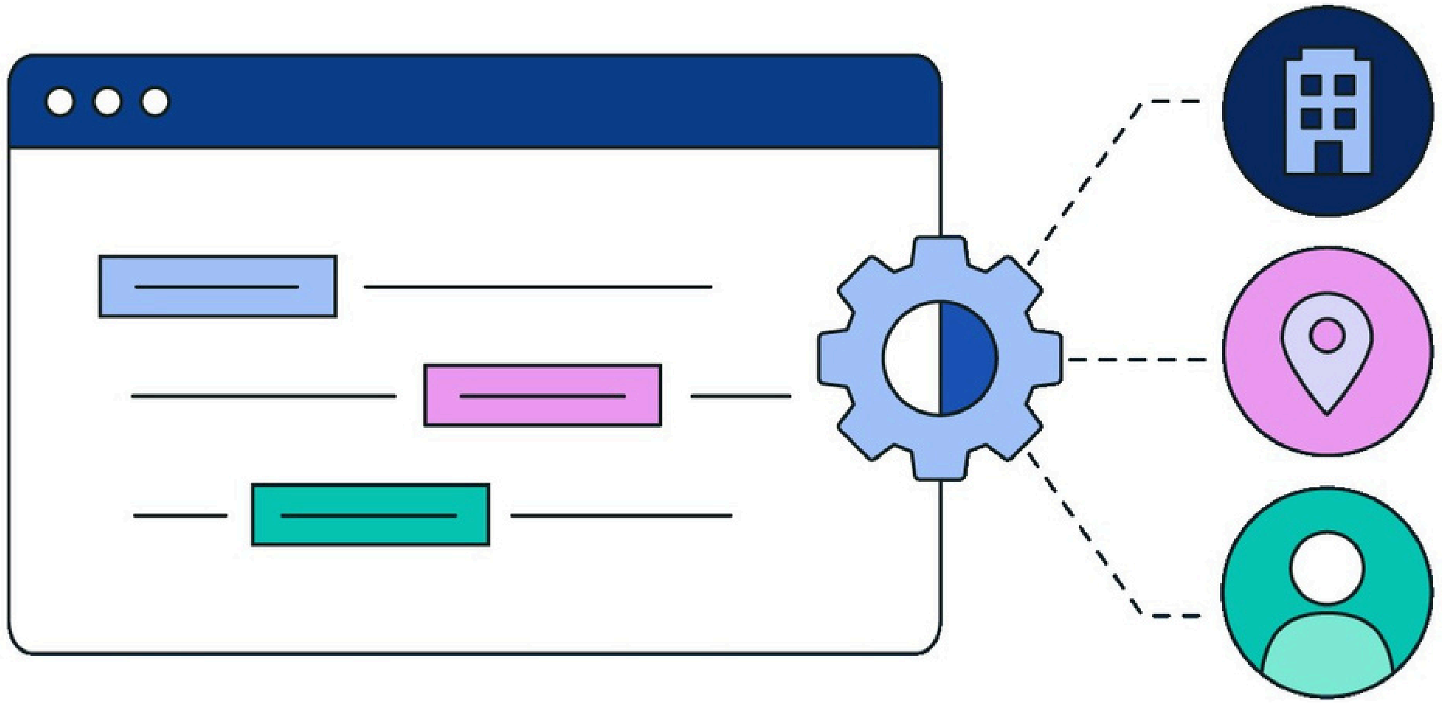
Improving Named Entity Recognition for Low-Resource Languages Using Large Language Models: A Ukrainian Case Study



https://medium.com/@jkshj21/validating-entities-in-dialogflow-cx-part-2-entities-in-pages-6bf0d51a7bbe

Vladyslav Radchenko
AppS, Ukrainian Catholic University

Nazarii Drushchak
AppS, Ukrainian Catholic University; SoftServe Inc.

Lviv – Ukraine | 31 July, 2025

Background: NER Basics

Named Entity Recognition (NER) is a subfield of NLP that focuses on identifying and classifying entities like people, organizations, and locations within unstructured text data

Київ LOC — столиця України LOC . Володимир Зеленський PER є президентом України LOC .

