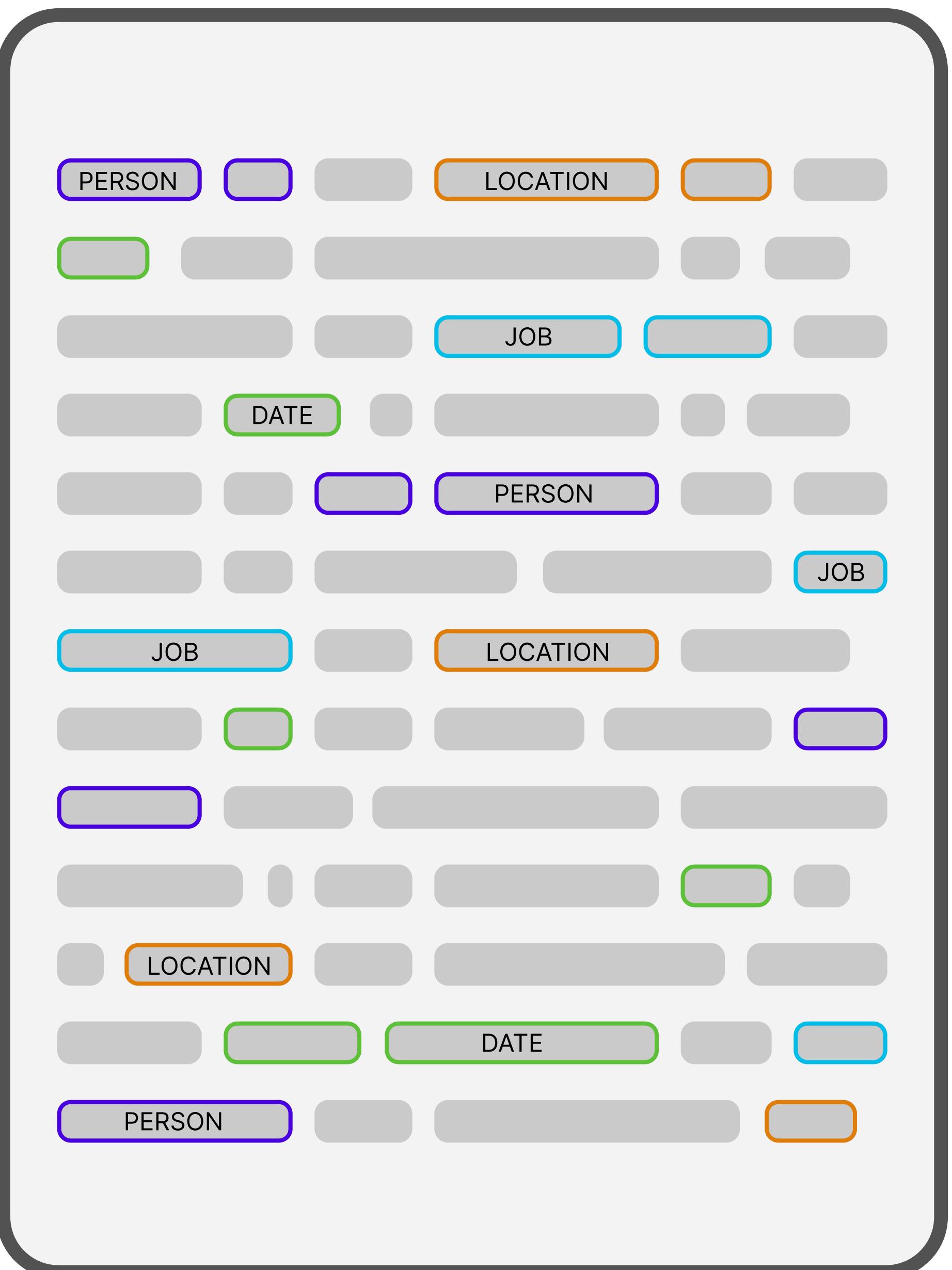


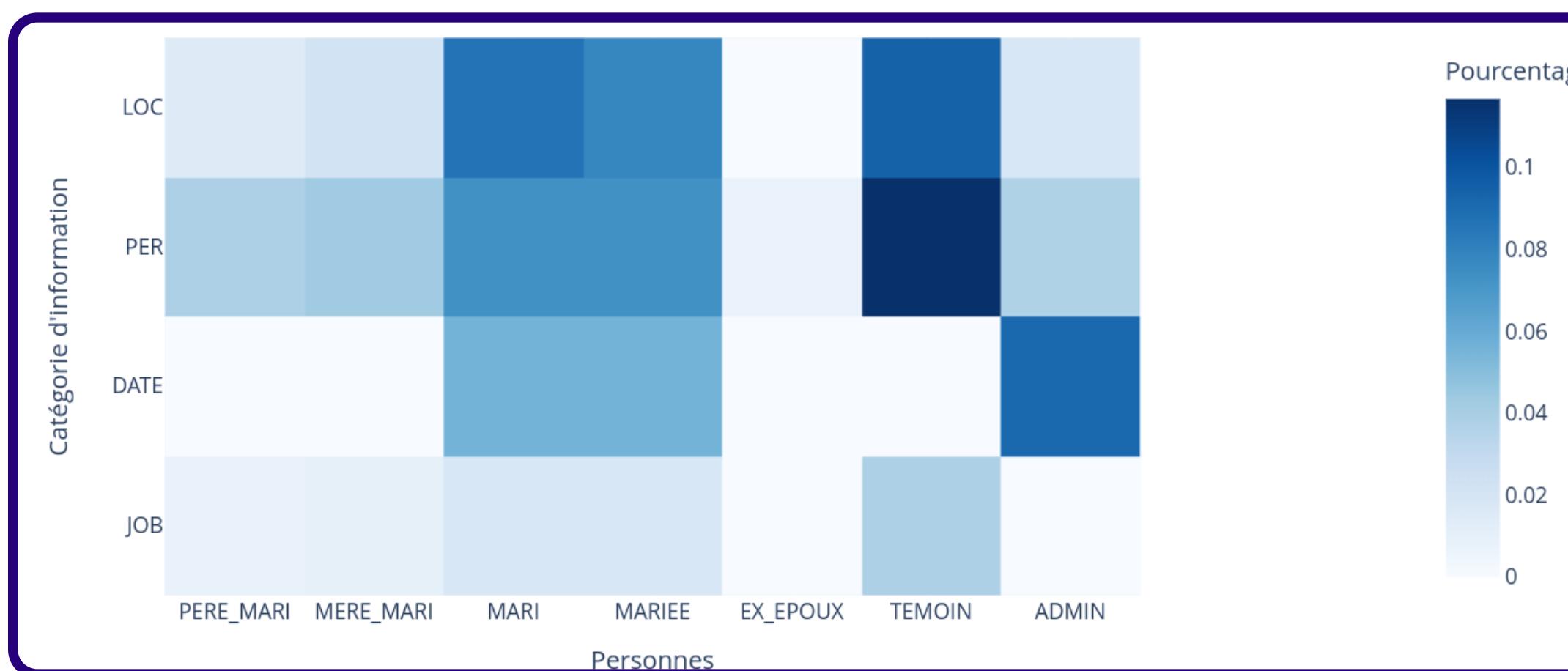
Extraction d'entités nommées dans le cadre du projet EXO-POPP

07/07/23



Le projet EXO-POPP

Niveau	Tags
1	Administratif Mari Epouse Témoin
1.1	Père Mère Ex-Epoux
2	Naissance Residence Décès Disparu Age Veuf
3	Profession Prenom Nom
4	Pays Département Ville Numéro voie Type voie Nom voie
	Année Mois Jour Heure Minute



→ Les Tags sont classés par niveaux : *Nom → ex-épouse → mari*

→ Il peut y avoir des changements dans les formulations, comme lors d'un décès, d'un scripteur différent ou des cas de pluriels :

...Denis, résidant 8 rue du palais, avec ses père et mère...

...Homer et Marguerite, marchands de bestiaux...

...fils de Simon, décédé, et Ophelie, résidant au 7 place de l'arche...

