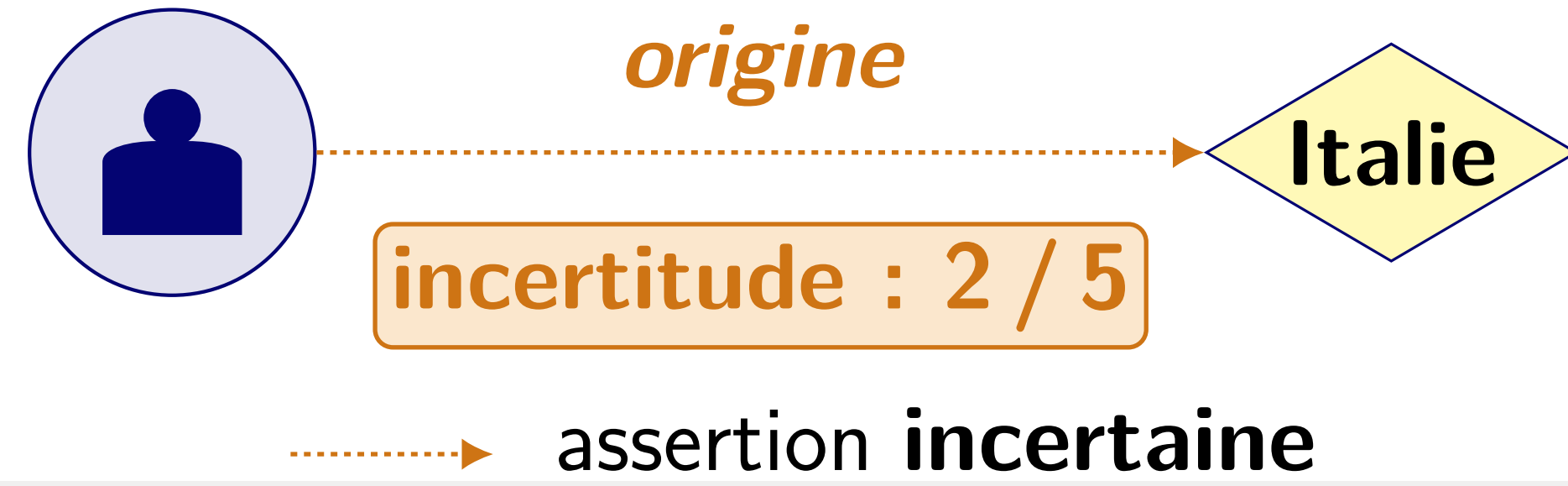


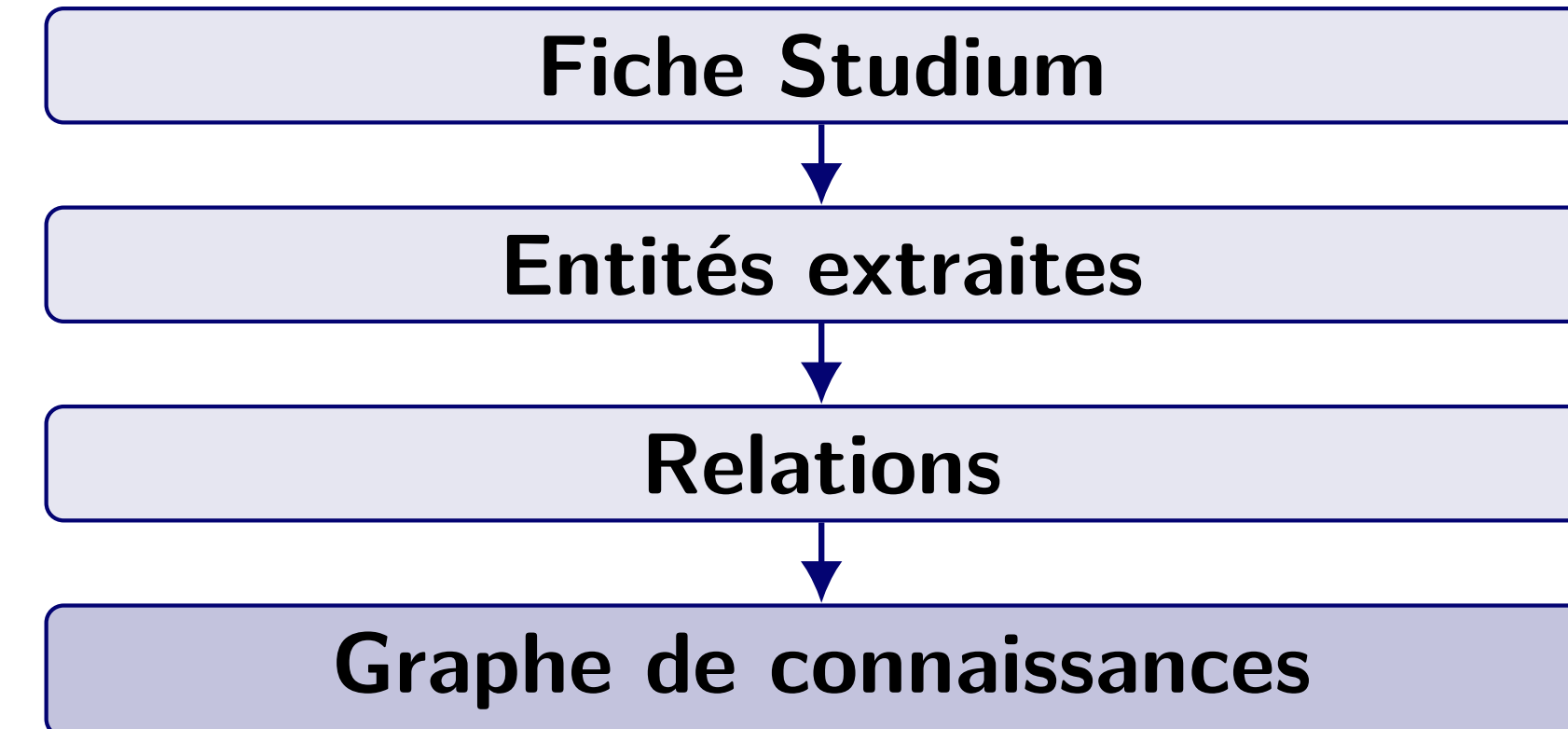
## Introduction

**Problème** : Les fiches historiques contiennent des informations **incertaines**, **imprécises** ou **manquantes**.

**Objectif** : Analyser ces informations en vue de leur intégration dans un **graphe de connaissances**.

