TECHWOLF

JobBERT code-along session
TechWolf @ DataCamp

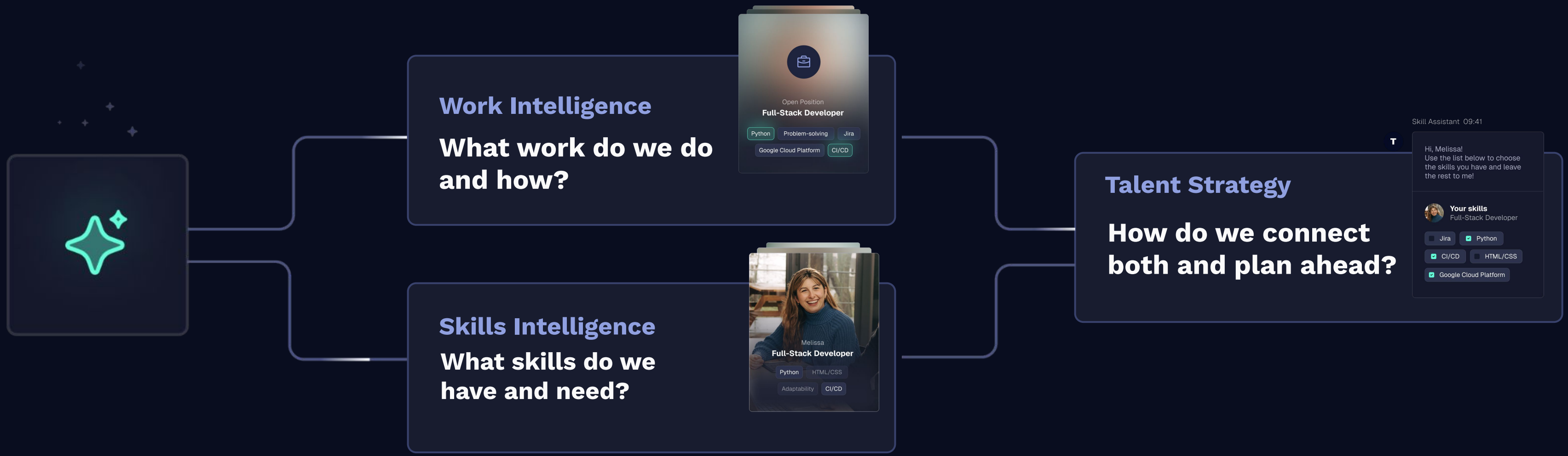
Jens-Joris Decorte, PhD

AI @ TechWolf

The #1 partner for workforce transformation

TechWolf is a **foundational data layer for work intelligence.**

We help our enterprises define and execute their talent strategy in the age of AI.



TechWolf at a glance

LEADING AI RESEARCH

University of Cambridge
Stanford University
Imperial College London
Ghent University
Vlaams Supercomputer Centrum

OUR CUSTOMERS

HSBC
Synopsys
IQVIA
Engie
Signify
KLM
bp
United
Workday
Garrett Advancing Motion
UCB
T Mobile
MetLife
Booking.com
VELUX
GSK
Red Hat
Atlas Copco
Ericsson
Just Eat

STRATEGIC PARTNERS

crunchr
visier
SAP
workday

INVESTORS

SAP
workday
servicenow

RECOGNIZED BY

World Economic Forum

PUBLISHING OVER PATENTING

Open-source

1.5 million downloads

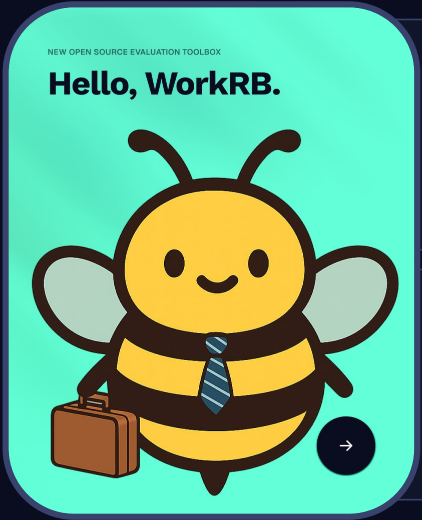
of our open-source datasets and models.

PUBLISHING OVER PATENTING

Open-source

1.5 million downloads

of our open-source datasets and models.



