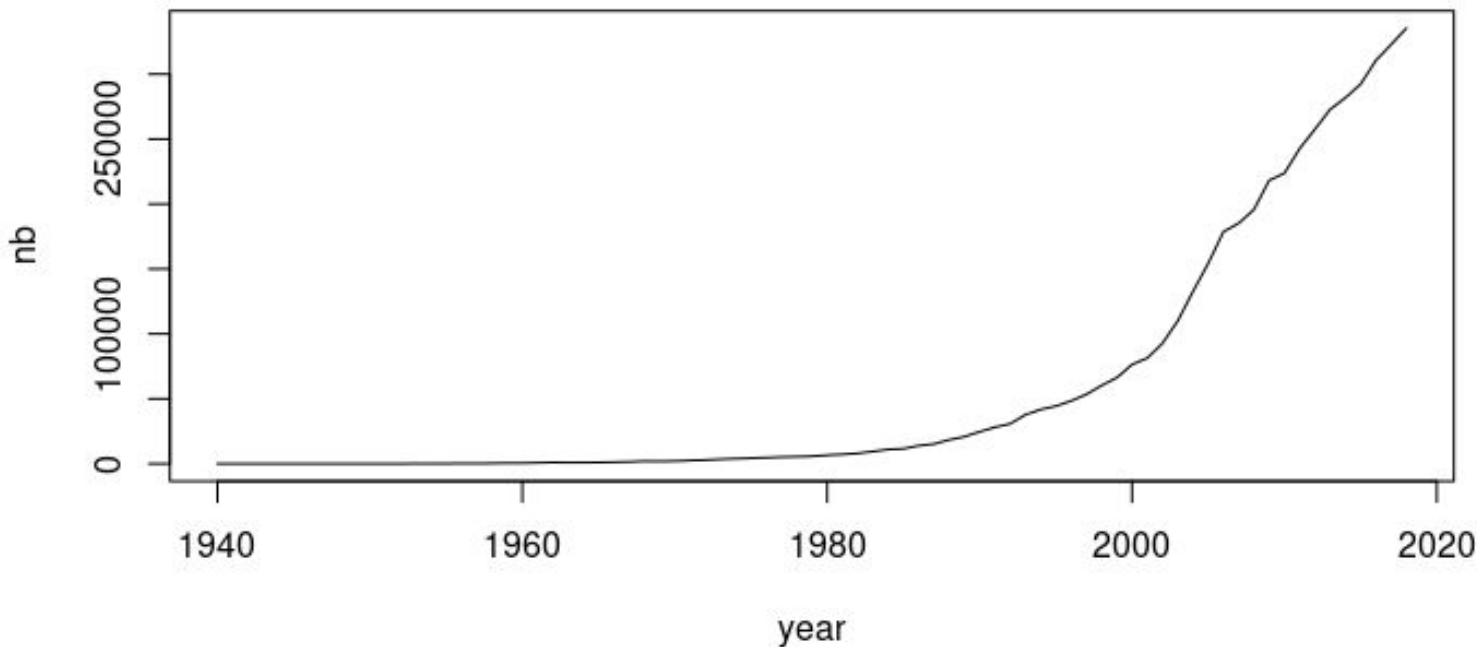
Quelques résultats inattendus

- Efficacité des plongements vectoriels denses à froid pré-calculés pour tous les titres et résumés
- Performance des bases de données relationnelles vis à vis d'approches NoSQL

App 1: Ressources Simple Text Wayback (Corpus)

Notices DBLP avant 2019 + Résumés + Citations ACM

+ MS Academic Graph 2019 + Sentence Transformers all-MiniLM-L6-v2



App 1: Ressources Wayback (Requêtes)

Thèmes extraits d'articles du Guardian publiés après 2019 :

1. How AI systems, especially virtual assistants, can perpetuate gender stereotypes?
2. Concerns related to the handling of sensitive information by voice assistants.
3. How children interact with voice assistants and the design of child-friendly interfaces.
4. Use of AI to improve success rates and speed in the pharmaceutical research field.
5. Application of machine learning algorithms to predict genomic features, functions, and the outcomes of gene-editing interventions like Crispr.
6. Ethical considerations, governance frameworks, and policies for the responsible development and deployment of AI technologies.
7. Use of NLP techniques to detect and analyze misinformation in textual content on social media platforms.
8. Understand the cryptographic underpinnings of blockchain technology, which is the foundation of Bitcoin and other cryptocurrencies
9. Computer science techniques to analyze spatial data and imagery, particularly for reconstructing crime scenes or human rights violation incidents
10. Study of robotic technologies and automated systems that are replacing human labor in various sectors.