Background: Applications of NER

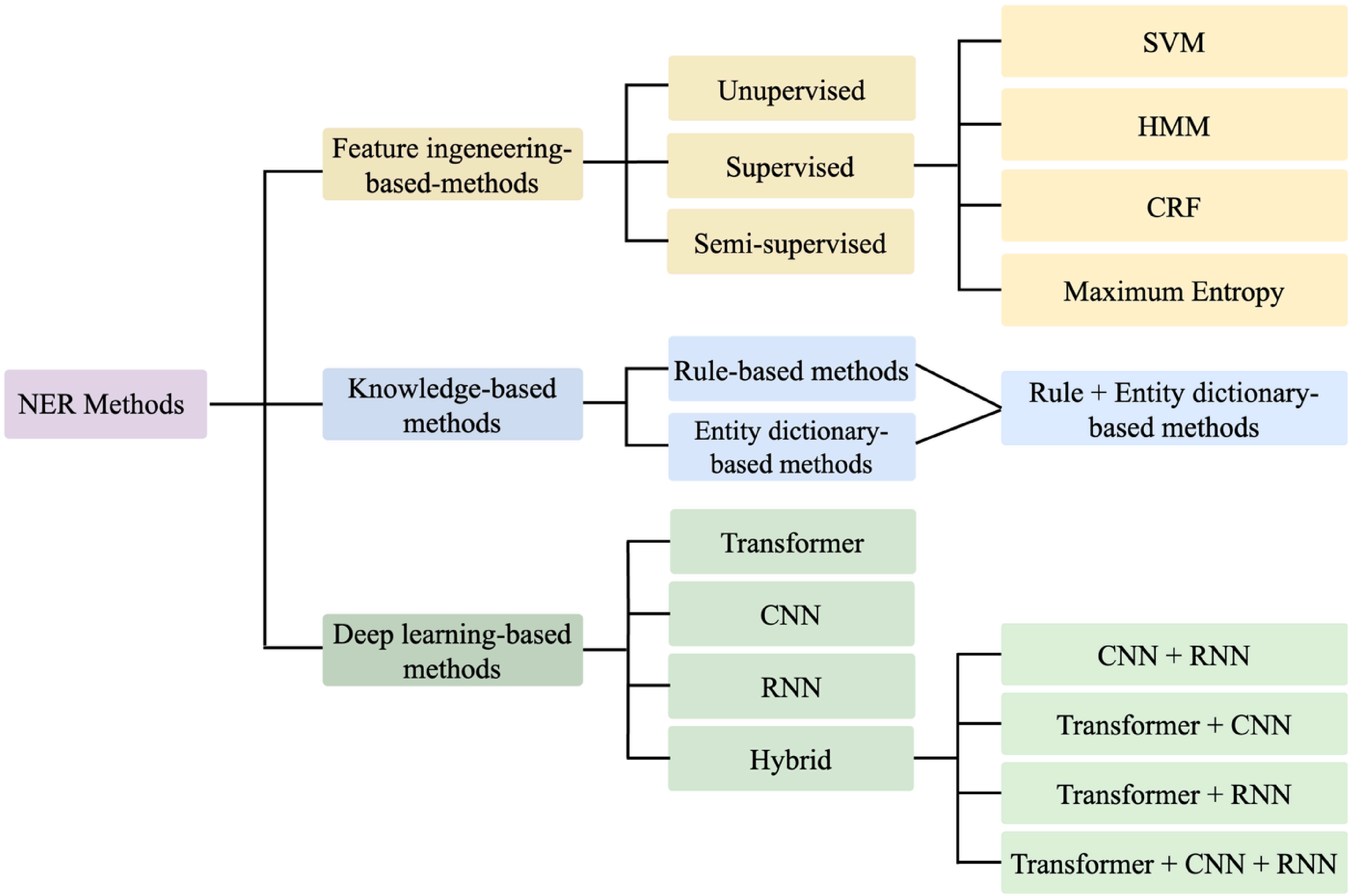
Data anonymization

Information extraction