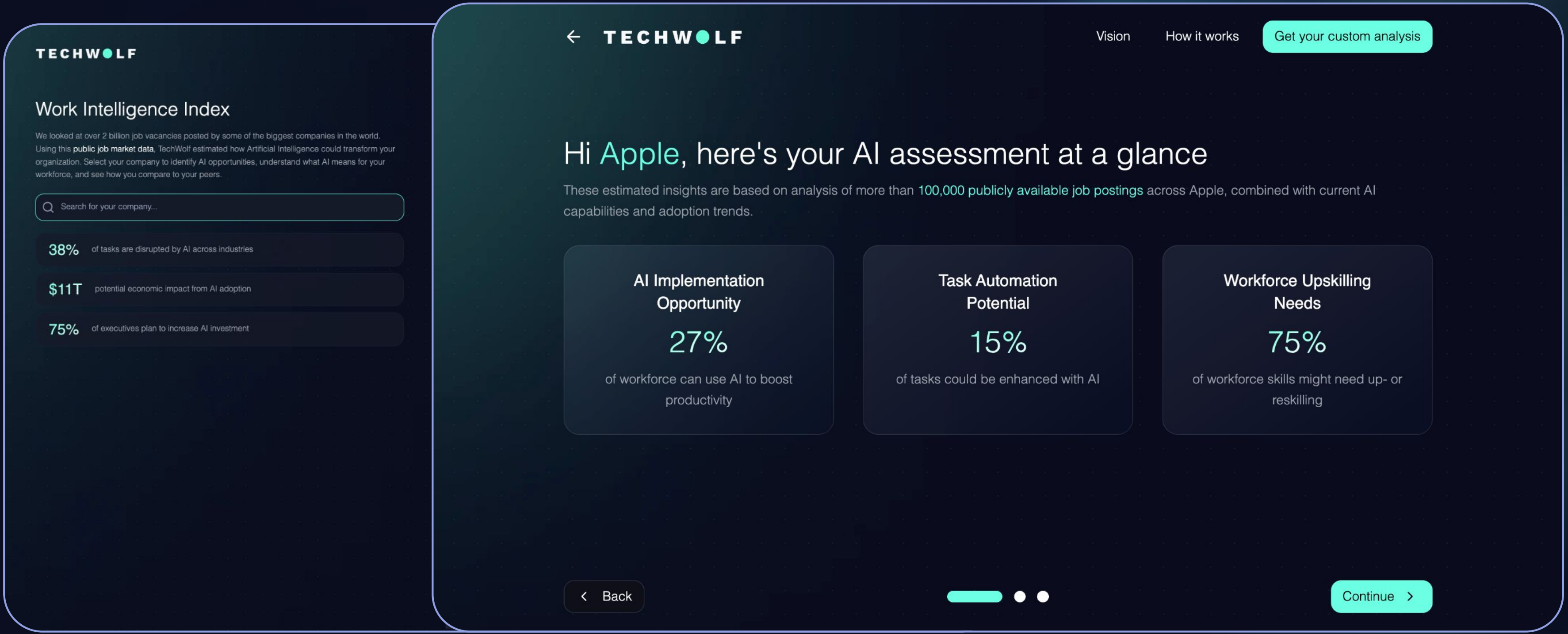
10+ tasks

and growing 👀

34 ⭐ stars

on our github repository.

Work Intelligence Index



Context for today: **Labor Market Analysis**

- Which skills are being asked for in the market?
- For which jobs?
- And how does that change over time and regions?

Labor Market Analysis

- Which skills are being asked for in the market?

  Skill Extraction

- For which jobs?

  Job Title Normalization

- And how does that change over time and regions?

  On large amounts of data (<u>efficiency</u>)

Technical resource: *Efficient Text Encoders for Labor Market Analysis; Decorte at al. 2025*

Efficient Text Encoders for Labor Market Analysis

JENS-JORIS DECORTE[1,2], JEROEN VAN HAUTTE[1], CHRIS DEVELDER[2], (Senior Member, IEEE), AND THOMAS DEMEESTER[2]
[1]TechWolf, 9000 Ghent, Belgium
[2]Internet and Data Science Laboratory (IDLab), Ghent University–imec, 9000 Ghent, Belgium

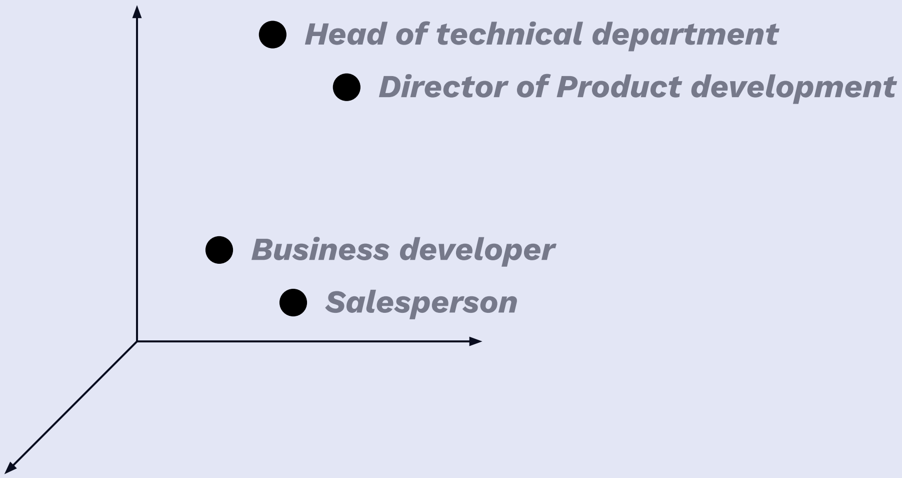
Corresponding author: Jens-Joris Decorte (jensjoris@techwolf.ai)