App 1 : Ressources Simple WayBack (Références)

query	nb
G01.C1	127
G02.C1	118
G03.C1	110
G04.C1	157
G05.C1	93
G06.C1	117
G07.C1	110
G08.C1	134
G09.C1	158
G10.C1	158

rel	nb
0	672
2	311
1	299

App 1 : Evaluation systèmes “officielle” 2024

Run	MRR	Precision		NDCG		Bpref	MAP
		10	20	10	20		
baseline_bool	0.6500	0.4100	0.2550	0.3167	0.2328	0.1216	0.0694
baseline_elasti	0.8192	0.5200	0.4050	0.4434	0.3623	0.3968	0.2000
baseline_meili	0.5124	0.1900	0.1000	0.1254	0.0869	0.0779	0.0209
baseline_vir_abstract	0.9500	0.8200	0.6150	0.6701	0.5543	0.5533	0.2996
baseline_vir_title	0.9333	0.8600	0.5650	0.7184	0.5415	0.5196	0.2633

Table 1: Evaluation of SimpleText Task 1 (Test qrels: Guardian).

App 1 : Evaluation systèmes “étendue” 2024

Run	MRR	Precision		NDCG		Bpref	MAP
		10	20	10	20		
baseline_bool	0.6500	0.4100	0.2550	0.2987	0.2264	0.1026	0.0590
baseline_elastic	0.8192	0.5200	0.4050	0.4065	0.3434	0.3534	0.1743
baseline_meili	0.5124	0.1900	0.1000	0.1161	0.0833	0.0689	0.0187
baseline_vir_abstract	0.9500	0.8200	0.6200	0.6276	0.5394	0.5164	0.2730
baseline_vir_title	0.9333	0.8700	0.5950	0.6806	0.5527	0.5244	0.2646

Table 1: Evaluation of SimpleText Task 1 (Test qrels: Guardian).

Back to SQL

```
create table st1_qrels_2024_08 as
(select topic, query,
 dblp_knn_title(qdblp_v_title(doc),10,-0.8) as doc, rel
 from st1_qrels_2024 where rel=2)
union (select * from st1_qrels_2024 where rel<2);

CREATE OR REPLACE FUNCTION public.dblp_knn_title(v_query vector(384), nb numeric)
RETURNS TABLE(id text, ip float, doc json)
LANGUAGE plpgsql
AS $$
BEGIN
    SET LOCAL ivfflat.probes = 65;
    SET LOCAL enable_seqscan = off;
    SET LOCAL min_parallel_table_scan_size = 1;
    SET LOCAL parallel_setup_cost = 1;
    RETURN QUERY
    SELECT J.id, (J.title_v <=> v_query) AS ip, D.doc AS doc
    FROM dblp_v AS J, dblp AS D
    WHERE D.id = J.id
    ORDER BY ip LIMIT nb;
END;
$$
;
```

App 1 : Peut-on automatiser les annotations ?

prefix = "Here is a societal question and a scientific paper in computer science. Please, do not recommend papers that are off topic. I do not have time to read them all.",

text = paste("To address the question: ", dfr\$query_text, ", should I read the paper which title is:", dfr\$title, "and which abstract is :", dfr\$abstract),

prompt = "Answer only returning a relevance score 0, 1 or 2. 0: Not really relevant, 1: relevant, 2: very relevant",

template = "{prefix}{text}\n{prompt}",

system = "You are a journalist writing about a tech topic that raises societal questions. You are looking for scientific publications that could feed your paper for a large audience."

App 1 : Evaluation des LLM locaux (Rollama)

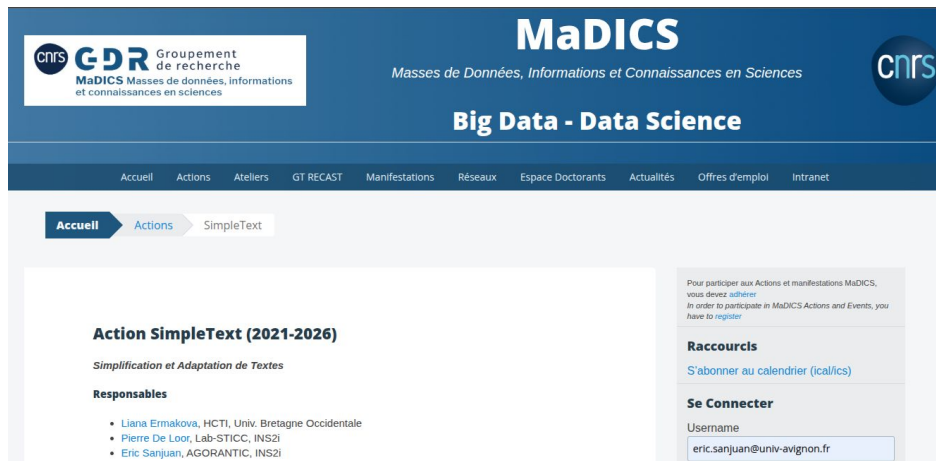
Model	tau	P-value	Accuracy (3)	Accuracy (2)
Qwen	-1,25 %	63,29 %	23,24 %	48,44 %
Qwq	30,00 %	***	32,17 %	52,73 %
Gemma3 :small	33,26 %	***	37,00 %	58,45 %
gemma3:12b	31,95 %	***	38,29 %	59,59 %
Phi4	41,54 %	***	45,16 %	62,79 %
Llama 4	40,04 %	***	52,11 %	69,58 %

