GLiNER: Generalist Model for Named Entity Recognition using Bidirectional Transformer

Urchade Zaratiana, Nadi Tomeh, Pierre Holat, Thierry Charnois

FI Group, LIPN - Université Sorbonne Paris Nord - CNRS

Introduction

- GLiNER utilizes encoder-only bidirectional transformer (BERT-like) for open-type Named Entity Recognition (NER)
- Outperforms zero-shot and fine-tuned LLM models across diverse datasets
- Fewer parameters, faster, superior performance

Architecture

Input format The input to our model comprises a unified sequence combining entity types (expressed in natural language) and the input text from which entities are to be extracted. The input format is as follows:

[ENT] t_0 [ENT] t_1 … [ENT] t_{M-1} [SEP] x_0 x_2 … x_{N-1}

[ENT] token represents a special token placed before each entity type and the [SEP] token functions as a delimiter, separating the sequence of entity types from the input text. They are initialized randomly at the start of training.

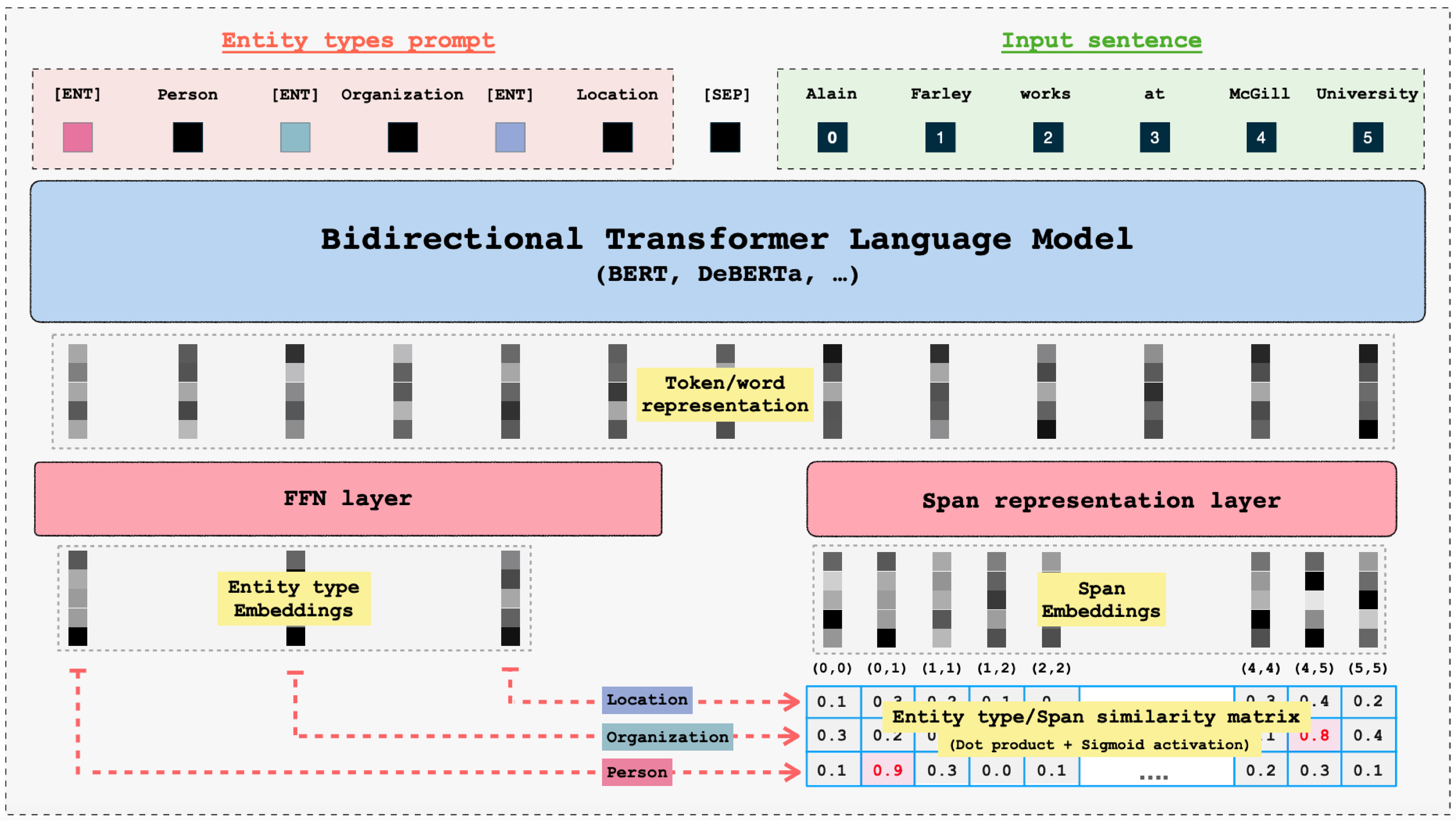
Token representation The token encoder processes the unified input to compute interactions between all tokens (both entity types and input text), producing contextualized representations. Let $\boldsymbol{p} = \{\boldsymbol{p}_i\}_0^{M-1} \in \mathbb{R}^{M \times D}$ represent the encoder's output for each entity type, corresponding to all the [ENT] token representations. Similarly, $\boldsymbol{h} = \{\boldsymbol{h}_i\}_0^{N-1} \in \mathbb{R}^{N \times D}$ denotes the representation of each word in the input text. For words tokenized into multiple subwords, we use the representation of the first subword, which is a standard choice in the NER literature (Zaratiana et al., 2022).