This work was supported in part by the Flemish Government, through Flanders Innovation and Entrepreneurship (VLAIO) under Project HBC.2020.2893; in part by the ‘‘Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen’’ Program; and in part by TechWolf.

ABSTRACT Labor market analysis relies on extracting insights from job advertisements, which provide valuable yet unstructured information on job titles and corresponding skill requirements. While state-of-the-art methods for skill extraction achieve strong performance, they depend on large language models (LLMs), which are computationally expensive and slow. In this paper, we propose ConTeXT-match, a novel contrastive learning approach with token-level attention that is well-suited for the extreme multi-label classification task of skill classification. ConTeXT-match significantly improves skill extraction efficiency and performance, achieving state-of-the-art results with a lightweight bi-encoder model. To support robust evaluation, we introduce Skill-XL a new benchmark with exhaustive, sentence-level skill annotations that explicitly address the redundancy in the large label space. Finally, we present JobBERT V2, an improved job title normalization model that leverages extracted skills to produce high-quality job title representations. Experiments demonstrate that our models are efficient, accurate, and scalable, making them ideal for large-scale, real-time labor market analysis.

Semantic Job Title Search

Semantic Job Title Search with embeddings



TECHWOLF

Job Titles Are **Short**

Responsibilities include creating forecasting models, assessing risk in investments and ensuring all accounting activities comply with regulations. To be successful in this role, you should have experience crafting financial strategies and managing accounting teams. Ultimately, you will maintain our company’s financial health and increase profitability in the long run ...

Head of Finance

- Few words
- Non-descriptive
- Rich meaning

Job Titles Are **Highly Contextual**

Single words are ambiguous. What do these words mean?

- Consultant, Engineer, Analyst, Manager…

An **API <u>developer</u>** is more related to a **software engineer** than it is to **business <u>developer</u>**.

Job Titles Are **Wild**

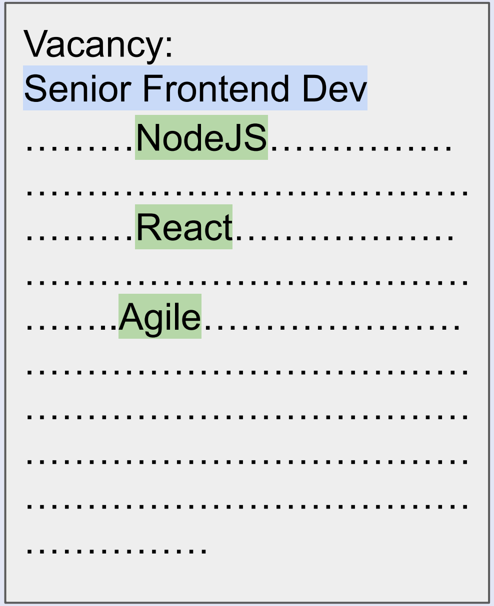
Elon Musk changes job title to 'Technoking of Tesla'