...fille de Bertrand et de Julie, sa veuve, résidant au 4 boulevard Pascal...

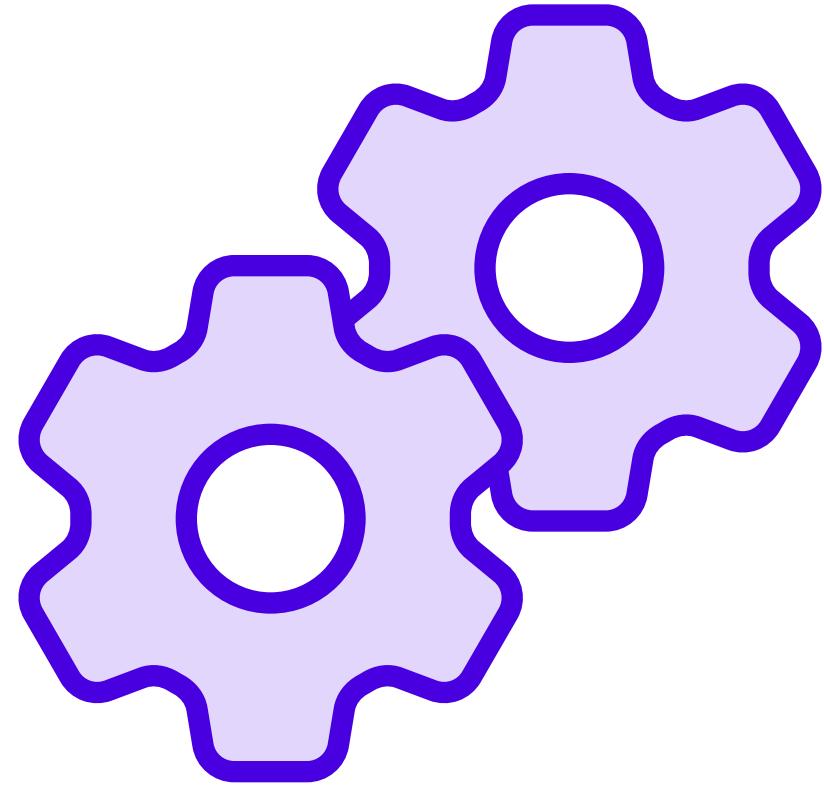
→ Les informations à relever sont nombreuses, souvent complexes et apparaissent parfois rarement.

neuf cent douze, vingt-sept ans, domicilié à Paris, 16, rue Hérole: fils de Ernest CARDOT et
de Julie GANIVET, son épouse, retraités, domiciliés à Saint Amand de Puisaye * Nièvre *