Social media monitoring

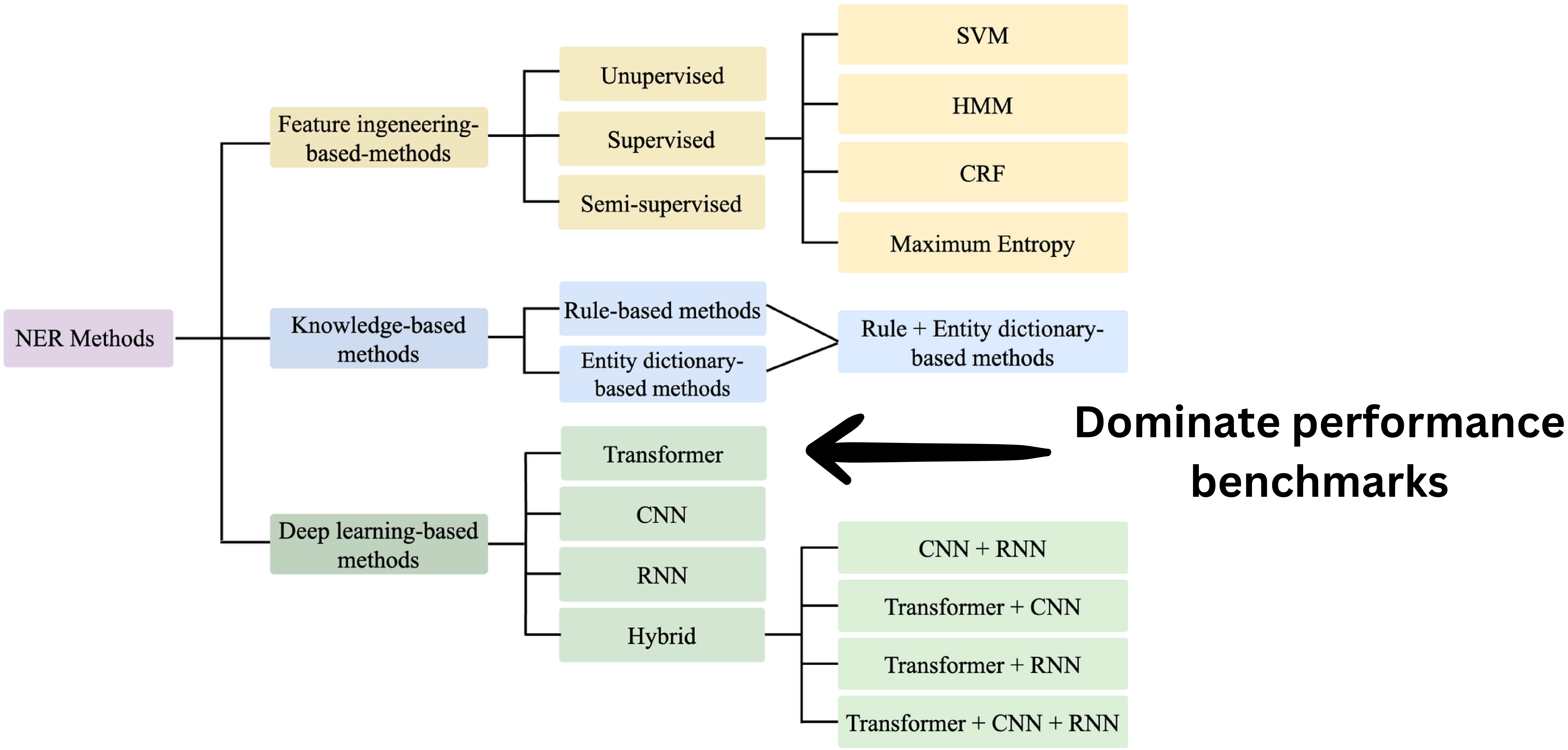
Image sources: stickers from Canva's Open Sticker Library.