JobBERT: Understanding Job Titles through Skills

Main idea: learn what job titles **mean.**

Definition of “meaning” ⇒ **skills**

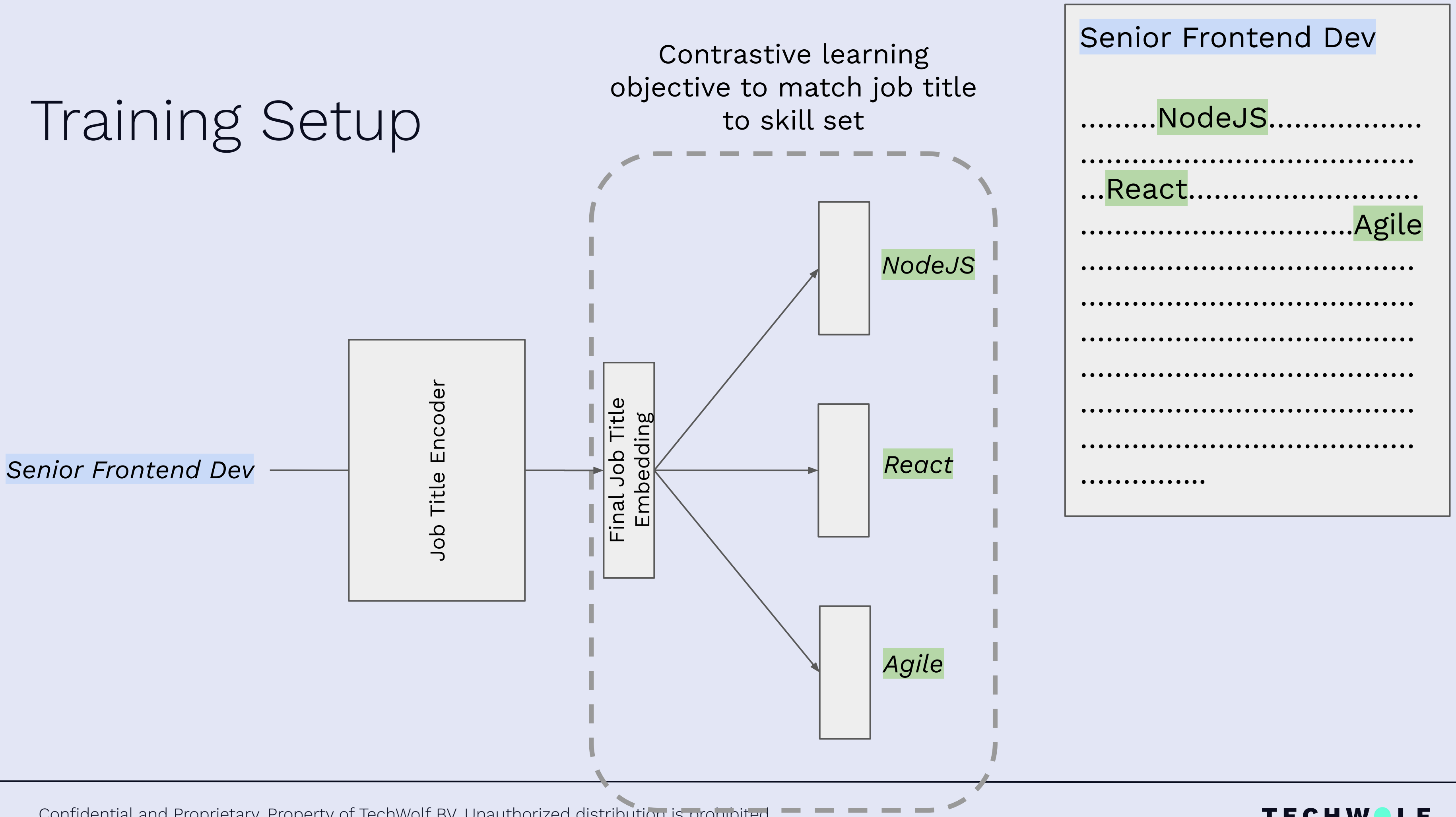
Dataset: **vacancies**



TECHWOLF

Training Setup

Contrastive learning objective to match job title to skill set



TECHWOLF

JobBERT V2

Improved training **data quality**, simplified **training method**, **SoTA results**

- Skills extracted with the ConTeXT-match skill extraction model
- JobBERT v2 is trained on 5.6M job ads (instead of 300M)
- Simple contrastive learning between job title and full comma-separated list of skills

TABLE 3. Performance comparison on the original JobBERT job title normalization benchmark. Metrics include Mean Reciprocal Rank (MRR) and Recall at Top-K (RP@5, RP@10). Results of previous methods are the reported scores. The recall is reported in percentage points.

Model	MRR	R@5	R@10
JobBERT V1 [7]	0.309	38.65	46.04
Doc2VecSkill [18]	0.341	45.95	54.00
JD Aggregation Network [20]	0.387	49.24	57.22
JobBERT V2 (Ours)	**0.390**	**50.08**	**58.47**

Skill Extraction

Skill Extraction: contextual phrase normalization

motivated

flexible

"You are hard working and able to adapt to different circumstances. Familiarity with Excel is required."

Microsoft Excel

TECHWOLF

Skill Extraction

2010-2020

- Traditionally relies on Named Entity Recognition and rule-based matching, detecting skill spans without normalizing them.
- Modern normalizing methods rely on:
  - one-vs-many classification or
  - bi-encoder based ranking systems.
- Highest accuracy: LLM-based multi-step systems.

Skill Extraction

- Traditionally relies on Named Entity Recognition and rule-based matching, detecting skill spans without normalizing them.
- Modern normalizing methods rely on:
  - one-vs-many classification or
  - bi-encoder based ranking systems.

2020-2022

- Highest accuracy: LLM-based multi-step systems.

