

# Approche hybride LLM–Ontologie pour la création d'un graphe de connaissances d'expertise d'accidents aéronautiques

**Yvan Valdes Tchoutouo Tcheunjio**

K. Berthe · H. Nassa Trawina · S. Sonfack Souchio

F. A. Mbiaya · F. Issa Soro · Ho Tuong Vinh

*VNU Hanoi – École Internationale · Robert Bosch · Univ. KI-ZERBO · Univ. Orléans (LIFO) · LBA*

**PFIA 2026 · APIA · Arras**



# Plan

**01** Contexte et problématique

**02** Ontologie BAEO

**03** État de l'art

**04** Solution proposée

**05** Expérimentation et résultats

**06** Discussion

**07** Conclusion et perspectives

# Contexte et problématique

En France, le **Bureau d'Enquêtes et d'Analyses (BEA)** documente chaque accident par des rapports détaillés, diffusés en **PDF non structurés**.

177

accidents

18

accidents mortels

30

décès

31

blessés graves

France, 2024 — Source : BEA, rapport annuel 2024. *En net recul vs 2023 : -16 % d'accidents, -45 % d'accidents mortels, -42 % de victimes.*

## Un double défi

- 1 la complexité de l'extraction d'expertise métier
- 2 la garantie de conformité au modèle de données (cohérence sémantique)

*Question : Comment transformer ces rapports en un graphe de connaissances interrogeable, fondé sur l'ontologie BAEO ?*

## Exploitation limitée

Corrélation à grande échelle difficile sur des récits narratifs.

## Expertise difficile d'accès

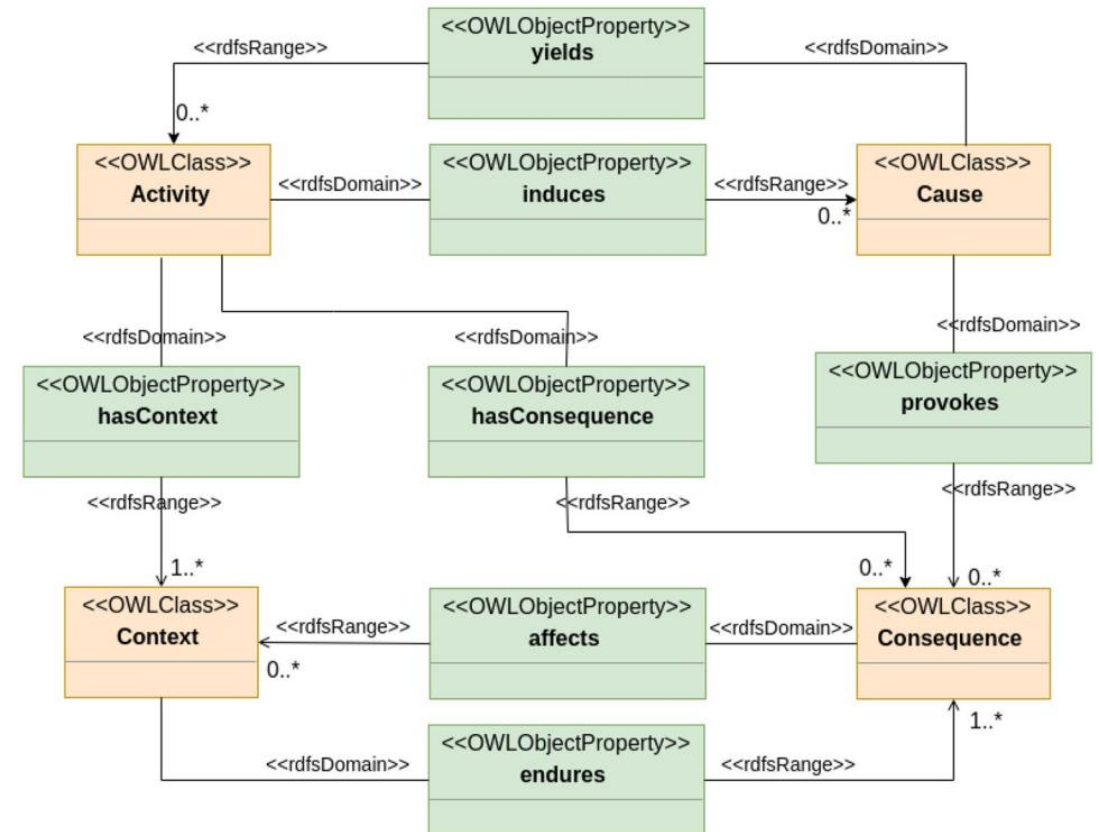
Un savoir métier riche mais peu interrogeable.