Background: Main Approaches to NER



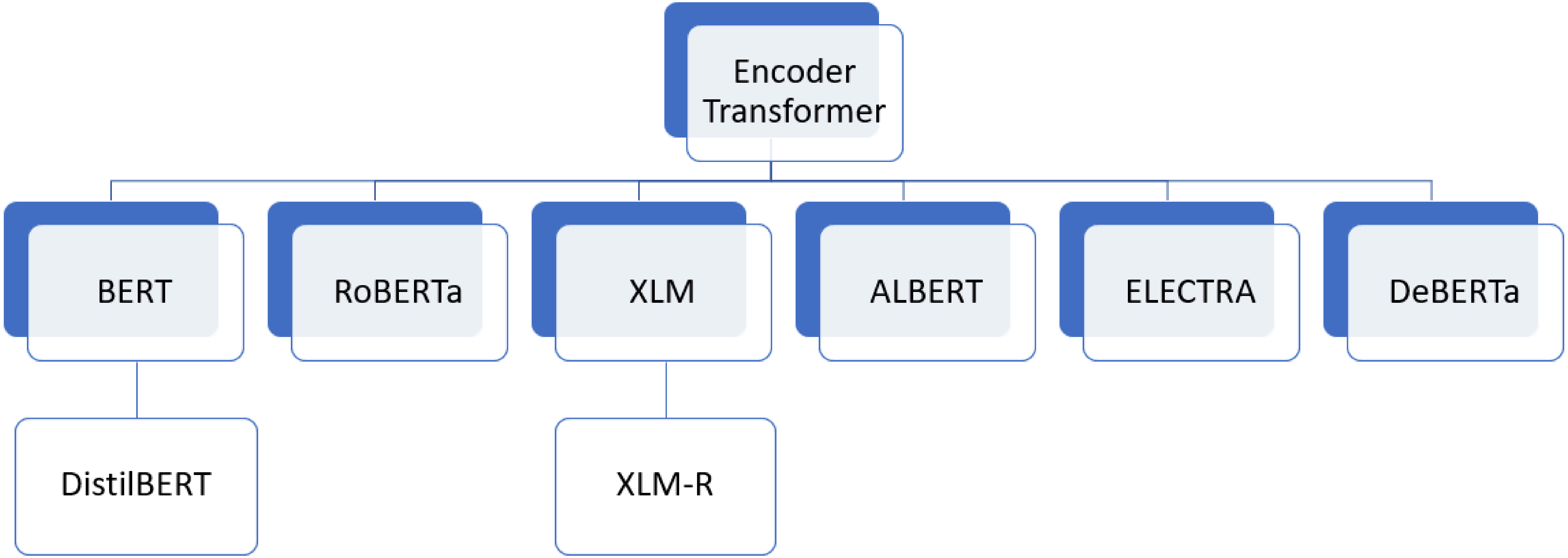
https://arxiv.org/html/2401.10825v3#S6

Background: Main Approaches to NER



https://arxiv.org/html/2401.10825v3#S6

Background: Transformers



https://vitalflux.com/encoder-only-transformer-models-examples/

Background: Challenges in Ukrainian NER

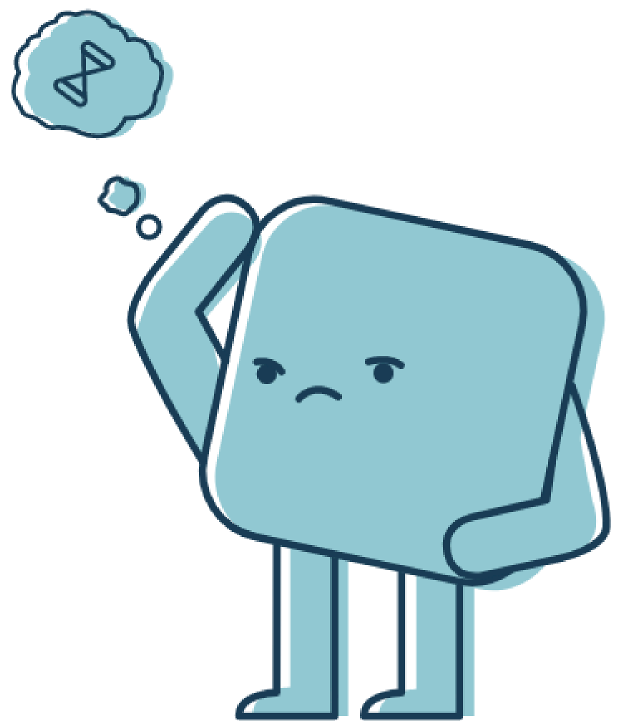
Data scarcity

Image source: Visual generated by ChatGPT

Background: Challenges in Ukrainian NER

Data scarcity

Image source: Visual generated by ChatGPT



Complex linguistic structure

Image source: sticker from Canva's Open Sticker Library.

Related Works: Large Language Models

- Pretrained on massive corpora → strong language understanding out of the box
- Task- and domain-agnostic → easily transferable across languages and applications
- Zero-/few-shot learning → minimizes need for costly annotated datasets

Related Works: Large Language Models

Comparable to fully supervised baselines, better in low-resource and few-shot setups

GPT-NER: Named Entity Recognition via Large Language Models

Shuhe Wang, Xiaofei Sun, Xiaoya Li, Rongbin Ouyang, Fei Wu, Tianwei Zhang, Jiwei Li, Guoyin Wang

https://arxiv.org/abs/2304.10428

State-of-the-art performance on few-shot NER, significant improvements on various datasets

PromptNER: Prompting For Named Entity Recognition

Dhananjay Ashok, Zachary C. Lipton

https://arxiv.org/abs/2305.15444

Research Gaps & Objectives

- Adapt large language models for NER in Ukrainian
- Benchmark open-source LLMs against proprietary models.
- Propose standardized evaluation pipeline for LLMs

Data: NER-UK 2.0 Overview

The largest publicly available manually annotated corpus for NER in Ukrainian

- Genres: News, legal documents, procurement contracts, social media.
- Entity Types: 13 categories.
- Corpus Size: 560 documents, ~22k labeled entities.