Skill Extraction

- Traditionally relies on Named Entity Recognition and rule-based matching, detecting skill spans without normalizing them.
- Modern normalizing methods rely on:
  - one-vs-many classification or
  - bi-encoder based ranking systems.
- Highest accuracy: LLM-based multi-step systems.

2022-2024

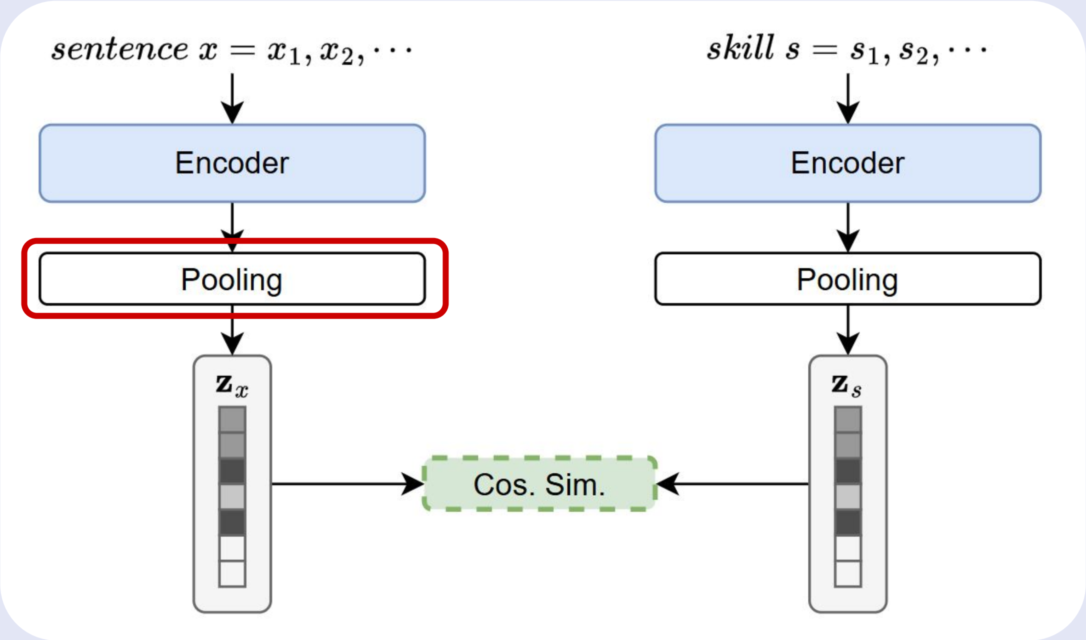
TECHWOLF

Skill Extraction

- Traditionally relies on Named Entity Recognition and rule-based matching, detecting skill spans without normalizing them.
- Modern normalizing methods rely on:
  - one-vs-many classification or
  - bi-encoder based ranking systems.
- Highest accuracy: LLM-based multi-step systems.

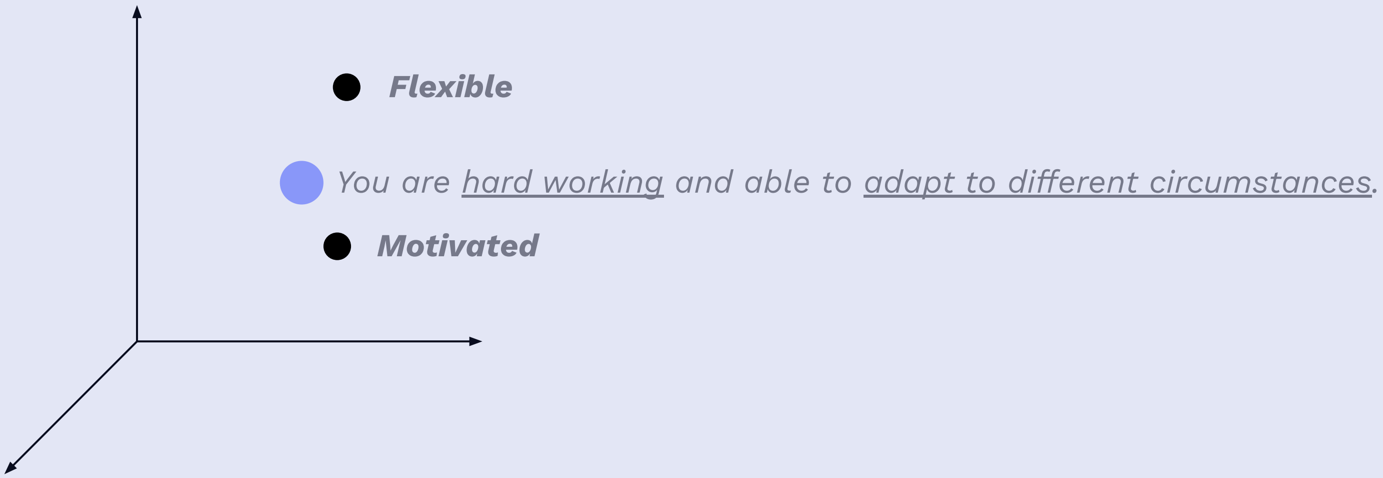
Best performance / efficiency trade-off: bi-encoder based ranking systems.