Entity and Span Representation In our model, we aim to encode entity types and span embeddings into a unified latent space. The entity representation is computed by refining the initial representation $\boldsymbol{p}$ using a two-layer feedforward network, resulting in $\boldsymbol{q} = \{\boldsymbol{q}_i\}_0^{M-1} \in \mathbb{R}^{M \times D}$. The representation of a span starting at position i and ending at position j in the input text, $\boldsymbol{S}_{ij} \in \mathbb{R}^D$, is computed as:

$$\boldsymbol{S}_{ij} = \mathsf{FFN}(\boldsymbol{h}_i \otimes \boldsymbol{h}_j) \qquad (1)$$

Entity Type and Span Matching To evaluate whether a span (i, j) corresponds to entity type t, we calculate the following matching score:

$$\phi(i, j, t) = \sigma(\boldsymbol{S}_{ij}^T \boldsymbol{q}_t) \in \mathbb{R} \qquad (2)$$



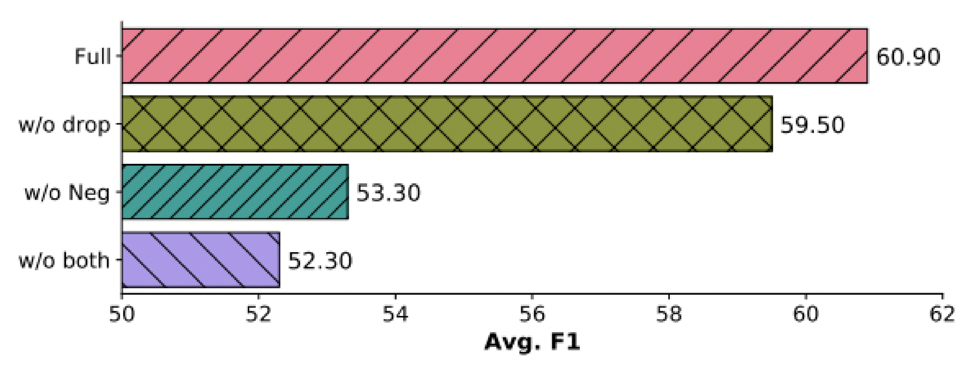
Main results

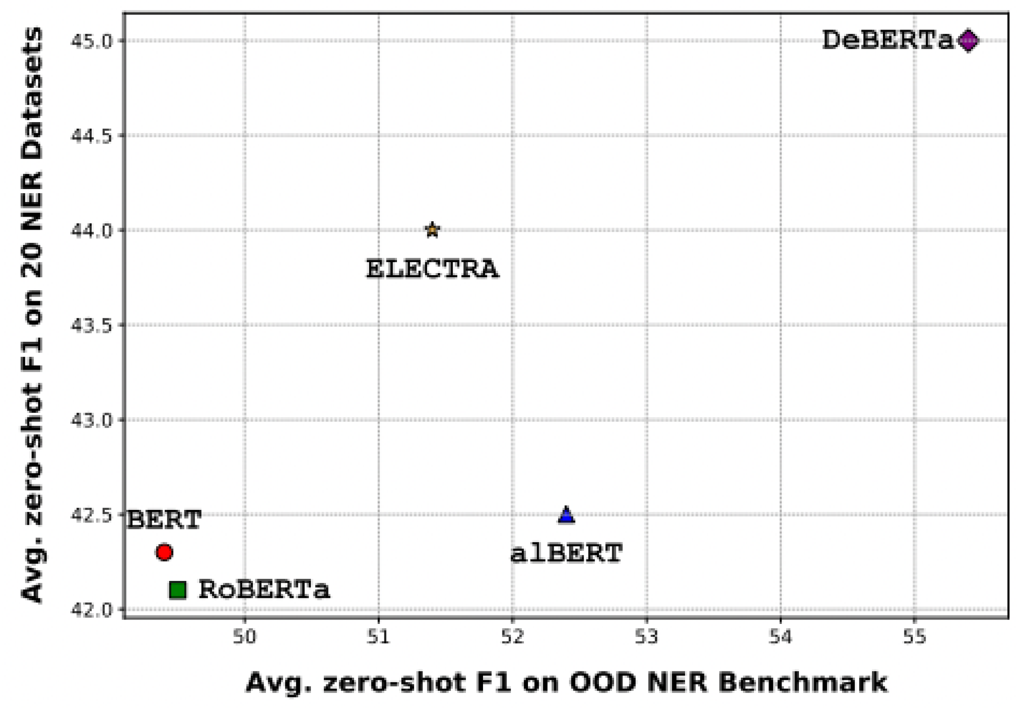
Model	Params	Movie	Restaurant	AI	Literature	Music	Politics	Science	Average
Vicuna-7B	7B	6.0	5.3	12.8	16.1	17.0	20.5	13.0	13.0
Vicuna-13B	13B	0.9	0.4	22.7	22.7	26.6	27.0	22.0	17.5
USM	0.3B	37.7	17.7	28.2	56.0	44.9	36.1	44.0	37.8
ChatGPT	–	5.3	32.8	52.4	39.8	66.6	68.5	**67.0**	47.5
InstructUIE	11B	**63.0**	21.0	49.0	47.2	53.2	48.1	49.2	47.2
UniNER-7B	7B	42.4	31.7	53.6	59.3	67.0	60.9	61.1	53.7
UniNER-13B	13B	48.7	36.2	54.2	60.9	64.5	61.4	63.5	55.6
GoLLIE	7B	**63.0**	**43.4**	**59.1**	62.7	67.8	57.2	55.5	58.0
GLiNER-S	50M	46.9	33.3	50.7	60.0	60.9	61.5	55.6	52.7
GLiNER-M	90M	42.9	37.3	51.8	59.7	69.4	68.6	58.1	55.4
GLiNER-L	0.3B	57.2	42.9	57.2	**64.4**	**69.6**	**72.6**	62.6	**60.9**