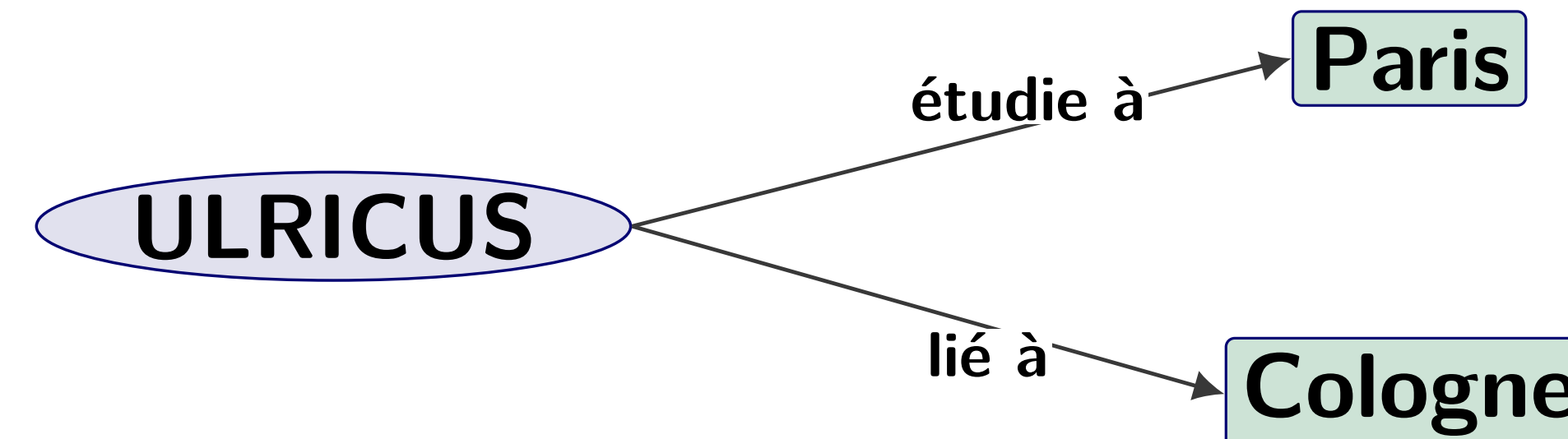


## Du texte au graphe



**Entités** : personne · lieu · institution · date

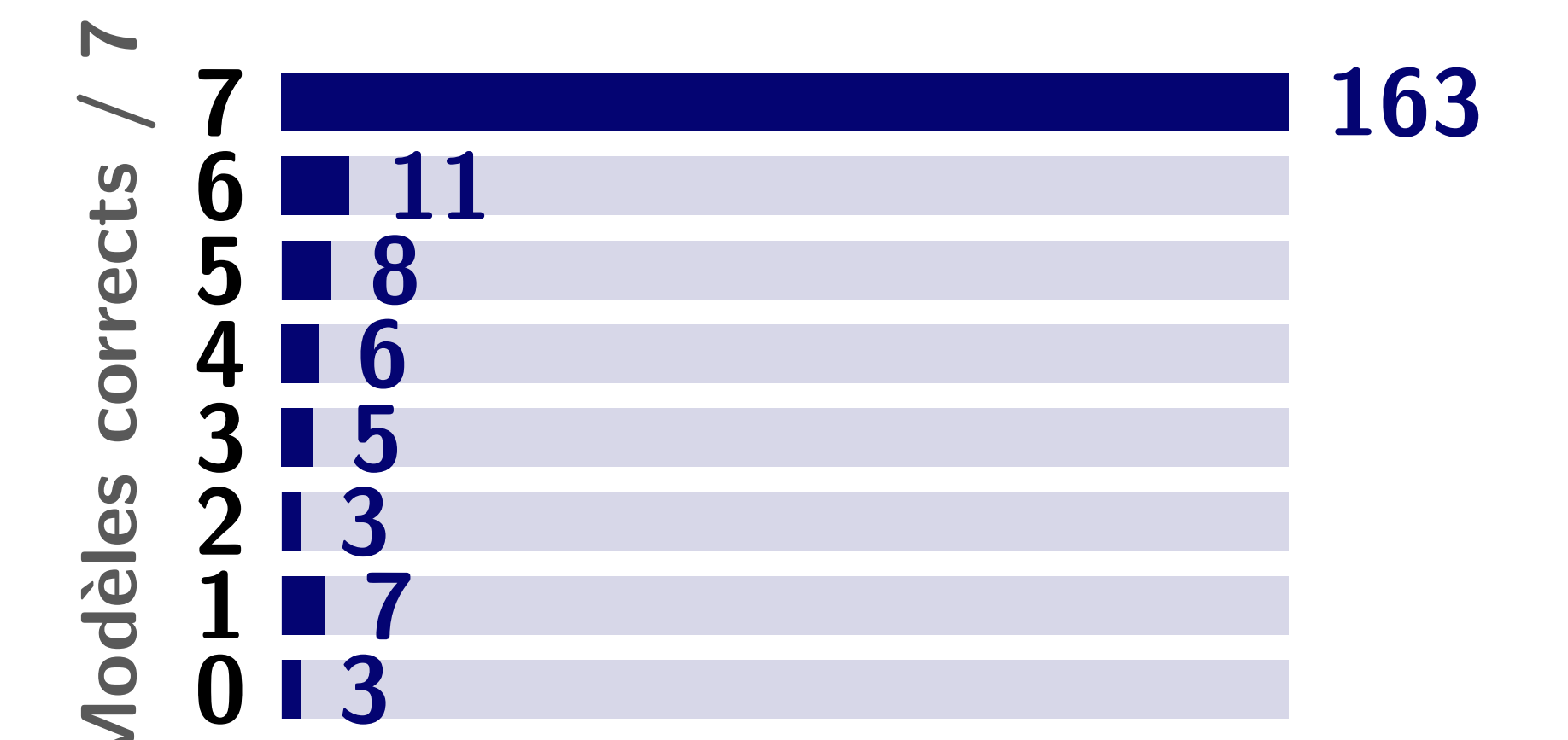
**Relations** : origine · affiliation · grade · relation · source



## Difficulté des exemples du benchmark

Nombre de modèles corrects par exemple

(annotation correcte = type + portée)



Lecture : 7/7 = réussi par tous les LLM ; 0/7 = échoué par tous.

## Données : Studium Parisiense

- ▶ ~20 000 fiches de personnes
- ▶ Membres de l'Université de Paris, XII<sup>e</sup>–XVI<sup>e</sup> siècle
- ▶ Noms, dates, carrières, institutions, relations, sources, etc.