App 1 : ressources et système de démonstration

<https://simpletext-project.com/>

<https://www.madics.fr/actions/simpletext/>

<https://wayback.simpletext-project.com/>



The screenshot shows the MaDICS website header with the CNRS logo and the text "MaDICS Masses de données, informations et connaissances en sciences". Below the header is a navigation bar with links: Accueil, Actions, Ateliers, GT RECAST, Manifestations, Réseaux, Espace Doctorants, Actualités, Offres d'emploi, Intranet. The main content area is titled "Action SimpleText (2021-2026)" and includes a description "Simplification et Adaptation de Textes". It also lists the responsible individuals: Liana Ermakova, HCTI, Univ. Bretagne Occidentale; Pierre De Loor, Lab-STICC, INS2I; and Eric Sanjuan, AGORANTIC, INS2I. On the right side of the page, there is a section for "Raccourcis" (Shortcuts) with a link to "S'abonner au calendrier (icalics)" and a "Se connecter" (Log in) section with a username field containing "eric.sanjuan@univ-avignon.fr".



[SimpleText Project](#)

Way Back data

TURN IT SIMPLE

- [Working Notes](#)
- [CLEF 2025 Slides](#)
- [Demo system](#)
- [Docker Image \(30 Go\)](#)
- [Database Dump](#)
- [task1 Topics related content](#)

contact: wayback@simpletext-project.com

Sommaire

Quatre parties

- Présentation de l'installation : RAG pour ETL
- Application 1 : Recherche bibliographique: CLEF Simple Text
- **Application 2 : Jurisprudence (cours d'appel)**
- Application 3 : Traitement de tickets de caisse pour enquête

App 2 : RI neuronale juridique

Open data décisions de justice

JSON extrait des PDFs Appel et Cassation

Résumés par extraction

Représentation dense unique

Larges Modèles de Langue

Segmentation en paragraphes

Mini Modèles de Langue

Recherche de passages

Recherche de décisions

Réordonnancement

Modèle relationnel

Jointure et agrégation

Suivi d'affaires (case retrieval)

App 2 : PoC

Open data décisions de justice

PostgreSQL + pg_vector + curl

Importation JSON

Opérateurs NoSQL

Segmentation en paragraphes

Vecteur dense MiniLM-L12-v2-mmarcoFR

Réduction k-means (ivvflat)