# Ontologie BAEO

Base Accident Expertise Ontology (Sonfack et al., 2023)

Ontologie de domaine (méthodologie MDA, OWL 2 DL) structurée en cinq concepts fondamentaux.

- **Activity** – Actions ou processus se déroulant durant l'accident (actions / processus )
- **Context** – Décrit l'environnement et les circonstances de l'accident (Agent, Location, Environment)
- **Cause** – Facteurs déclencheurs identifiés par les experts (Human / System / Natural)
- **Consequence** – Fatality, Failure...
- **Time** – OWL-Time (relations d'Allen)



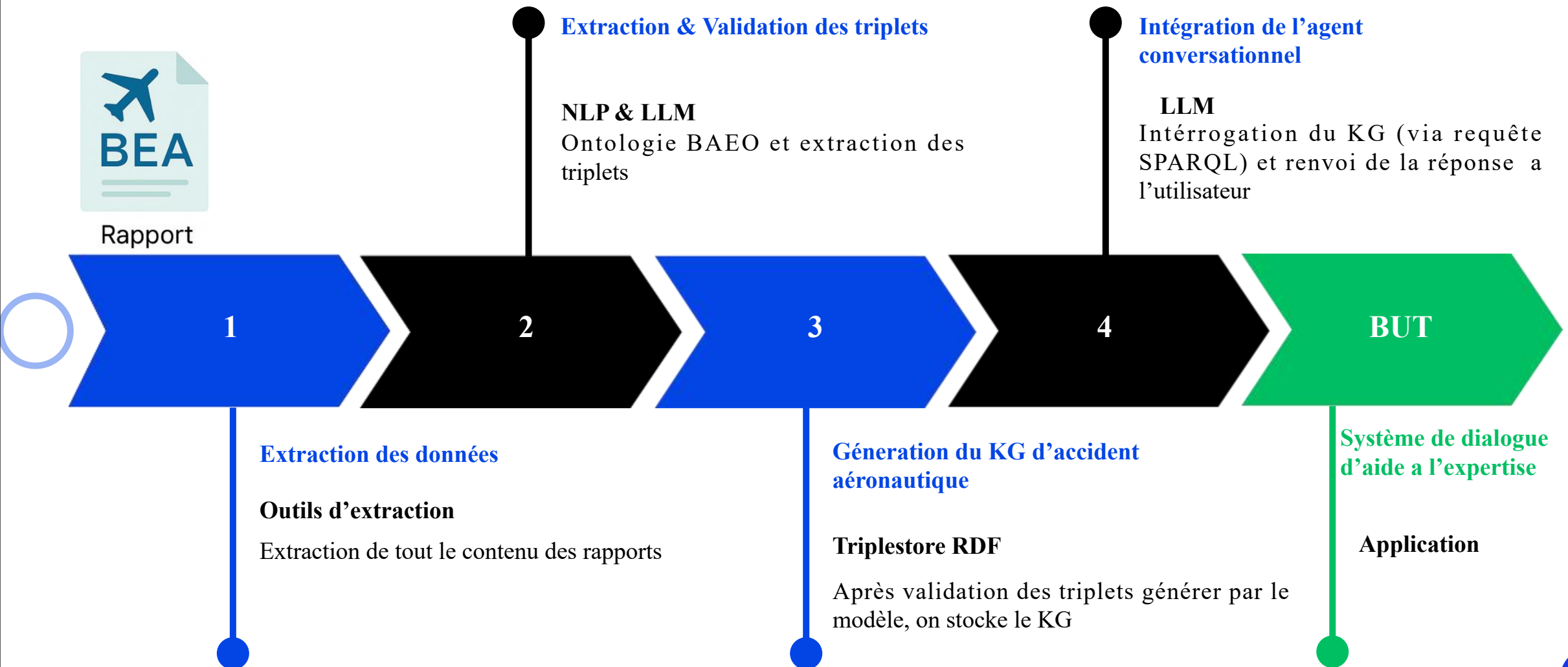
Modèle conceptuel de BAEO

| Travaux                       | Apport                                        | Limite                                 |
|-------------------------------|-----------------------------------------------|----------------------------------------|
| Sonfack et al., 2023          | Ontologie BAEO pour l'expertise d'accidents   | Peuplement manuel chronophage          |
| Feng et al., 2024             | KG par LLM guidé par une ontologie (Wikidata) | Performance variable selon l'ontologie |
| Sun et al., 2024 (Docs2KG)    | Extraction multimodale de KG (PDF, Excel)     | Non spécialisé domaine                 |
| Melnyk et al., 2022 (Grapher) | Génération bout-en-bout entraînée             | Pas de garde-fou ontologique           |