Table 1: **Zero-Shot Scores on Out-of-Domain NER Benchmark.** We report the performance of GLiNER with various DeBERTa-v3 (He et al., 2021) model sizes. Results for Vicuna, ChatGPT, and UniNER are from Zhou et al. (2023); USM and InstructUIE are from Wang et al. (2023); and GoLLIE is from Sainz et al. (2023).

Dataset	InstructUIE w/o	UniNER-7B w/	GLiNER-L w/	GLiNER-L w/o
ACE05	79.9	**86.7**	82.8	81.3
AnatEM	88.5	88.5	**88.9**	88.4
bc2gm	80.7	82.4	**83.7**	82.0
bc4chemd	87.6	**89.2**	87.9	86.7
bc5cdr	89.0	**89.3**	88.7	88.7
Broad Twitter	80.3	81.2	82.5	**82.7**
CoNLL03	91.5	**93.3**	92.6	92.5
FabNER	78.4	**81.9**	77.8	74.8
FindVehicle	87.6	**98.3**	95.7	95.2
GENIA	75.7	77.5	**78.9**	77.4
HarveyNER	**74.7**	74.2	68.6	67.4
MIT Movie	89.6	**90.2**	87.9	87.5
MIT Restaurant	82.6	82.3	**83.6**	83.3
MultiNERD	90.3	93.7	**93.8**	93.3
ncbi	86.2	87.0	**87.8**	87.1
OntoNotes	88.6	**89.9**	89.0	88.1
PolyglotNER	53.3	**65.7**	61.5	60.6
TweetNER7	**65.9**	65.8	51.4	50.3
WikiANN	64.5	**84.9**	83.7	82.8
wikiNeural	88.3	**93.3**	91.3	91.4
Average	81.2	**84.8**	82.9	82.1

Table 4: **In-domain Supervised Finetuning.** All the





	Language	Sup.	ChatGPT	GLiNER En	GLiNER Multi
Latin	German	64.6	37.1	35.6	**39.5**
	English	62.7	37.2	**42.4**	41.7
	Spanish	58.7	34.7	38.7	**42.1**
	Dutch	62.6	35.7	35.6	**38.9**
Non-Latin	Bengali	39.7	23.3	0.89	**25.9**
	Persian	52.3	25.9	14.9	**30.2**
	Hindi	47.8	27.3	11.3	**27.8**
	Korean	55.8	**30.0**	20.5	28.7
	Russian	59.7	27.4	30.3	**33.3**
	Turkish	46.8	**31.9**	22.0	30.0
	Chinese	53.1	18.8	6.59	**24.3**
Average		54.9	29.9	23.6	**32.9**

Table 3: **Zero-Shot Scores on Different Languages.**

Demo **Paper** **Code**