Bi-encoder information bottleneck

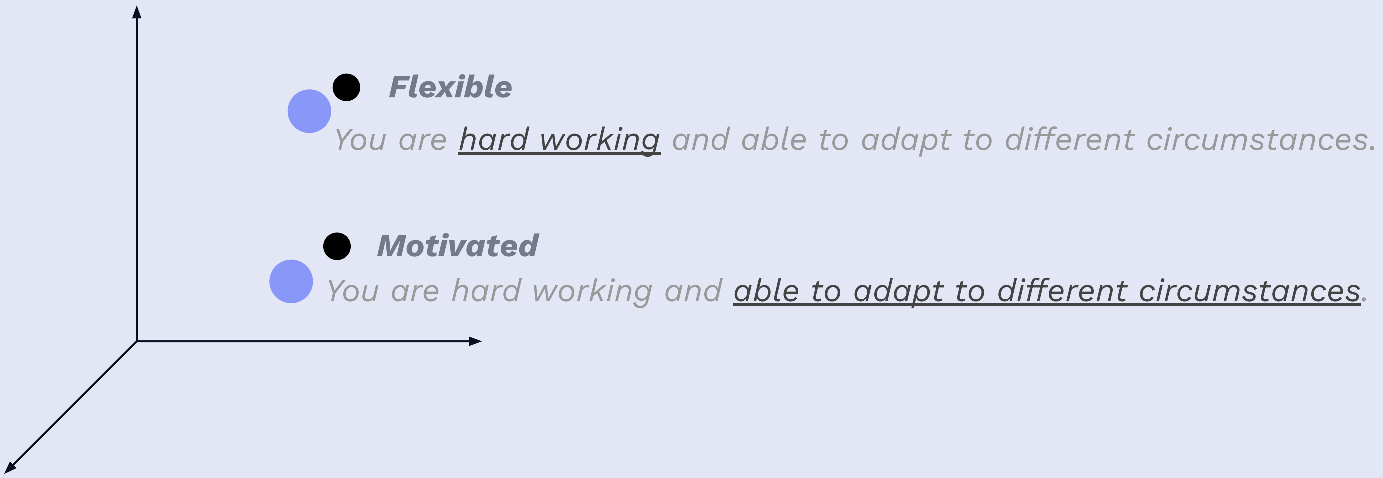


TECHWOLF

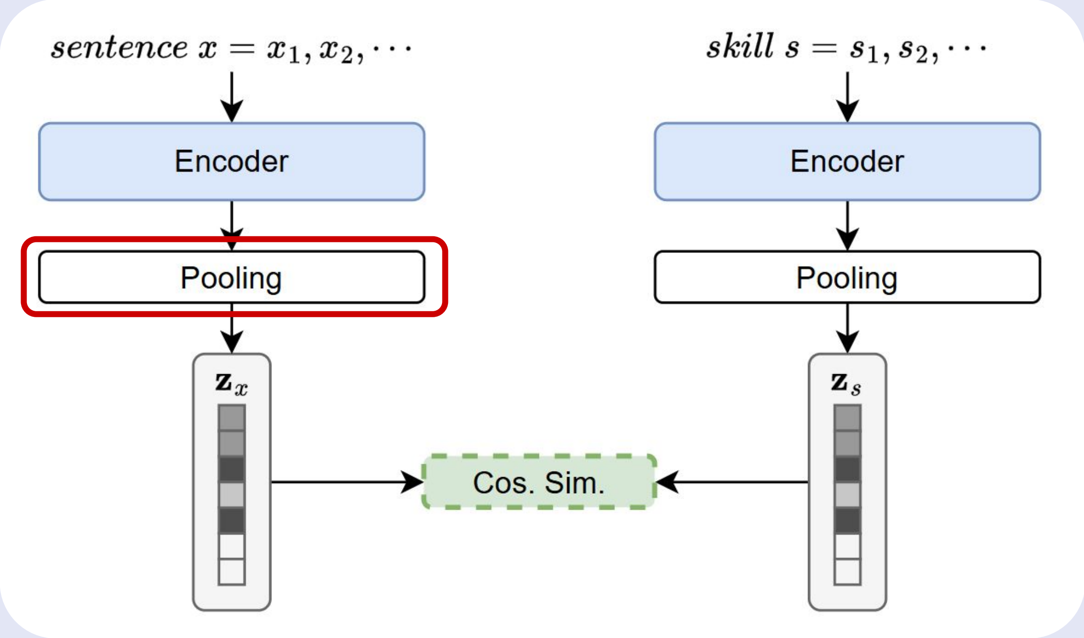
Bi-encoder information bottleneck



Bi-encoder information bottleneck



Bi-encoder information bottleneck



Averaging over entire sentences overlooks valuable token-level information.