*Notre positionnement : automatiser le peuplement de BAEO via une approche hybride LLM–ontologie garantissant la cohérence structurelle.*

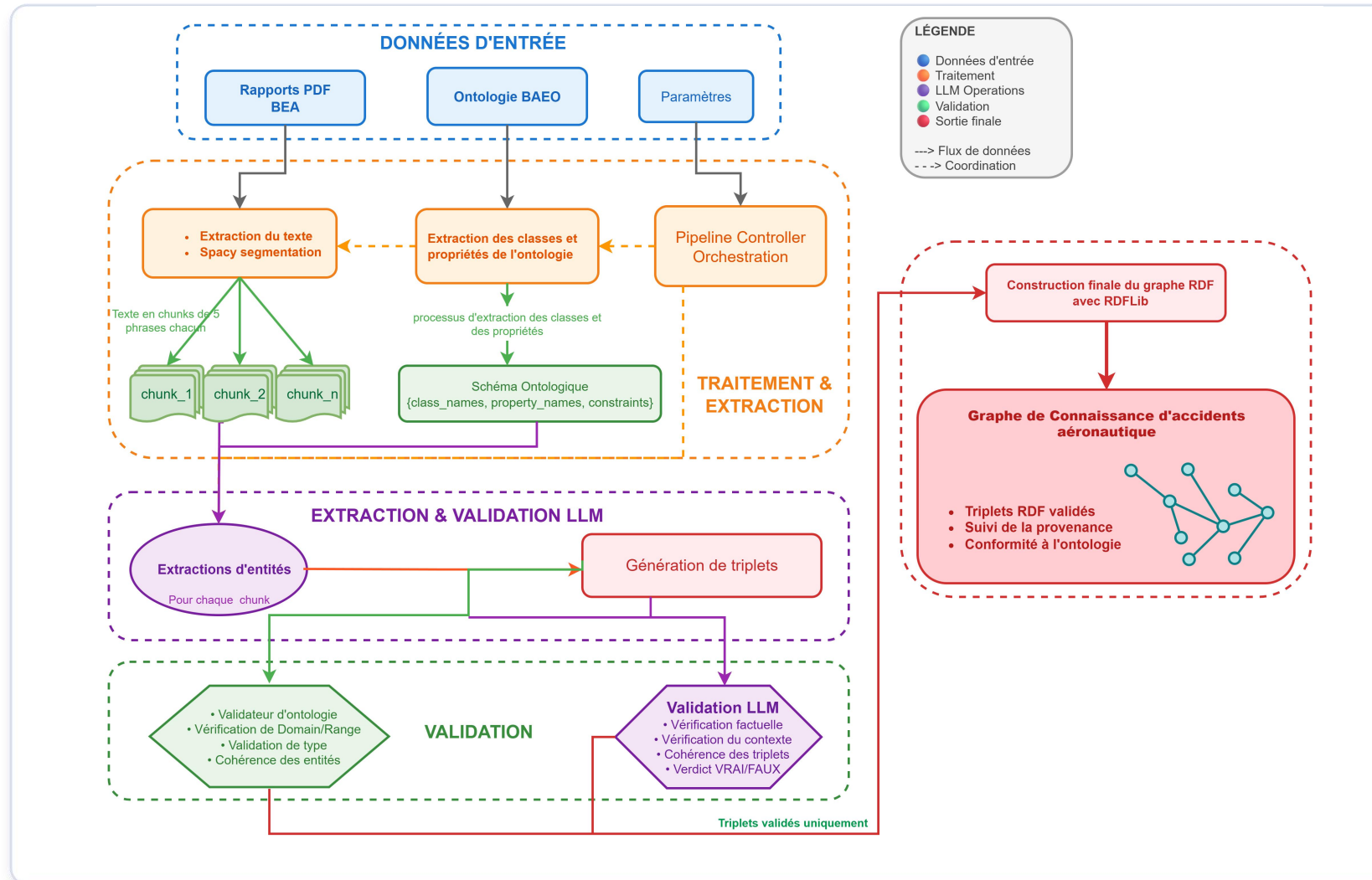
# Solution : vue d'ensemble

*Du rapport BEA au graphe interrogeable*



# Solution : architecture du pipeline

Trois phases s'appuyant directement sur BAEO



## À retenir

- **Pipeline Controller**  
orchestre les trois phases
- **Schéma BAEO injecté**  
class\_names / property\_names dans le prompt
- **Double validation**  
ontologique + factuelle (VRAI/FAUX)
- **Sortie**  
graphe RDF conforme, provenance tracée

# Extraction guidée par l'ontologie

Phases 1–2 et prompt Chain-of-Thought

1

**Prétraitement** — spaCy (fr\_core\_news\_lg) : extraction PDF, segmentation par phrases (chunks), contexte glissant de 3 chunks pour la coréférence.

