LELA: an LLM-based Entity Linking Approach with Zero-Shot Domain Adaptation

Samy Haffoudhi, Fabian Suchanek, Nils Holzenberger

Institut Polytechnique de Paris

October 28, 2026





Which “Paris” does this text refer to?

France hosted the 2024 Olympics in Paris.

Entities in the knowledge base

Entity linking

Paris (city)	Capital city of France.
Paris (novel)	1897 novel by Émile Zola.
Paris (Texas)	City in Texas, United States.

The mention is already identified.

Why is Entity Linking difficult?

1. Mention ambiguity

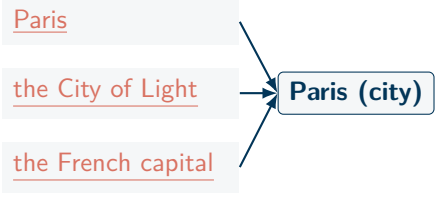
The same mention can refer to different entities.

Paris is the capital of France. → **Paris (city)**

Paris won the match 2–0. → **Paris Saint-Germain**

2. Mention variability

Different mentions can refer to the same entity.



String matching is not enough

Linking requires the mention, its context, and information about candidate entities.

A new knowledge base should not require new annotations

Deployment scenario

A chemical manufacturer wants to link product names in laboratory reports to products in its internal catalog.

Existing supervised methods require annotated mention–entity pairs. In specialized domains, **creating these labels requires expert time**.

Setting	KB at test time	Domain at test time
Classical EL	Same	Same
Zero-shot EL	New	Same / related
True zero-shot EL	**New**	**New**

LELA's goal: true zero-shot entity linking

Link entities in a new domain and target knowledge base, without in-domain annotations or fine-tuning.

Training here refers to supervised EL fine-tuning after pretraining.

Can you disambiguate this mention?

After adding APC, the blood sample took longer to form a clot. The preparation contained no intact cells.

Candidate entity	KB description
Antigen-presenting cell	Immune cell that presents antigens to other immune cells.
Adenomatous polyposis coli gene	Gene involved in regulating cell growth.
Activated protein C	Enzyme that inhibits proteins responsible for blood clotting.

Illustrative sentence. Definitions adapted from NCI, NCBI: APC and NCBI: PROC.

Can you disambiguate this mention?

After adding APC, the blood sample took longer to form a clot. The preparation contained no intact cells.

Candidate entity	KB description
Antigen-presenting cell	Immune cell that presents antigens to other immune cells.
Adenomatous polyposis coli gene	Gene involved in regulating cell growth.
Activated protein C	Enzyme that inhibits proteins responsible for blood clotting.

Illustrative sentence. Definitions adapted from NCI, NCBI: APC and NCBI: PROC.

Can you disambiguate this mention?

After adding APC, the blood sample took longer to form a clot. The preparation contained no intact cells.

Candidate entity	KB description
Antigen-presenting cell	Immune cell that presents antigens to other immune cells.
Adenomatous polyposis coli gene	Gene involved in regulating cell growth.
Activated protein C	Enzyme that inhibits proteins responsible for blood clotting.

Correct link: Activated protein C

Context and entity descriptions enable reasoning about unfamiliar entities.

LELA: a coarse-to-fine pipeline

France hosted the 2024 Olympics in Paris.

1. Retrieve	2. Rerank	3. Select
Find candidate entities.	Keep the top k candidates.	Reason jointly and choose.

Coarse to fine

Knowledge base $\longrightarrow$ Candidates $\longrightarrow$ Top k $\longrightarrow$ **Paris (city)**

LELA, step 1: retrieve candidate entities

France hosted the 2024 Olympics in Paris.

Search the knowledge base

- Index each entity as a document: **title + description**.
- Query using the marked mention, Paris.
- Retrieve candidates using BM25 or a dense retriever.

Retrieved candidates

Paris (city)
Capital city of France.

Paris (novel)
1897 novel by Émile Zola.

Paris (Texas)
City in Texas, United States.

...

Retrieval determines candidate recall.

Illustrative candidates. Retrieved entities become the input to the reranker.

LELA, step 2: rerank each candidate

France hosted the 2024 Olympics in Paris.

One candidate per model call

Supply the full context and a candidate's name and description.

Example candidate

Paris (city)

Capital city of France.

Does the marked mention refer to this entity?