Goal: adapt training to match skills with a dynamic range of tokens.

Hypothesis: improved recall and performance.

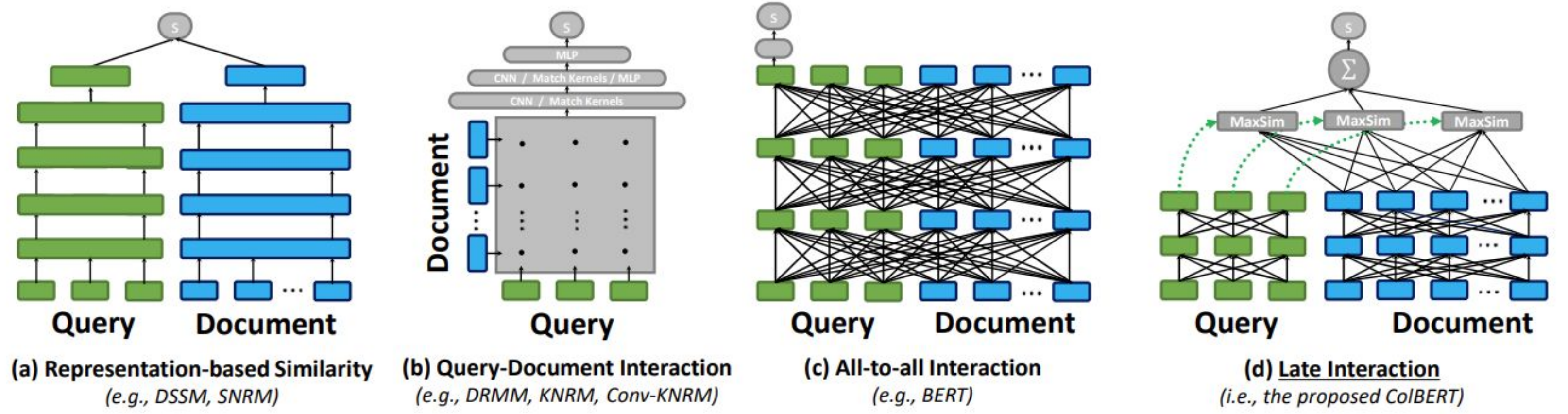
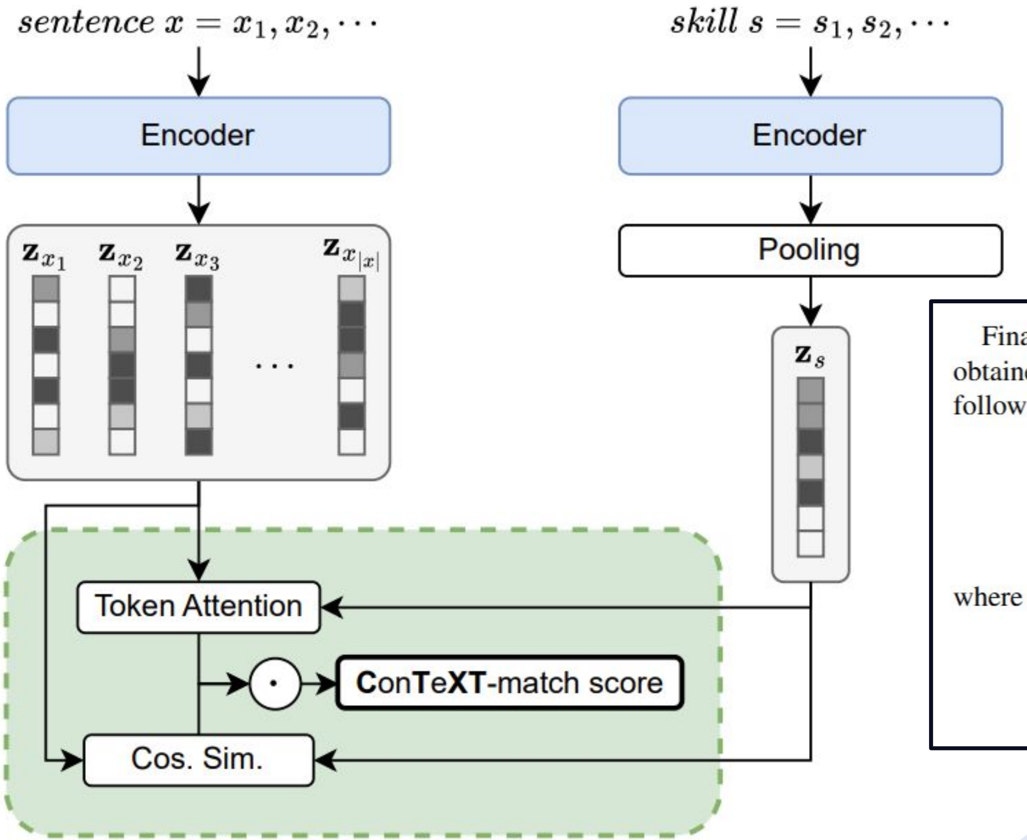