2

**Extraction guidée par LLM** — un Ontology Handler extrait via SPARQL les classes et propriétés de l'ontologie baeo\_application.ttl et les injecte dans le prompt.

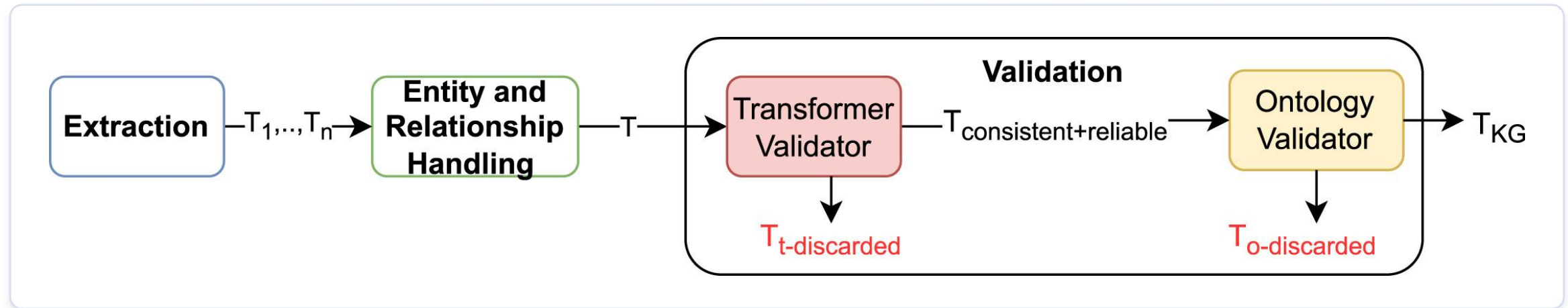
## Prompt Chain-of-Thought contraint

Identification → validation ontologique → vérification factuelle → enrichissement → JSON.

**type / predicate doivent provenir de l'ontologie** ⇒ *réduction des hallucinations structurelles.*



# Validation et insertion (Phase 3)



*Processus de validation automatique des triplets*

## 1 Valideur ontologique

filtre les triplets dont le type sujet/objet contredit rdfs:domain / rdfs:range.

## 2 Validation factuelle (LLM)

vérifie que chaque fait est présent dans le texte. Seuls les triplets doublement validés sont insérés (RDFLib).

## 1 Corpus

3 rapports BEA, dont BEA2016-0759 (Piper PA34-200T HB-LSD, Bâle-Mulhouse, 07/12/2016), ~160 chunks.

## 2 Modèles évalués

Mistral-7B-Instruct-v0.3, Llama-3.1-8B-Instruct et Mistral-large.

## 3 Évaluation

100 triplets + 20 requêtes SPARQL (5 catégories : Chronologie des événements, Facteurs humains, Facteurs techniques (Aéronef), Facteurs environnementaux, Analyse et conclusions).

Métriques : précision, rappel, F1.

Code & ontologie : [github.com/Tchoutouo/KG-Builder](https://github.com/Tchoutouo/KG-Builder)



## Résultats : comparaison des modèles

| Modèle                | Précision | Rappel | F1     |
|-----------------------|-----------|--------|--------|
| Mistral-large         | 96 %      | 91 %   | 93,4 % |
| Mixtral-8x7B          | 89 %      | 84 %   | 86,4 % |
| Llama-3.1-8B-Instruct | 74 %      | 68 %   | 70,9 % |

96%

précision Mistral-large (Mistral AI)

**Mistral-large** suit nettement les contraintes ontologiques ; les 20 requêtes SPARQL retournent toutes des résultats corrects.

**Llama-3.1-8B** : hallucinations de schéma, coréférence défailante et JSON malformé (taille réduite, 8B).

# Interrogation du graphe (SPARQL)

Exemples de requêtes sur le graphe Mistral-large

Anonyme X Anonyme X Anonyme X +

1 # 4. Quelle a été la durée de la montée rapide après le virage ?