Score from yes/no logits

Read the next-token logits for `yes` and `no`:

$$s(c) = \frac{e^{\ell_{\text{yes}}}}{e^{\ell_{\text{yes}}} + e^{\ell_{\text{no}}}}$$

Repeat for each retrieved candidate.

Keep the k highest-scoring candidates ($k = 10$ in our experiments).

The reranker scores candidates independently. Joint comparison follows.

LELA, step 3: reason over the shortlist

France hosted the 2024 Olympics in Paris.

All k candidates in one prompt

Include the context and every shortlisted candidate's index, name and description.

Example completion

```
<think>
... Candidate 0, Paris (city), matches the Olympics in France. The novel and the city in Texas do not. ...
</think>

{"answer": 0}
```

Final link in this example: Paris $\longrightarrow$ **Paris (city)**.

Self-consistency

Sample k_{sc} reasoning and answer completions.

Extract the final `answer` index from each:

$$i_j \in \{0, \ldots, k-1\}.$$

Vote over these indices:

$$\hat{i} = \text{mode}(i_1, \ldots, i_{k_{\text{sc}}})$$

We use $k_{\text{sc}} = 10$.

Illustrative completion with Paris (city) indexed as 0. A None option can also be added.

When selection corrects the reranker

> The game's plot revolves around Ogel making Evil Orbs that turn regular townspeople into mindless zombies, and rescuing captured Alpha Team members ...

The reranker's two highest-ranked candidates

- **Mind Control Orb**: the object used to turn townspeople into zombies.
- **Skeleton Drone**: the minions that the townspeople become.

ZESHEL/LEGO example. Context shortened; candidate descriptions paraphrased; two of ten candidates shown.

When selection corrects the reranker

The game's plot revolves around Ogel making Evil Orbs that turn regular townspeople into mindless zombies, and rescuing captured Alpha Team members . . .

The reranker's two highest-ranked candidates

- **Mind Control Orb**: the object used to turn townspeople into zombies.
- **Skeleton Drone**: the minions that the townspeople become.

Reranker's choice: **Mind Control Orb** (incorrect)

ZESHEL/LEGO example. Context shortened; candidate descriptions paraphrased; two of ten candidates shown.

When selection corrects the reranker

> The game's plot revolves around Ogel making Evil Orbs that turn regular townspeople into mindless zombies, and rescuing captured Alpha Team members ...

The reranker's two highest-ranked candidates

- **Mind Control Orb**: the object used to turn townspeople into zombies.
- **Skeleton Drone**: the minions that the townspeople become.

Reranker's choice: **Mind Control Orb** (incorrect)

Selection LLM's choice: **Skeleton Drone** (correct)

Selection LLM: reasoning excerpt

"Candidate 1 (Mind Control Orb): This is the **tool used to create the zombies**. Candidate 2 (Skeleton Drone): This is **what the townspeople become** (the minions). [...]
Later in the text, it says: 'The zombies were renamed to skeleton drones' [...]
So, 'zombies' refers to the minions. Candidate 2 is the most accurate description of these minions."

ZESHEL/LEGO example. Context shortened; candidate descriptions paraphrased; two of ten candidates shown.

How do we test transfer to new domains?

Benchmark	Challenge	Metric
ZESHEL	Transfer across domains	Normalized accuracy
GLADIS	Acronym disambiguation	Accuracy
ESCO	Occupational skills + None-linking	Accuracy
ZELDA	Wikipedia-based entity linking	Accuracy

Two comparison groups

- True zero-shot competitors.
- Approaches using benchmark-specific or supervised EL training.

Shared settings: top $k = 10$ candidates; $k_{sc} = 10$ sampled answers.

Models: `Qwen3-Reranker-4B` for reranking; `Magistral-Small-2509` for selection.

LLM-SFT baseline: an LLM fine-tuned for candidate selection on Wikipedia.

ZESHEL normalized accuracy evaluates only mentions whose correct entity is present in the candidate set.

ZESHEL: 83.11% in true zero-shot EL

Macro-averaged normalized accuracy across four test domains.

Method	Setting	Accuracy (%)
BLINK	True zero-shot	74.27
LLM-SFT (our baseline)	True zero-shot	77.06
GenDecider	Zero-shot	82.75
LELA (Magistral)	**True zero-shot**	**83.11**

Best reported true zero-shot result

- +8.84 percentage points over the previous reported best.
- +6.05 points over our supervised transfer baseline.