Table 1: Entity Type Distribution in NER-UK 2.0

ART	**DATE**	**DOC**	**JOB**	**LOC**	**MISC**	**MON**
635	2,047	142	1,982	3,000	515	943
ORG	**PCT**	**PERIOD**	**PERS**	**QUANT**	**TIME**	**Total**
5,213	263	596	6,235	382	40	**21,993**

Methodology: Experiments Set Up

- **Baseline Setting:** Fine-Tuning Encoder-Only Models.
- **Minimal supervision:**
    - LLM Prompting.
    - Generalist models Zero-Shot.
- **LLM Supervised Fine-tuning.**

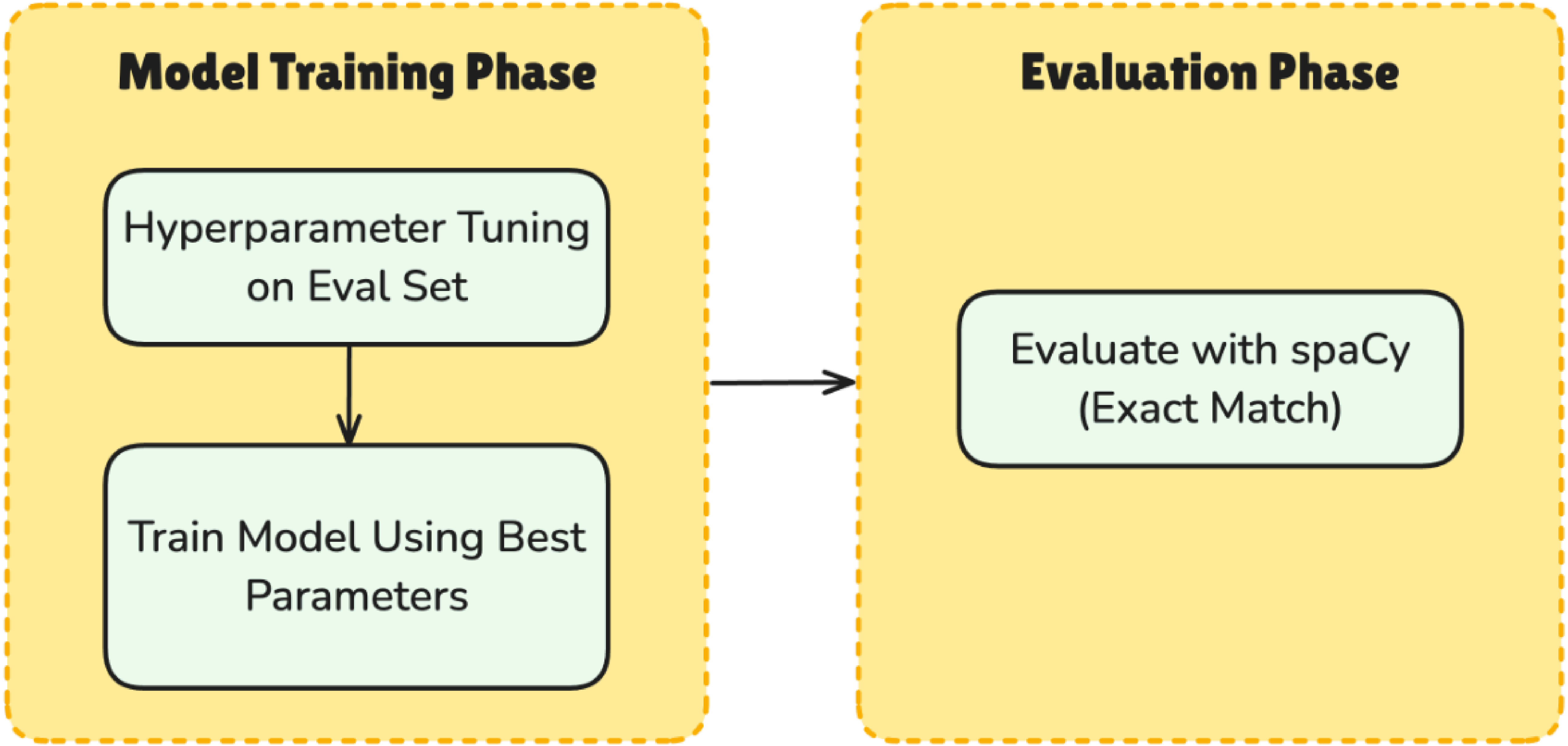
Methodology: Evaluation

- **Metric:** Weighed F1 Score
- **Strategy:** Exact Matching
- **Framework:** Spacy

Experiments: Model Selection

Model	Number of Parameters	Model Category
gpt-4o-2024-11-20	-	Proprietary LLM
Gemma-3-27B-IT	27.4B	Open-Source LLM
Gemma-2-27B-IT	27.2B	Open-Source LLM
Gemma-2-9B-IT	9.2B	Open-Source LLM
Phi-4	14.7B	Open-Source LLM
Qwen-2.5-14B-Instruct	14.8B	Open-Source LLM
Qwen-2.5-7B-Instruct	7.6B	Open-Source LLM
DeepSeek-R1-Distill-Qwen-14B	14.8B	Open-Source LLM
Gemma-2-2B-IT	2.6B	Open-Source LLM
Qwen-2.5-3B-Instruct	3.0B	Open-Source LLM
Llama-3.2-3B-Instruct	3.2B	Open-Source LLM
Phi-3-mini-4k-instruct	3.8B	Open-Source LLM
Llama-3.1-8B-Instruct	8.3B	Open-Source LLM
Aya-expanse-8b	8.0B	Open-Source LLM
Aya-101	13.0B	Open-Source LLM
roberta-large-NER	561M	Encoder-only
xlm-roberta-large	561M	Encoder-only
NuNER-Zero	449M	Encoder-only
Modern-BERT-large	396M	Encoder-only
gliner-multi-v2.1	209M	Encoder-only
gliner-multi-pii-v1	209M	Encoder-only
uk-ner-web-trf-13class	110M	Encoder-only