Recherche de décisions

Produit scalaire requête x passages

Agrégation

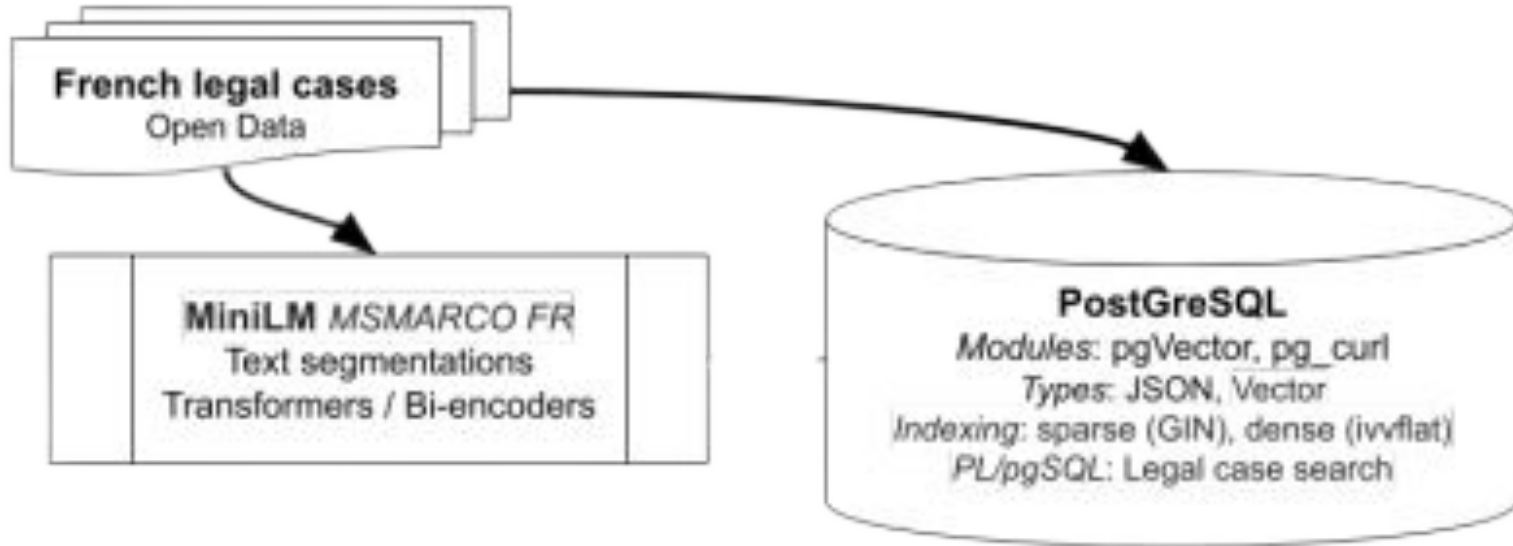
Recherche de passages

Similarité avec la requête

Recherche de Jurisprudence

Contexte instance

App 2 : Expérimentation



App 2 : Modèle relationnel

Pas que relationnel

Documents

- Id
- JSON

Extraits

- Id
- Id_doc
- position
- texte

Vecteurs

- id_extrait
- Modèle
- vecteur

Catégories

- Id_extrait
- catégorie

```
SELECT
J.id AS id_dec,
R.line AS line,
R.score AS score,
J.json->'metadata'->>'date' AS date,
J.json->'metadata'->>'short_title' AS title,
J.json->'raw_text'->>(line-1) AS passage
FROM
(SELECT id_dec, sum(ip) as score, line
FROM lib_knn(getemb(%s),%s)
GROUP BY id_dec, line
ORDER BY score) AS R,
judilibre_json AS J
WHERE R.id_dec=J.id;
```

App 2 : Exemple de fonction

```
-- ### knn judilibre vectors from vector
CREATE OR REPLACE FUNCTION lib_knn(v_query vector(384), nb numeric)
    RETURNS
    TABLE(id_dec character varying(100), chunk int, ip float, passage text, line int)
    LANGUAGE plpgsql
    AS $$
    BEGIN
        SET LOCAL ivfflat.probes = 65;
        SET LOCAL enable_seqscan = off;
        SET LOCAL min_parallel_table_scan_size = 1;
        SET LOCAL parallel_setup_cost = 1;
        RETURN QUERY
        SELECT J.id_dec, J.chunk, (J.embedding <#> v_query) AS ip, J.passage, J.line
        FROM judilibre_v AS J
        ORDER BY ip LIMIT nb;
    END;
    $$
    ;
```

App 2 : Évaluation

Systèmes :

- **0 [CLS]** biencoder-all-MiniLM-L12-v2-mmarcoFR (<https://huggingface.co/antoinelouis>)
- **1 Booléen** avec index généralisé (PostgreSQL)
- **2 mean pooling** biencoder-mMiniLMv2-L12-mmarcoFR (<https://huggingface.co/antoinelouis>)

Site: <https://guacamole.univ-avignon.fr/jplab/>

API:

https://guacamole.univ-avignon.fr/jpvir_test?corpus=0&phrase=contestation%20amende%20pour%20stationnement&length=100

50 requêtes moteur de recherche Juri'Predis

- CONTRAVENTION ET COMPETENCE POLICE MUNICIPALE
- peut-on couper l'électricité chez un occupant sans droit ni titre
- coupure d'électricité chez un occupant sans droit ni titre
- doutes raisonnables volonté quitter état membre visa mariage
- doutes raisonnables volonté quitter territoire état membre visa
- demande de libération du logement de fonction pour loger le successeur

App 2 : Experimentation

<https://guacamole.univ-avignon.fr/jplab/>

Jurisprudences Personnalisées



0



contestation contravention pour stationnement dangereux

5



Rechercher

Id: JURITEXT63c10971bf9fd47c90a13c0d

Titre: CA Paris, 12 janv. 2023, RG n° 21/19806

Date: 2023-01-12

Passage: Les stipulations du contrat de cession d'actions prévoient donc expressément que pour être valide, une réclamation doit préciser le fondement, l'acte ou l'événement sur lequel elle est fondée, indiquer le montant ou une estimation du montant réclamé et comporter les documents qui permettent d'établir l'existence de la perte.