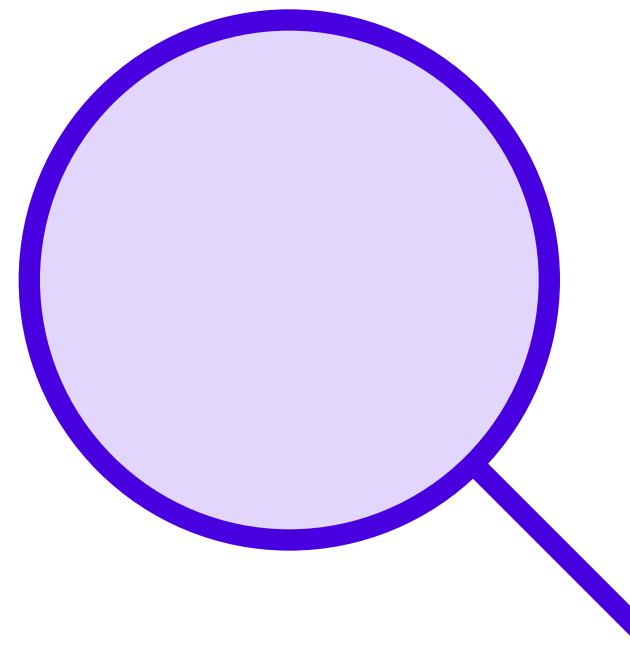
Étiquettes utilisées dans : retraités,

< MARI ✕ PÈRE ✕ PROFESSION ✕ ÉPOUSE ✕

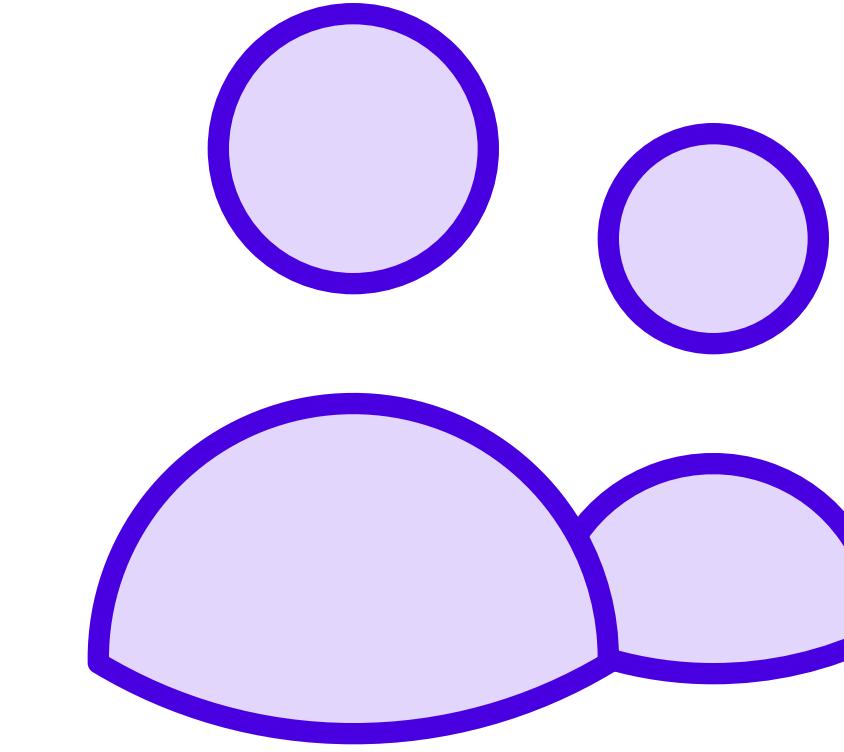
Les limites d'une approche classique



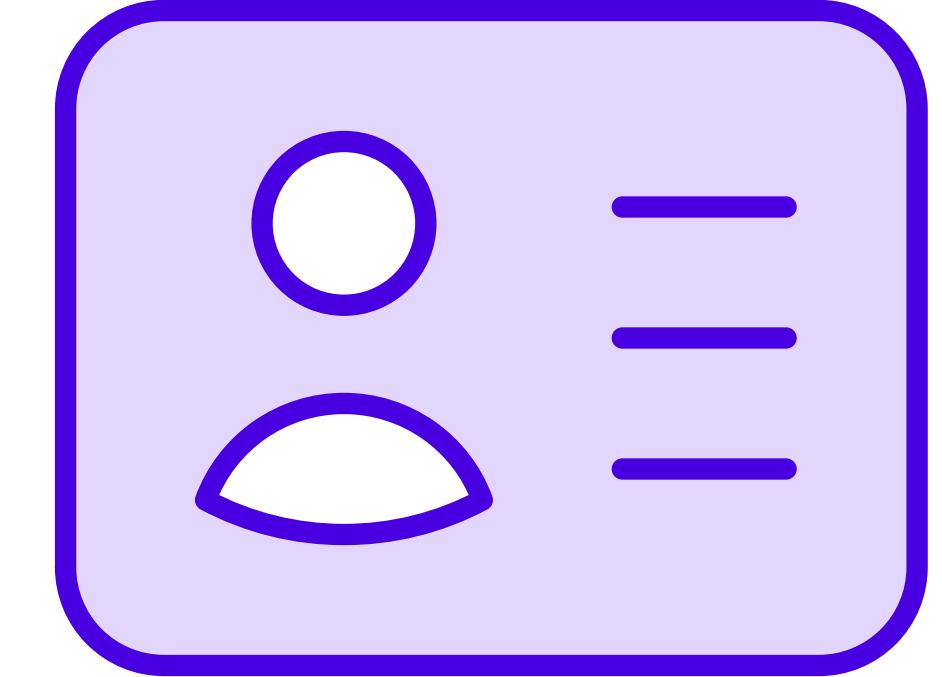
Demandeur en ressources



Les entités rares posent problème