Experiments: Baseline Setting

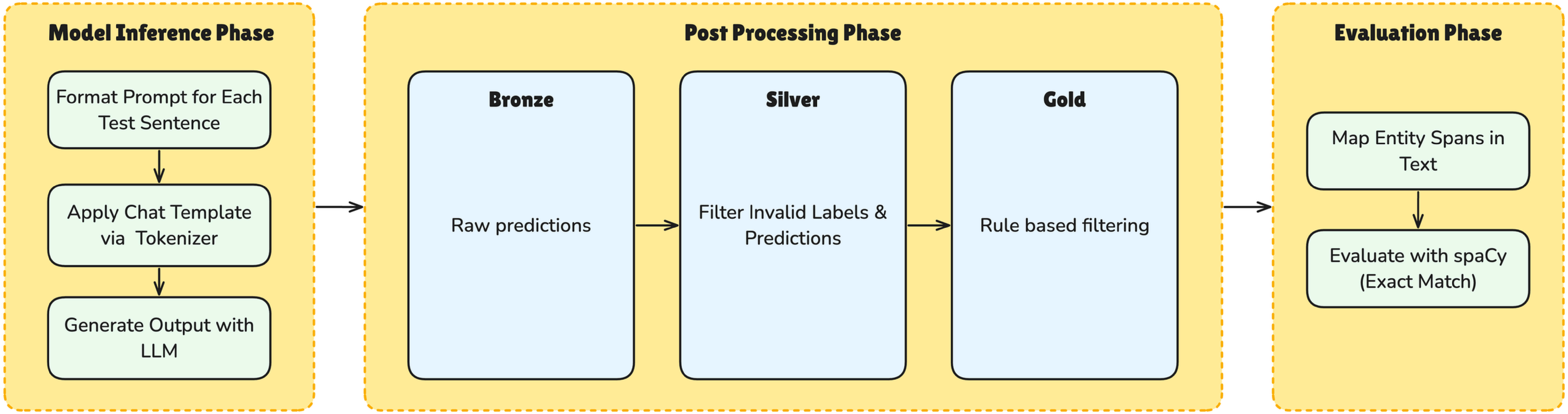


Experiments: Baseline Setting

Entity	roberta-large-NER	xlm-roberta-large	gliner-multi-v2.1	Modern-BERT-large	uk-ner-web-trf-13class
JOB	**0.699**	0.689	**0.699**	0.470	0.696
PERIOD	0.743	0.742	0.712	0.596	**0.769**
QUANT	0.915	**0.929**	0.819	0.803	0.860
DOC	0.561	0.556	0.456	0.271	**0.574**
LOC	0.916	**0.918**	0.880	0.720	0.899
DATE	0.895	0.896	0.881	0.839	**0.908**
ORG	0.916	0.913	0.875	0.791	**0.918**
PERS	**0.968**	**0.968**	0.951	0.862	0.967
TIME	0.500	0.609	0.471	0.000	**0.700**
MON	0.955	**0.960**	0.906	0.915	0.919
MISC	0.344	**0.386**	0.249	0.138	0.359
ART	0.737	**0.759**	0.639	0.508	0.757
PCT	**1.000**	0.989	0.961	0.977	0.973
Overall	**0.890**	0.889	0.855	0.762	0.887