Ligne: 167

Score: -28.410415649414062



Jurisprudences Personnalisées



2



contestation contravention pour stationnement dangereux

5



Rechercher

Id: JURITEXT64019dd4546e3305deed5bca

Titre: CA Aix-en-Provence, 2 mars 2023, RG n° 21/18317

Date: 2023-03-02

Passage: Le dommage imminent s'entend du dommage qui n'est pas encore réalisé mais qui se produira sûrement si la situation présente doit perdurer. Il s'ensuit que, pour que la mesure sollicitée soit prononcée, il doit nécessairement être constaté, avec l'évidence qui s'impose à la juridiction des référés, l'imminence d'un dommage, d'un préjudice ou la méconnaissance d'un droit, sur le point de se réaliser et dont la survenance et la réalité sont certaines ; un dommage purement éventuel ne saurait être retenu pour fonder l'intervention du juge des référés.

Ligne: 83

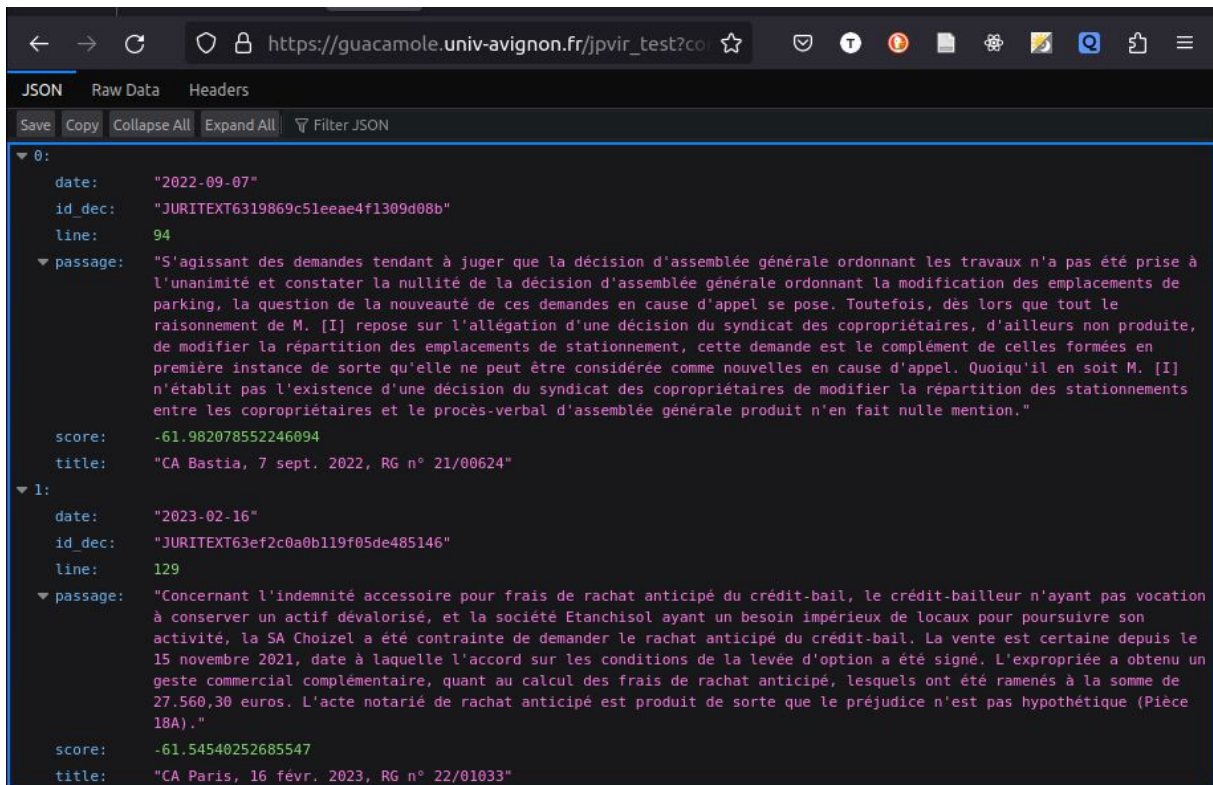
Score: -0.22006282210350037



App 2 : Experimentation

https://guacamole.univ-avignon.fr/jpvir_test?