Figure 2: Schematic diagrams illustrating query–document matching paradigms in neural IR. The figure contrasts existing approaches (sub-figures (a), (b), and (c)) with the proposed late interaction paradigm (sub-figure (d)).

Omar Khattab and Matei Zaharia, 2020

Contrastive Token-level Explainable Text matching



ConTeXT-match

Finally, the match between the sentence and the skill is obtained as the weighted average of the token similarities, as follows:

$$\mathrm{match}(x, s) = \sum_{j=1}^{|x|} \alpha_j \cdot \cos(x_j, s), \tag{3}$$

where the weights α_j sum to one, as defined by

$$\alpha_j = \frac{\exp\left(\mathbf{z}_{x_j} \cdot \mathbf{z}_s\right)}{\sum_{k=1}^{|x|} \exp\left(\mathbf{z}_{x_k} \cdot \mathbf{z}_s\right)}. \tag{4}$$

Results

TABLE 2. **Performance comparison of skill extraction methods across the SkillSpan-ESCO test sets (HOUSE, TECH, TECHWOLF). Metrics include Mean Reciprocal Rank (MRR) and Recall Precision at Top-K (RP@1, RP@5, RP@10). Following previous work, the RP@K scores are reported in percentage points. Average positive training examples per skill are reported in the first column (N). Results of previous methods are the reported scores, and empty cells mean the metric was not reported. The strongest results per metric are shown in bold, and the second strongest are underlined. Whenever our results are significantly ($p < 0.05$) stronger than the second-best method, we indicate this with *.**

		HOUSE				TECH				TECHWOLF			
	N	MRR	RP@1	RP@5	RP@10	MRR	RP@1	RP@5	RP@10	MRR	RP@1	RP@5	RP@10
Encoder classifiers													
Decorte et al. [12]	365	0.299	–	30.82	38.69	0.339	–	31.71	39.19	–	–	–	–
Clavié et al. [4]	40	0.326	27.20	37.60	46.47	0.299	27.16	33.41	39.86	–	–	–	–
Encoder rankers													
Clavié et al. [4]	40	0.355	26.44	35.22	43.73	0.405	32.84	49.67	58.66	–	–	–	–
Decorte et al. [14]	**10**	0.426	27.10	45.94	53.87	0.521	38.46	54.19	61.52	0.506	37.42	52.64	60.10
ConTeXT-match (Ours)	**10**	**0.530***	38.42	51.09	65.84	**0.632***	**50.99***	**63.98**	**73.99***	**0.562***	**43.15***	**57.69***	**66.08**
LLM-based systems													
Clavié et al. [4] GPT3.5	40	0.427	36.92	43.57	51.44	0.488	40.53	52.50	59.75	–	–	–	–
Clavié et al. [4] GPT4	40	0.495	**40.70**	53.34	61.02	0.537	46.52	61.50	68.94	–	–	–	
IReRa [5]	n/a	–	–	**56.50**	**66.51**	–	–	59.61	70.23	–	–	57.04	65.17

TECHWOLF

(a) Lead the group in charge of cost and risk management objectives

cost management										
lead	the	group	in	charge	of	cost	and	risk	management	objectives
lead a team										
lead	the	group	in	charge	of	cost	and	risk	management	objectives
risk management										
lead	the	group	in	charge	of	cost	and	risk	management	objectives

(c) Responsible for diagnosing, repairing, and maintaining cars

diagnose problems with vehicles										
responsible	for	dia	gno	sing	,	repairing	,	and	maintaining	cars
carry out repair of vehicles										
responsible	for	dia	gno	sing	,	repairing	,	and	maintaining	cars
maintain vehicle service										
responsible	for	dia	gno	sing	,	repairing	,	and	maintaining	cars

(b) You will write software in Java, Python and C++

C++										
you	will	write	software	in	java	,	python	and	c	+	+
authoring software											
you	will	write	software	in	java	,	python	and	c	+	+
Java (computer programming)											
you	will	write	software	in	java	,	python	and	c	+	+
Python (computer programming)											
you	will	write	software	in	java	,	python	and	c	+	+

Let’s put this into practice