TABLE 5.1: Entity-wise F_1 Scores for Encoder-Only Models

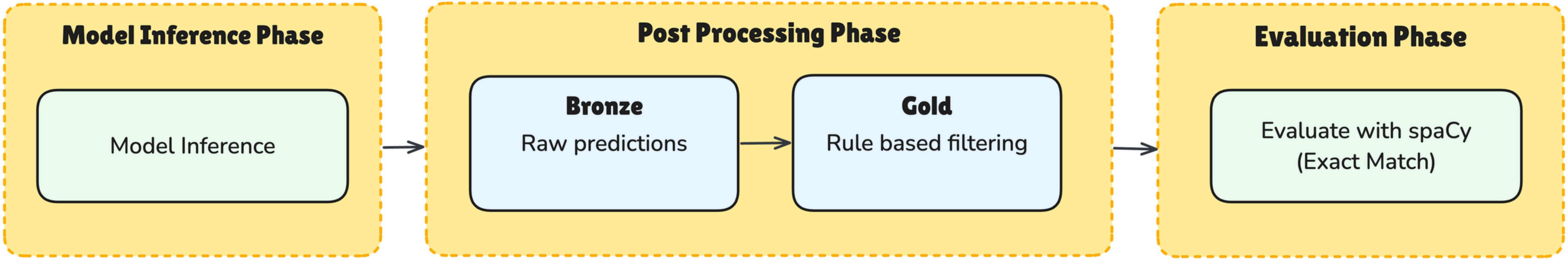
Experiments: LLM Prompting Pipeline



Experiments: LLM Prompting Pipeline

Model	Bronze			Silver			Gold		
	Zero-Shot	Few-Shot	CoT	Zero-Shot	Few-Shot	CoT	Zero-Shot	Few-Shot	CoT
GPT-4o	**0.67**	**0.71**	**0.60**	**0.68**	**0.71**	**0.61**	**0.72**	**0.71**	**0.68**
Gemma-3-27B-IT	0.39	0.67	0.40	0.41	0.69	0.43	0.56	**0.71**	0.58
Gemma-2-27B-IT	0.45	0.62	0.38	0.49	0.66	0.40	0.58	0.70	0.51
Gemma-2-9B-IT	0.42	0.49	0.42	0.46	0.54	0.47	0.55	0.62	0.60
Phi-4	0.38	0.48	0.36	0.43	0.53	0.41	0.52	0.61	0.51
Qwen-2.5-14B-Instruct	0.42	0.50	0.36	0.44	0.53	0.38	0.53	0.57	0.48
Qwen-2.5-7B-Instruct	0.34	0.36	0.30	0.36	0.38	0.33	0.45	0.45	0.44
DeepSeek-R1-Distill-Qwen-14B	0.34	0.11	0.35	0.36	0.13	0.38	0.42	0.13	0.46
Gemma-2-2B-IT	0.16	0.30	0.25	0.20	0.37	0.28	0.28	0.47	0.36
Qwen-2.5-3B-Instruct	0.18	0.33	0.20	0.22	0.37	0.23	0.28	0.45	0.30
Llama-3.2-3B-Instruct	0.17	0.28	0.13	0.24	0.41	0.23	0.30	0.45	0.25
Phi-3-mini-4k-instruct	0.16	0.27	0.19	0.19	0.32	0.24	0.23	0.39	0.29
Llama-3.1-8B-Instruct	0.14	0.23	0.14	0.18	0.29	0.18	0.25	0.37	0.23
Aya-expanse-8b	0.23	0.03	0.23	0.31	0.03	0.28	0.34	0.03	0.29
Aya-101	-	0.31	-	-	0.38	-	-	0.41	-