corpus=0&phrase=contestation%20amende%20pour%20stationnement&length=100



```
JSON  Raw Data  Headers
Save  Copy  Collapse All  Expand All  Filter JSON

▼ 0:
  date:      "2022-09-07"
  id_dec:    "JURITEXT6319869c51eeae4f1309d08b"
  line:      94
  passage:   "S'agissant des demandes tendant à juger que la décision d'assemblée générale ordonnant les travaux n'a pas été prise à l'unanimité et constater la nullité de la décision d'assemblée générale ordonnant la modification des emplacements de parking, la question de la nouveauté de ces demandes en cause d'appel se pose. Toutefois, dès lors que tout le raisonnement de M. [I] repose sur l'allégation d'une décision du syndicat des copropriétaires, d'ailleurs non produite, de modifier la répartition des emplacements de stationnement, cette demande est le complément de celles formées en première instance de sorte qu'elle ne peut être considérée comme nouvelles en cause d'appel. Quoiqu'il en soit M. [I] n'établit pas l'existence d'une décision du syndicat des copropriétaires de modifier la répartition des stationnements entre les copropriétaires et le procès-verbal d'assemblée générale produit n'en fait nulle mention."
  score:     -61.982078552246094
  title:     "CA Bastia, 7 sept. 2022, RG n° 21/00624"

▼ 1:
  date:      "2023-02-16"
  id_dec:    "JURITEXT63ef2c0a0b119f05de485146"
  line:      129
  passage:   "Concernant l'indemnité accessoire pour frais de rachat anticipé du crédit-bail, le crédit-bailleur n'ayant pas vocation à conserver un actif dévalorisé, et la société Etanchisol ayant un besoin impérieux de locaux pour poursuivre son activité, la SA Choizel a été contrainte de demander le rachat anticipé du crédit-bail. La vente est certaine depuis le 15 novembre 2021, date à laquelle l'accord sur les conditions de la levée d'option a été signé. L'expropriée a obtenu un geste commercial complémentaire, quant au calcul des frais de rachat anticipé, lesquels ont été ramenés à la somme de 27.560,30 euros. L'acte notarié de rachat anticipé est produit de sorte que le préjudice n'est pas hypothétique (Pièce 18A)."
```

```
score:     -61.54540252685547
title:     "CA Paris, 16 févr. 2023, RG n° 22/01033"
```

App 2 : Evaluation https://coria.asso-aria.org/2024/articles/position_18/

Table 1: Evaluation based on expert qrels . *** indicates highly significant differences ($p < 0.01$, t-test)

System	Description	MAP	P[5]	P[10]	Recip_Rank	IPRec[0]	IPRec[10]	NDCG
M	MeiliSearch Document	0.1344	0.3407	0.3259	0.5180	0.5654	0.3825	0.2456
B	Boolean Passage Search	0.0580	0.1556	0.1185	0.2509	0.2735	0.1840	0.1360
T	Sentence Transformer	0.3252***	0.4815	0.4667	0.7727	0.7948	0.6258	0.4675***
E	Sentence bi-Encoder	0.1214	0.2519	0.2444	0.4050	0.4662	0.3617	0.2476
TB	Transformer + Boolean	0.3099	0.4593	0.4481	0.7171	0.7636	0.6073	0.4647
EB	bi-Encoder + Boolean	0.1332	0.2667	0.2593	0.4174	0.4829	0.3770	0.2808
TE	Transformer + Encoder	0.2283	0.4593	0.4481				

Table 2: Evaluation based on keyword overlap

System	Description	MAP	P[5]	P[10]	Recip_Rank	IPRec[0]	IPRec[10]	NDCG
M	MeiliSearch Document	0.0027	0.0043	0.0036	0.7407	0.0241	0.0213	0.0043
T	Sentence Transformer	0.0530	0.0645	0.1064	0.7727	0.1887	0.1351	0.1115
E	Sentence bi-Encoder	0.0204	0.0511	0.0291	0.4050	0.1309	0.0523	0.0636
TE	Transformer + Encoder	0.2283	0.1021	0.0464	0.7171	0.2026	0.1124	0.1148

App 2 : Requêtes et intensions

clause de non concurrence agent commercial

- clause concurrence agent commercial (42)

caution disproportion manifeste 12000 15000 aucun patrimoine

- caution disproportion aucun patrimoine (8)

référé 521-3 écoulement des eaux et fuite canalisation

- eaux fuite canalisation (97)

avenant compromis de vente droit de rétractation

- compromis vente rétractation (16)

harcèlement moral critiques

- harcèlement moral critiques (56)

chevauchement deux arrêts de travail

- chevauchement arrêt travail (2)

conduite sous l'empire d'un état alcoolique procédure de contrôle et vices de procédure

- conduite alcool (97)

COMMISSION DISCIPLINE BACCALAUREAT SANCTION FRAUDE BAC

- commission BAC (5)