2

3 PREFIX baeo: <http://www.enit.fr/2022/03/baeo#>

4 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

5

6 SELECT ?label

7 WHERE {

8   ?duration a baeo:Duration ;

9       rdfs:label ?label .

10   FILTER(CONTAINS(lcase(str(?label)), "montée post-virage"))

11 }

12

Exécuter

raccourcis clavier

Tableau Réponse brute Table de pivotement Graphique Google

Téléchargement

Filter les résultats des requêt

Liste de résultats de 1 à 1 pour un total de 1. La requête a pris 0.1s, il y a quelques instants.

|   | label                                                |
|---|------------------------------------------------------|
| 1 | "Durée de 10 secondes pour la montée post-virage"@fr |

Chronologie : montée rapide après le virage

Anonyme X Anonyme X Anonyme X +

1 # 5. Quel était l'âge et la date de la licence du pilote du HB-LSD ?

2 PREFIX baeo: <http://www.enit.fr/2022/03/baeo#>

3 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

4

5 SELECT ?age ?licenceInfo

6 WHERE {

7   ?pilote a baeo:Human ;

8       rdfs:label ?label ;

9       baeo:hasAge ?age .

10 OPTIONAL {

11   ?pilote baeo:hasTRS ?licence .

12   ?licence rdfs:comment ?licenceInfo .

13   FILTER(CONTAINS(lcase(str(?licenceInfo)), "licence"))

14 }

15 FILTER(CONTAINS(lcase(str(?label)), "pilote") && CONTAINS(lcase(str(?label)), "61 ans"))

16

Exécuter

raccourcis clavier

Tableau Réponse brute Table de pivotement Graphique Google

Téléchargement

Filter les résultats des requêt

Liste de résultats de 1 à 2 pour un total de 2. La requête a pris 0.1s, il y a quelques instants.

|   | age               | licenceInfo                                                                                                                                                                                                                                                                                                                                      |
|---|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | "61"^^xsd:integer | "Document officiel autorisant le pilote à opérer des aéronefs en tant que pilote privé. Délivrée par l'Office Fédéral de l'Aviation Civile (OFAC) de Suisse en 1993, avec une validité confirmée au moment de l'accident. La licence est un prérequis légal pour les qualifications ultérieures mentionnées (monomoteur, multimoteur, etc.)."@fr |
| 2 | "61"^^xsd:integer | "Licence délivrée en 1993 par l'OFAC suisse, valide au moment de l'accident"@fr                                                                                                                                                                                                                                                                  |

Facteurs humains : licence du pilote (HB-LSD)

Y. V. Tchoutouo et al. · PFIA 2026 · Graphe de connaissances d'accidents aéronautiques

13

**96 % de précision** confirme la viabilité des approches hybrides LLM–KG : synergie entre compréhension contextuelle et rigueur structurelle de BAEO.

**Cadre d'usage :** un outil d'aide à l'expertise post-accident, pas un système décisionnel autonome — l'expert humain reste souverain.

## Limites

- tableaux et schémas dans les PDF rompant la continuité sémantique ;
- forte sensibilité au modèle (chute de Llama-3.1-8B) ;
- relations temporelles implicites et nuances d'incertitude partiellement capturées.

# Conclusion et perspectives

Un pipeline automatisé transforme des rapports BEA non structurés en un graphe RDF d'expertise (**96 % de précision avec Mistral-large**) ; l'extraction guidée par BAEO réduit les biais des LLM tout en préservant la richesse de l'expertise métier.

## Perspectives

- **Multimodalité**  
images, diagrammes et trajectoires de vol
- **Agent conversationnel**  
interroger la base en langage naturel
- **Validation humaine en boucle**  
interface pour les enquêteurs du BEA
- **SHACL + LLM juge / Fine-tuning**  
contraintes expressives et LLM léger



# Merci pour votre attention !

[yvalttnomdecode@gmail.com](mailto:yvalttnomdecode@gmail.com)

[github.com/Tchoutouo/KG-Builder](https://github.com/Tchoutouo/KG-Builder)

*PFIA 2026 · APIA · VNU – IS · IFI*



# PFIA 2026

PLATFORME  
INTELLIGENCE  
ARTIFICIELLE

CONFERENCES

APIA  
CNIA  
IC  
JFSMA  
RDPIA  
RJCIA

29 juin  
**ARRAS**  
3 juillet



[pfia2026.pfia.fr](https://pfia2026.pfia.fr)