Normalized accuracy: evaluated only when the correct entity is among the supplied candidates.
BLINK: Wu et al. (2020). GenDecider: Zhou et al. (2024).

GLADIS: acronym disambiguation across domains

Map an acronym to its intended long form using the context.

Method	General	Scientific	Biomedical	Macro
AcroBERT (fine-tuned)	76.1	27.6	**69.2**	57.7
AcroBERT (zero-shot)	72.0	27.4	61.4	53.6
LLM-SFT (true zero-shot)	59.6	17.4	31.4	36.1
LELA (Magistral)	**78.9**	**46.9**	61.0	**62.3**

Gains in macro accuracy

- +26.2 percentage points over LLM-SFT.
- +4.6 points over fine-tuned AcroBERT.

Scores: accuracy (%). LLM-SFT is fine-tuned for candidate selection on Wikipedia and transferred to GLADIS.
GLADIS and AcroBERT: Chen et al. (2023).

Reasoning-based selection makes the difference

Controlled ablations on ZESHEL using the same `Qwen3-30B-A3B` selection model.

Variant	Accuracy (%)	Change (pp)
Full LELA	**76.79**	—
Without reasoning	54.23	−22.56
Without self-consistency	71.03	−5.76
Without descriptions	65.13	−11.66
Without reranking	72.35	−4.44
Without selection LLM	27.15	−49.64

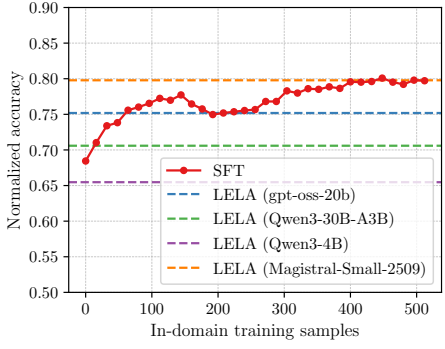
Reasoning contributes beyond joint candidate comparison

Disabling reasoning while keeping the same selection model reduces accuracy by 22.56 percentage points.

Metric: macro-averaged normalized accuracy. Qwen supports switching between reasoning/non-reasoning modes.

What is zero annotation worth?

Over 400 annotations for the tested supervised baseline to match LELA (Magistral).



LELA uses **zero target-domain training examples**.

Star Trek domain of ZESHEL. The supervised baseline is first trained on Wikipedia, then fine-tuned on increasing amounts of in-domain data.

Distilling reasoning into smaller models

Generate teacher reasoning from 10,000 Wikipedia mentions, retain correct outputs, and fine-tune smaller students.

Model	Configuration	Accuracy (%)
Qwen3-4B	No distillation	68.00
gpt-oss-20b	Teacher	78.30
Qwen3-0.6B	Distilled student	72.78
Qwen3-1.7B	Distilled student	75.15
Qwen3-4B	**Distilled student**	**78.61**

The 4B student matches the teacher

+10.61 percentage points over the base `Qwen3-4B`, with no target-domain EL annotations.

Metric: ZESHEL macro-averaged normalized accuracy. Students are fine-tuned on out-of-domain Wikipedia data. Evaluation remains true zero-shot on the target domains.

When is LELA the right choice?

Best fit

- New domains and proprietary KBs.
- Annotation requires scarce expert time.
- Offline or batch processing.

Trade-offs

- Higher inference cost.
- Strongest supervised methods remain ahead in-domain.
- None-linking remains difficult.

The practical trade-off

Spend inference compute when expert annotation is the scarcer resource.

Runtime depends on the model and experimental configuration.

LELA: take-home messages

LELA is an **entity linking system**.

Works with **any knowledge base** (YAGO by default).

1. **Modular architecture**
   Compatible with different retrievers, rerankers and LLMs.
2. **No target-domain training or fine-tuning**
   No annotated mention–entity pairs from the target domain.
3. **Strong empirical performance**
   Best reported results on ZESHEL, ESCO and GLADIS.

Main takeaway

Inference-time scaling can compensate for the absence of domain-specific supervision.

Paper

arxiv.org/abs/2601.05192

Code

github.com/dig-team/LELA

Presenter's website

samyhaff.github.io

YAGO ecosystem

yago-knowledge.org/ecosystem