App 1 & 2 : ressources sur

<https://www.madics.fr/actions/simpletext/>

Contener image debian autonome (30 Go)

- *Logiciels* : PostGreSQL + pg_vector + services mini LM + R + Rstudio + Shiny + Rollama
- *Corpus*: DBLP (< 2019) + JudiLibre (< 2019)
- *Données*: SimpleTest task1 2024 + evaluations Master Gouvernance des Données avignon
- *A venir*: collaboration NHS Scotland

Apps 1 & 2: Conclusions et perspectives

- Les mini LMs ont permis l'intégration dans les SGBDR de systèmes RI efficaces.
- Les LLMs internes peinent à se substituer aux annotateurs humains et ... les externes non plus à un coût raisonnable.
- Les SGBDR permettent de gérer les documents comme de larges ensembles de vecteurs à l'échelle de l'open data juridique.
- Perspectives d'applications à la gestion de l'information au sein de la NHS Scotland.
- Comment approcher la double nature discrète/continue des mots (graphie/vecteur) ?

Sommaire

Quatre parties

- Présentation de l'installation : RAG pour ETL
- Application 1 : Recherche bibliographique: CLEF Simple Text
- Application 2 : Jurisprudence (cours d'appel)
- **Application 3 : Traitement de tickets de caisse pour enquête**

App 3 : Stage BUT 3 SD

INRAE Montpellier - MOISA

Cheick Habib Aidara

Pascaline Rollet

<https://sd2.iut-avignon.fr>

RAPPORT DE STAGE

Automatisation du processus d'extraction d'informations à
partir de tickets de caisse

AIDARA Cheick Habib

01/04/2025 - 31/07/2025

Tuteurs MoISA : Pascaline ROLLET - Sabrina EYMARD-DUVERNAY

Professeurs Référents : Eric SANJUAN

Formation : IUT d'Avignon - BUT3 Science des Données - VCOD

Établissement : INRAE Montpellier - UMR MoISA

App 3 : Données

- Photos confidentielles de tickets de caisse confidentiels
- Historique de précédentes catégorisations manuelles d'items extraits manuellement de tickets de caisse.

Tableau des Détails et Similarités Article et Prix par Scanner

Code_message	Article_Ticket_2e_scan	Article_Ticket_1ere_scan	Prix_2e_scan	Prix_1ere_scan	Articles_Ticket_original	Prix_Atendu	Similarite_Article_Scan1	Similarite_Prix_Scan2	Similarite_Article_Scan2	Similarite_Prix_Scan2
093L0028	SAC BRETEL UNIT HIPPO	Sac Caisse Bretel Unit Hippo	0,20 €	0,20 €	SAC CASSIE BRETEL UNIT HIPPO	0,20 €	0.8889	1	0.6154	1.0000
093L0028	VERMICELLES PANZI 500G	Vermicelles Panzani 500g	1,83 €	1,83 €	VERMICELLES PANZANI 500G	1,83 €	0.9583	1	0.9167	1.0000
093L0028	TUILES GOUT CREME D.	Tuiles gout Chrs Oignon	3,14 €	3,14 €	TUILES-GOUT CREME OIGNON GUTOS	3,14 €	0.7000	1	0.6000	1.0000
093L0028	MONSTER KETCHUP	Monster Huncketchup 85g	1,51 €	1,51 €	MONSTER MUNCH KETCHUP 85G	1,51 €	0.8800	1	0.6000	1.0000
093L0028	FARFALLE PANZI ST CELLO 500G	Farfalles Panzani St Cello 500g	1,41 €	1,41 €	FARFALLE PANZANI ST CELLO 500G	1,41 €	0.9333	1	0.9000	1.0000
093L0028	SACS POUILLIEN COUL 1350L	Sacs Poux Coules 13x25	2,70 €	2,70 €	SACS POUILLIEN COUL V 1350L	2,70 €	0.6154	1	0.8846	0.9890
093L0028	EAU LITE CITRON U 750ML	Liq Vaiselle Citron 750ml	1,40 €	1,40 €	LIQ VAISSELLE CITRON U 750ML	1,40 €	0.8889	1	0.6296	1.0000
093L0028	LT UHT 1,2L	Lait 1,2% G&L	12,56 €	12,56 €	LT UHT 1,2% NATIN LEG BLE G&L	12,56 €	0.3704	1	0.2963	0.9984
093L0028	DEUTS FRAYS	Beuf Frites Telle Cates	1,89 €	1,89 €	OEUTS FRAYS TT CALBRE X30 PPN	1,89 €	0.4333	1	0.3000	1.0000
093L0028	EMMENT PAST 30X200S	Emen Frites Past 200g	1,77 €	1,77 €	EMMENT. RAPE PAST.30% 200G PSM	1,77 €	0.5185	1	0.5026	1.0000