Les données doivent être annotées



Les tags complexes posent problème

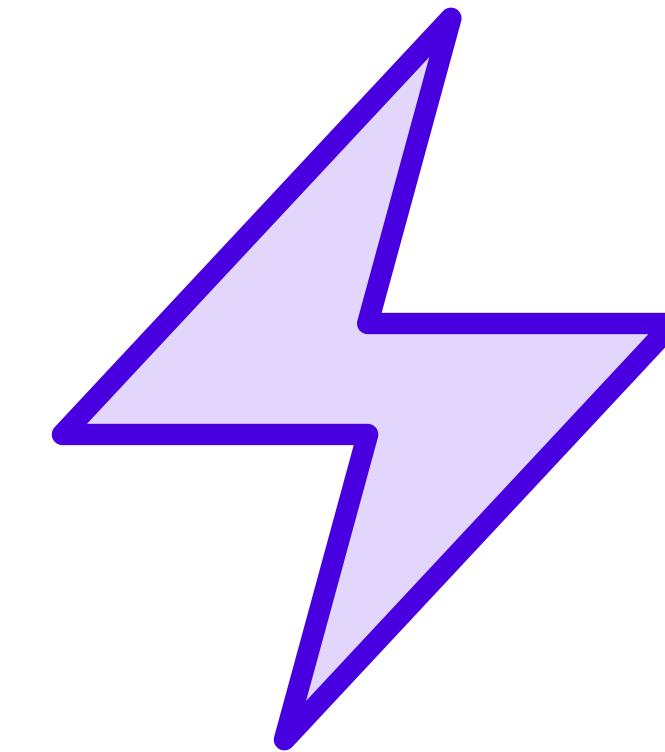
→ Le projet EXO-POPP pose problème dans ces quatres domaines

→ Les Large Language Models sont une piste à explorer

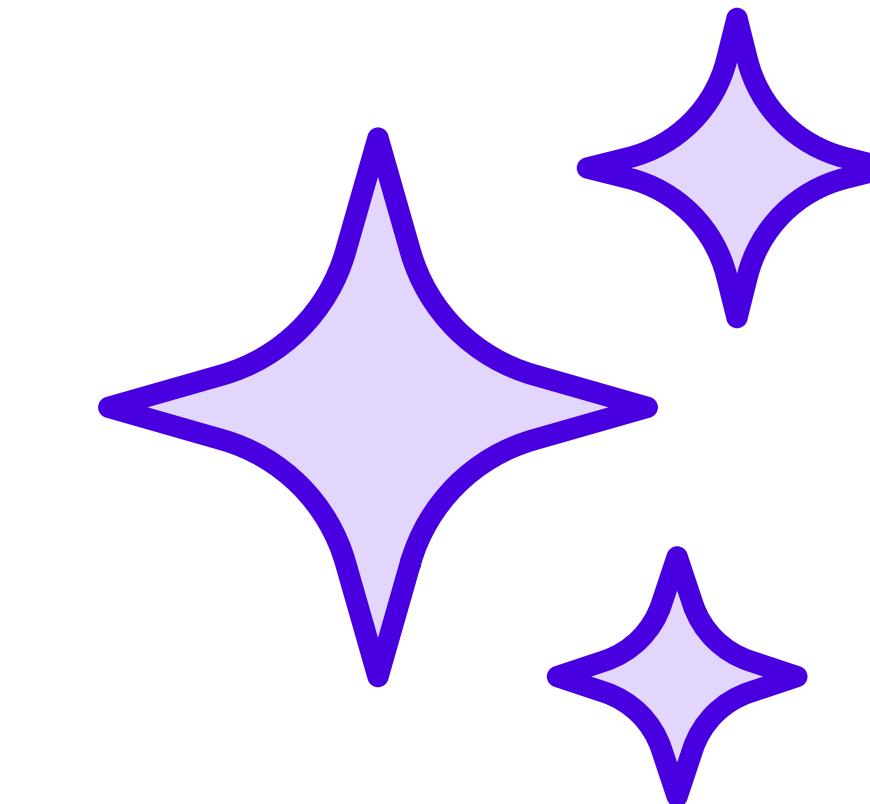
GPT3.5, EN QUOI C'EST INTERRESSANT ?