Experiments: Generalist Models Pipeline

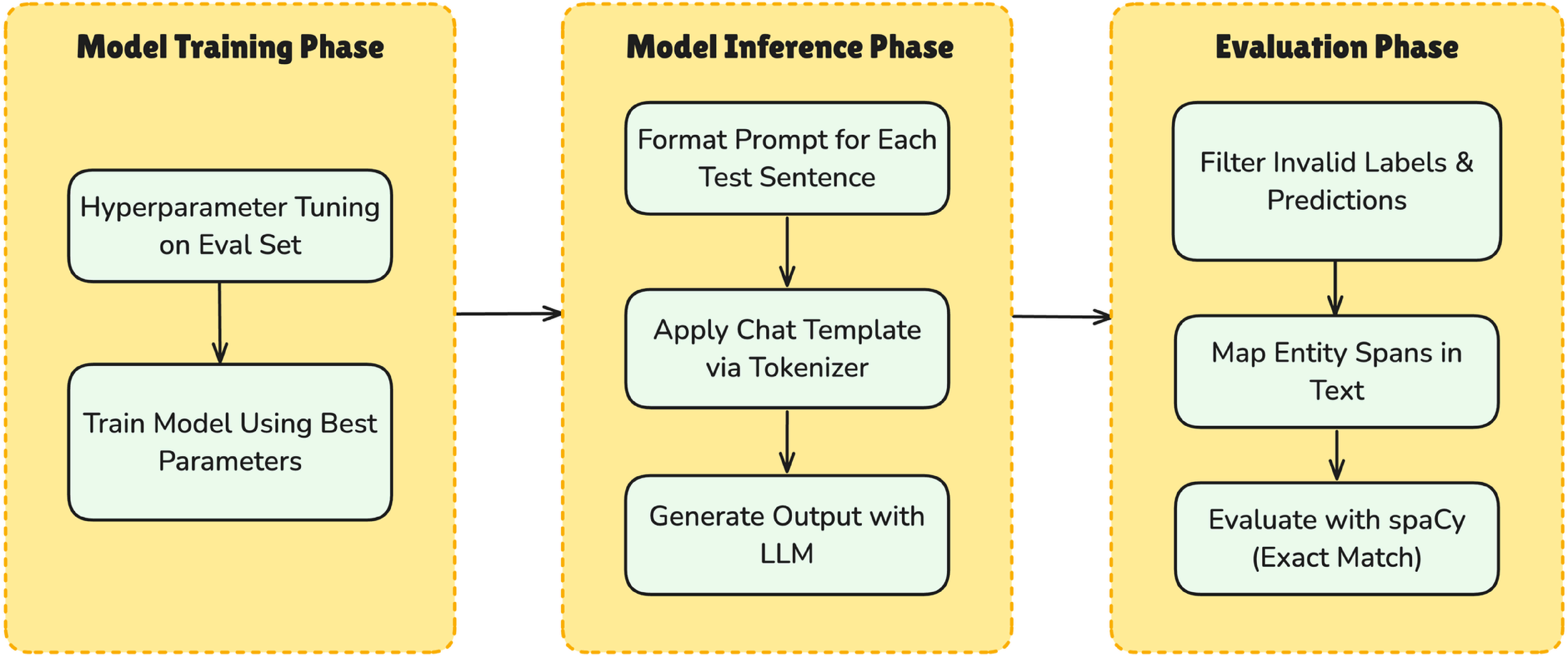


Experiments: Generalist Models Pipeline

Model	Bronze	Silver	Gold
gliner-multi-v2.1	**0.53**	Not applied	**0.67**
gliner-multi-pii-v1	0.46	Not applied	0.62
NuNER-Zero	0.41	Not applied	0.58

Table 5.3: Zero-Shot Performance of Generalist Models

Experiments: SFT Pipeline



Experiments: SFT Pipeline

Entity	Qwen2.5-14B-Instruct	Phi-4	Gemma-2-27B-IT	Gemma-3-27B-IT
JOB	0.624	0.638	**0.662**	0.642
PERIOD	0.667	0.714	0.742	**0.747**
QUANT	0.812	0.833	0.864	**0.897**
DOC	0.479	0.464	**0.537**	0.514
LOC	0.890	0.907	0.903	**0.929**
DATE	0.866	0.885	0.900	**0.906**
ORG	0.898	0.911	0.918	**0.923**
PERS	0.955	**0.967**	0.966	0.965
TIME	0.400	0.571	**0.824**	0.632
MON	0.950	0.958	**0.964**	0.953
MISC	**0.390**	0.314	0.311	0.350
ART	0.725	**0.774**	0.740	0.716
PCT	0.977	0.966	**0.994**	0.989
Overall	0.867	0.882	0.886	**0.888**

TABLE 5.5: Entity-wise F_1 Scores for Supervised Fine-Tuned LLMs

Conclusion

Entity	Tuning		Prompting		
	roberta-large-NER	**Gemma-3-27B-IT**	**GPT-4o**	**Gemma-3-27B-IT**	**GLiNER**
JOB	**0.699**	0.642	0.332	0.381	0.141
PERIOD	0.743	**0.747**	0.263	0.280	0.105
QUANT	**0.915**	0.897	0.475	0.000	0.155
DOC	**0.561**	0.514	0.122	0.000	0.111
LOC	0.916	**0.929**	0.775	0.782	0.705
DATE	0.895	**0.906**	0.650	0.738	0.663
ORG	0.916	**0.923**	0.809	0.757	0.672
PERS	**0.968**	0.965	0.900	0.870	0.863
TIME	0.500	**0.632**	0.308	0.111	0.154
MON	**0.955**	0.953	0.916	0.525	0.812
MISC	0.344	**0.350**	0.077	0.000	0.000
ART	**0.737**	0.716	0.289	0.000	0.175
PCT	**1.000**	0.989	0.910	0.949	0.867
Overall	**0.890**	0.888	0.724	0.713	0.669

Future Work

- Repurposed LLMs as text encoders through modifications like enabling bidirectional attention
- Explore Reinforcement Learning from Human Feedback techniques
- Expand and diversify annotated corpora

Many Thanks for Your Time

Happy to Take Your Questions

fb.com/csatucu

@ucu_apps

apps@ucu.edu.ua

apps.ucu.edu.ua

Faculty of Applied Sciences
Ukrainian Catholic University
Kozelnytska st. 2a, Lviv,
79076, Ukraine