App 3 : Processus

- Extraction d'informations des tickets : articles, prix, date, lieu, et montant total figurant sur les tickets.
- Classification automatique : attribuer automatiquement chaque aliment ou boisson à un groupe alimentaire
- Spécification des articles NON ALIMENTAIRE pour calculer indicateurs de qualité nutritionnelle.
- Contrôle qualité : Identification et annotation manuelles des cas tangents

App 3 : Shiny <https://sd2.iut-avignon.fr/dataviz> avec Llama4

The screenshot displays a Shiny application interface with two main panels. The left panel, titled "Tickets à traiter", lists 11 tickets. Ticket #2, labeled "Traitement...", is selected and highlighted with a black circle. Above the list are buttons for "Tout sélectionner" and "Traiter la sélection". The right panel, titled "Prévisualisation du ticket", shows a preview of a receipt. Below the preview, the section "Extraction des articles" contains the text "Traitement du ticket..." which is also circled in black. This section includes instructions for the extraction process, such as ignoring promotions and using the "EUR" column for pricing. At the bottom of the right panel is a button labeled "Lancer l'extraction sur le ticket affiché".

Tickets à traiter

Tout sélectionner Traiter la sélection

- Ticket #1 En attente
- Ticket #2 Traitement...**
- Ticket #3 En attente
- Ticket #4 En attente
- Ticket #5 En attente
- Ticket #6 En attente
- Ticket #7 En attente
- Ticket #8 En attente
- Ticket #9 En attente
- Ticket #10 En attente
- Ticket #11 En attente

Prévisualisation du ticket

Extraction des articles

Traitement du ticket...

Prérompt d'extraction

Voici un ticket de caisse.
Retourne tous les articles affichés sur le ticket, sous forme de liste texte.

Ignore totalement tout ce qui concerne les promotions, remises, réductions, cartes de fidélité, ainsi que les informations liées au magasin (adresse, SIRET, TVA, totaux, etc.).

Pour chaque article :

- Utilise uniquement le prix indiqué dans la colonne "EUR", qui correspond au prix final total de la ligne. Ignore complètement la colonne "P.U.EUR", même si cette dernière est renseignée. Ne fais aucun calcul.

Lancer l'extraction sur le ticket affiché

App 3 : Classification mixte KNN pgVector / LLM gemma 3

```
classify_article <- function(article_name, current_db_conn, session_obj, ref_articles_data) {  
  classified_category <- NA_character_  
  
  tryCatch({  
    options(rollama_server = "http://iut-sd2.univ-avignon.fr:11434")  
    article_embedding <- embed_text(article_name, model = "nomic-embed-text")  
    article_embedding_str <- paste0("[", paste(article_embedding, collapse = ", "), "]")  
  
    knn_result <- dbGetQuery(current_db_conn,  
                             sprintf("SELECT article, distance, categorie FROM knn_articles('%s'::vector, 1);",  
                                     article_embedding_str))  
  
    if (nrow(knn_result) > 0) {  
      closest_distance <- knn_result$distance[1]  
      closest_category <- knn_result$categorie[1]  
  
      if (closest_distance <= SIMILARITY_THRESHOLD) {  
        classified_category <- closest_category  
        message(paste("Classifié (Similarité Vectorielle):", article_name, "-->", classified_category, "(Distance:", closest_distance, ")"))  
        session_obj$sendCustomMessage('showNotification', list(message = paste('Classifié par SV:', article_name), type = 'info', duration = 5))  
      } else {  
        message(paste("Similarité Vectorielle insuffisante pour:", article_name, "(Distance:", closest_distance, ">", SIMILARITY_THRESHOLD,  
1
```

Apps 1 & 2 & 3: Conclusions

- Utilité des installations locales sur station de travail internes
- Maturité de l'environnement logiciel R, PostgreSQL, Ollama pour le déploiement de processus ETL assistés par LLMs.
- Disponibilité des techniciens pour leur mise en oeuvre formés par les IUT BUT Sciences des données.

Apps 1 & 2 & 3: Perspectives (admin@adesis.fr)

- Association Loi 1901 pour la Diffusion dans l'Économie de Solutions d'Intelligence Souveraine
<https://adesis.fr>
- Développement réseau SS2I pour déploiement, maintenance, formations à ces solutions:
 - <https://aiedge4u.com/>
 - <https://www.dalibo.com/>
 - <https://www.enterprisedb.com/>
 - <https://www.jumpingrivers.com/>
 - etc.