Excellent adaptation
à une tâche



Besoin de très peu
d'exemples



Détournement des
capacités génératives

⇒ Plutôt que de taguer des éléments, pourquoi ne pas générer directement les réponses à ce que l'on cherche ?

NER AVEC GPT3.5 - Few-Shot Learning

Les actes tapuscrits (entre 1930 et 1940) sont toujours organisés avec la même structure :

Paragraphe 1 5 labels

Le jour, la date et l'heure du mariage

Paragraphe 2 37 labels

Toutes les infos concernant le mari et sa famille

Paragraphe 3 37 labels

Toutes les infos concernant la mariée et sa famille

Paragraphe 4 0 label

La présence d'un contrat de mariage et les noms des mariés

Paragraphe 5 21 labels

Toutes les infos concernant les témoins et l'adjoint au maire

→ En séparant les actes selon celle-ci, on peut fragmenter la collecte des informations et ainsi faire rentrer plusieurs exemples dans une requête

NER AVEC GPT3.5 - Few-Shot Learning

One-shot Learning

Acte
exemple
Complet

Labels
exemple
Complets

Acte
question
Complet

Labels
réponse

Few-shot Learning

Exemple
Paragraphe 1

Labels §1

Exemple
Paragraphe 1

Labels §1

Exemple
Paragraphe 1

Labels §1

Question
Paragraphe 1

Réponse §1

Exemple
Paragraphe 2

Labels §2

Question
Paragraphe 2

Réponse §2

Exemple
Paragraphe 3

Labels §3

Question
Paragraphe 3

Réponse §3

Exemple
Paragraphe 4

Labels §4

Question
Paragraphe 4

Réponse §4

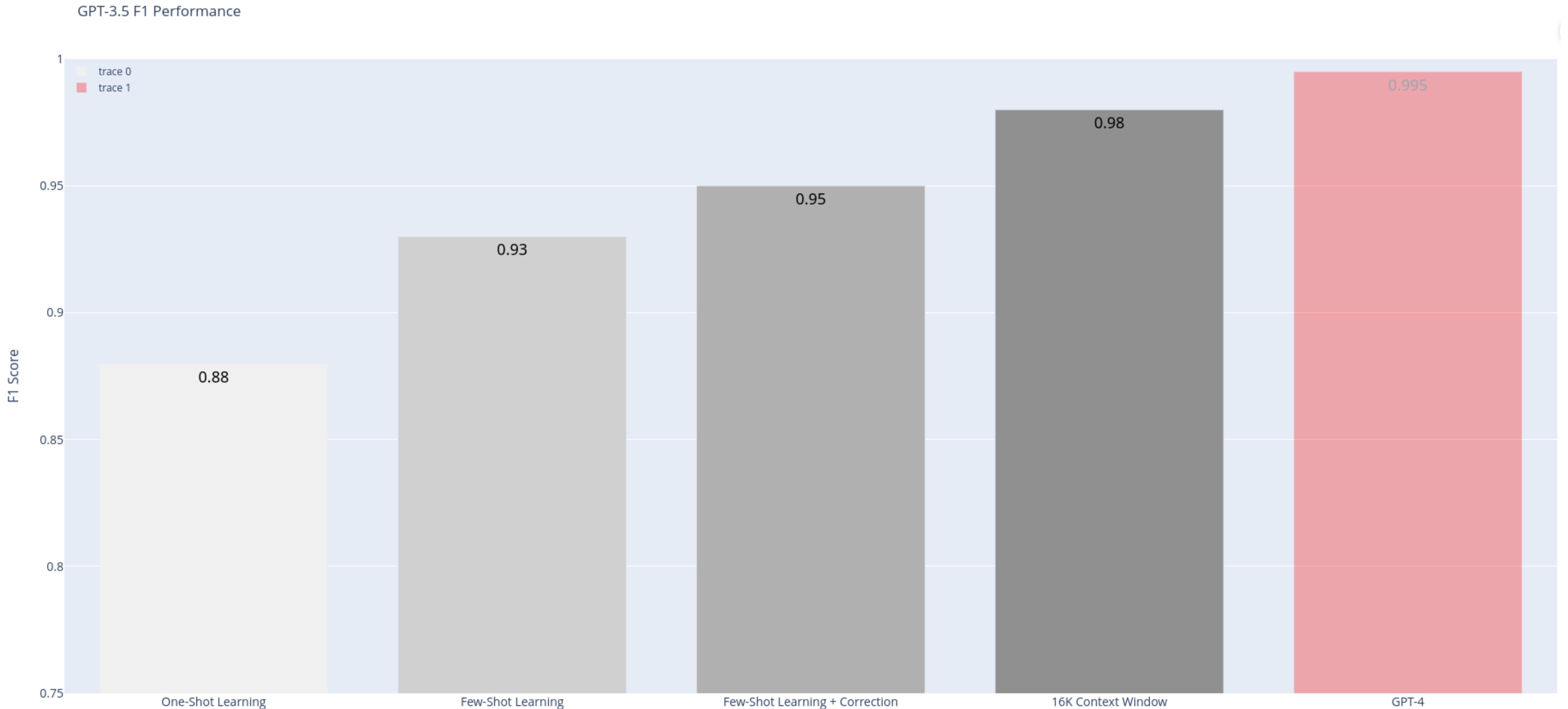
Exemple
Paragraphe 5

Labels §5

Question
Paragraphe 5

Réponse §5

Performances obtenues au cours du temps



Utilisation : format CSV

Utilisation : format ‘EMOJI’ pour comparer avec le DAN

  vingt-huit     juin   31   1 mil neuf cent trente   ,   douze heures   douze minutes   devant Nous ont comparu