## Exemples de données Studium

**Variante de nom** : Ulrich de STRASBOURG  
Uldaricus ENGELBERTI

**Lieu de naissance** : Italie ? **Date de vie** : -1278

**Université** : ?

**Activité** : vers 1248–1254 **Grade** : Bachelier en théologie (Paris)

**Commentaire** : « Son parent est **peut-être** Guillelmus de VISSACO, sacristain de Cluny »

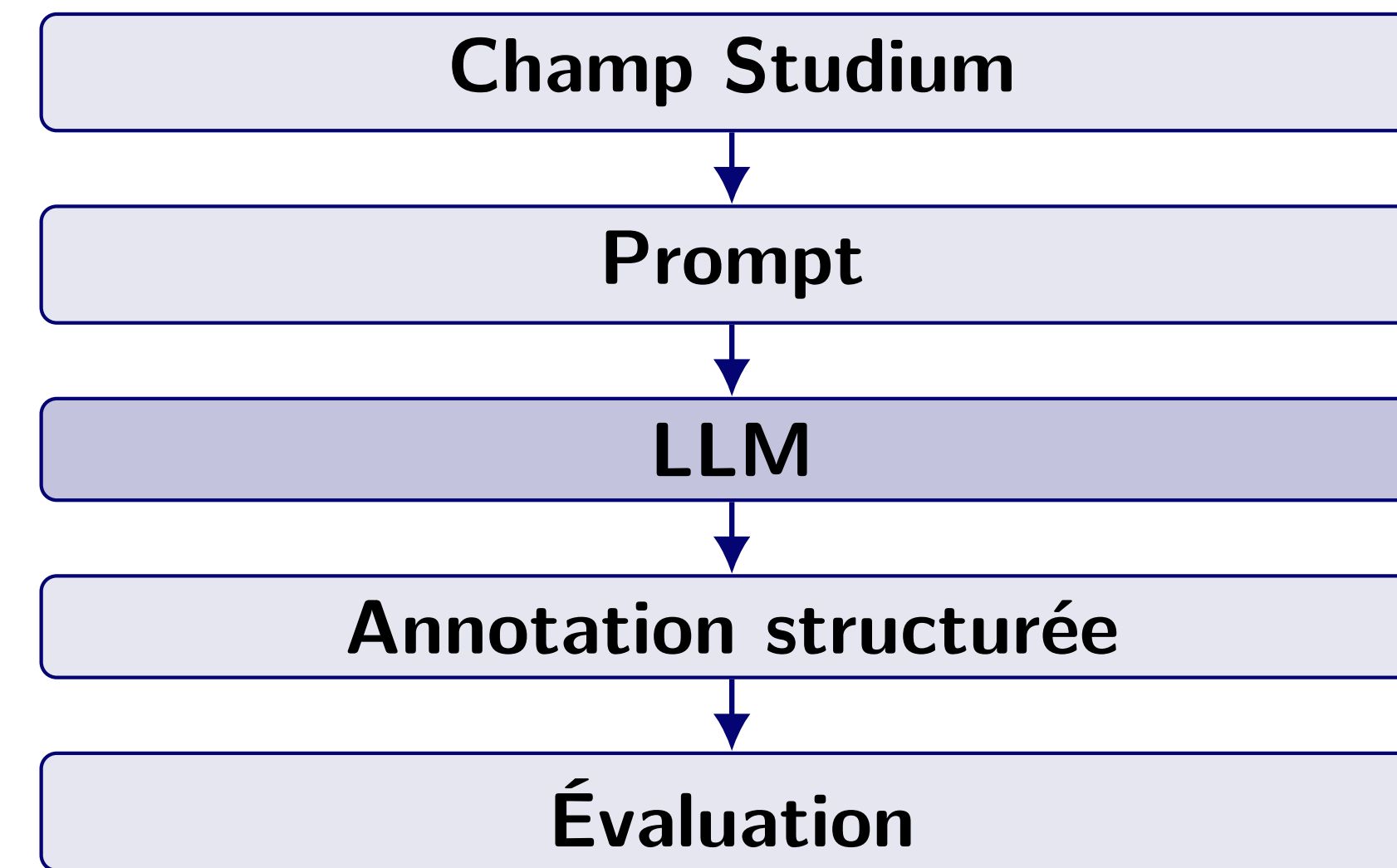
**Commentaire** : « Il est mort à Paris avant d'avoir complété son cursus pour la maîtrise ? »