publiquement en la maison Commune:

Gilbert Alexandre Jean VALO, fraiseur né à Saint Aubin des Bois, (Eure et Loir), le vingt-huit février mil neuf cent-huit vingt-deux ans, domicilié 143 rue Legendre, fils de jean marie Joseph VALO décédé et de Anne Marie LOSSOUARN, sa veuve, commerçante, domiciliée à Saint Aubin (Eure et Loir), présente et consentante, d'une part ./-

Et Marcelle Fernande WENDER, couturière, née à Paris XX^e arrondissement, le 151 rue de la Roquette, avec ses père et mère, fille de Fernand Edmond WENDER, tourneur, et de Marie BACHELET, son épouse, couturière, présents et consentants, d'autre part.

Aucune opposition n'existant, les futurs époux les père et mère de la future épouse déclarent qu'il n'a pas été fait de contrat de mariage.;   Gilbert Alexandre Jean    VALO   et   Marcelle Fernande    WENDER   ont déclaré l'un après l'autre vouloir se prendre pour époux et Nous avons prononcé Au nom de la loi qu'ils sont unis par le mariage.

En présence de:  Jeanne   LOUSSOUARN   employée à la Ville de Paris  ,  114 
   rue   Jean Jaurès  à   Levallois Perret  et de  Jean   CHEMIN 
  commis principal aux Postes  , Médaille Militaire, Croix de Guerre,  16    rue   , témoins majeurs, qui, lecture faite, ont signé avec les époux, la mère de l'époux les père et mère de l'épouse et Nous,  Louis 
  PINOTEAU , adjoint au Maire du onzième arrondissement de Paris , Chevalier de la Légion d'Honneur ./.

Conclusion

L'utilisation de LLM permet de palier à certains aspects problématiques en NER :

- Manque de données annotées
- Grand nombre d'entités à extraire
- Temps et ressources nécessaires conséquents

⇒ Un coût très réduit et de bonnes performances peuvent représenter une alternative préliminaire à une annotation manuelle.