**Identité** : « L'Adam de KALKER n'est **peut-être qu'un fantôme** »

incertitude    imprécision    variante / information manquante

Extraits de fiches différentes

## Pipeline



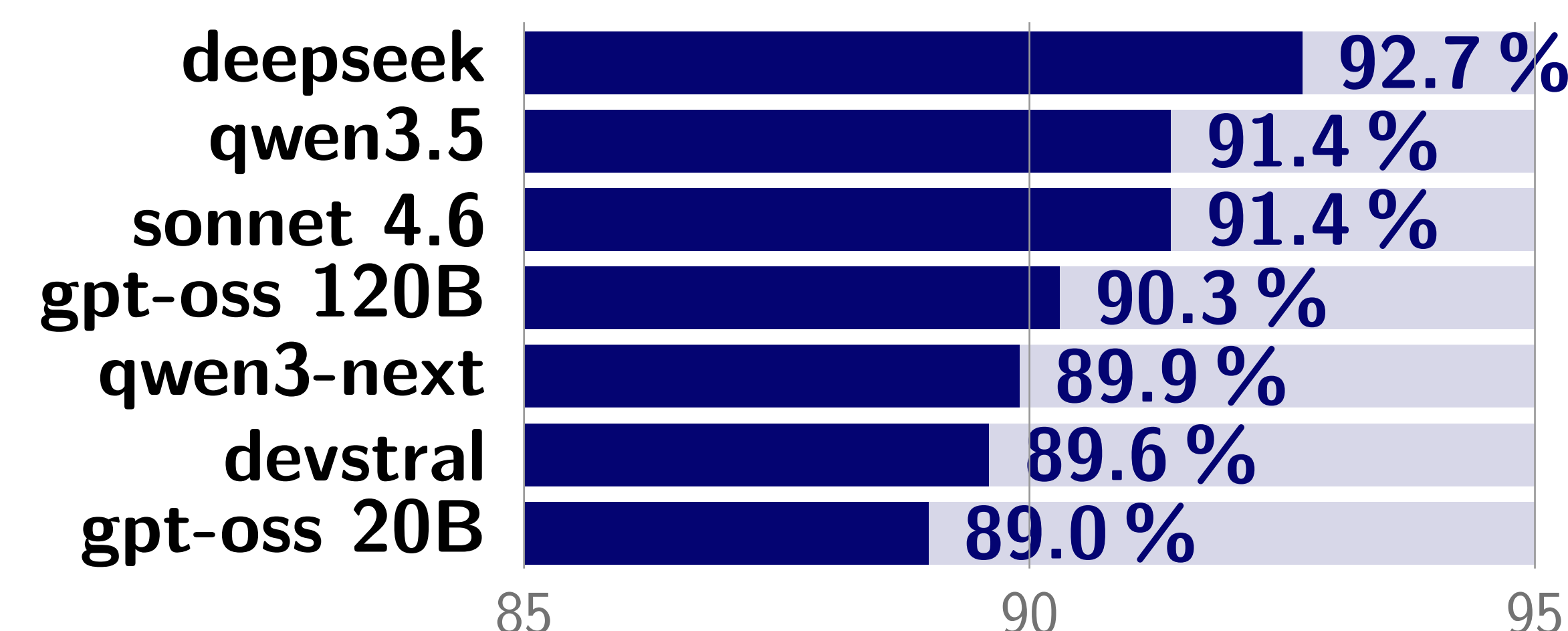
206 exemples    158 positifs    48 négatifs

Évaluation : type · portée · certitude

## Résultats

7 LLM testés sur 206 exemples. **Détection binaire** (incertitude, imprécision, manquant, aucune) : **95-99 %**.

F1 sur annotation : **type + portée**



## Cas limites

**Paris** : lieu ou institution ?

Cursus ▶ Université : Paris (Nation de France) ?

**Paris = ville**  
→ lieu

**Paris = Université**  
→ institution

**Annotation de référence = institution**  
Des modèles répondent lieu, d'autres institution.

« Il a également séjourné auprès de son cousin à Avignon »

**Annotation possible** :

- ▶ cousin comme **entité** → information manquante
- ▶ **relation** familiale seule → information suffisante

La granularité du graphe change l'annotation.

## Perspectives

- ▶ Enrichir le benchmark
- ▶ Analyser les **erreurs de portée**
- ▶ Tester d'autres **prompts** et d'autres **modèles**
- ▶ Intégrer ces annotations dans le **graphe**

