

# DocIE-Probing

## Probing Representations for Document-level Event Extraction

**Work by** Barry Wang, Xinya Du, Claire Cardie  
**Published** @EMNLP Findings 2023 | **Presenting** @PEER 2024

## TST3-MUC4-0061

Farabundo Marti National Liberation Front (FMLN) detachments have conducted the largest military operation in the entire history of the Salvadoran conflict in the country's capital. An offensive was launched simultaneously throughout the country's territory; however, San Salvador's city blocks, which were the scene of fierce fighting between FMLN detachments and government troops, were the epicenter of Sunday's events.

The intense firefight involving artillery and helicopters lasted approximately 18 hours and encompassed virtually the entire capital, completely paralyzing life in the city. According to Reuter, attempts were made to storm President Alfredo Cristiani's official and personal residences; however, it is reported that the president was not hurt. Reports of the losses on both sides are contradictory.

A state of siege has been introduced in El Salvador and constitutional guarantees have been suspended. A curfew is in force in the capital, and the international airport has been closed. According to President Cristiani, whose speech was broadcast on the radio, these steps are needed to "defend the population against left-wing terrorism." According to the President, the military are in complete control of the situation; however, it is reported from San Salvador that these words by A. Cristiani were heard on the airwaves accompanied by continual explosions and shooting.

International observers regard events in San Salvador in the context of the interrupted talks between the government and the insurgents. The third round of these talks should have been held recently in Caracas, but opposition representatives refused to take part in them after a left-wing trade union's headquarters was subjected to artillery bombardment resulting in the deaths of at least 10 people. According to the insurgents, the government decided on dialogue with them solely out of propaganda considerations, with no intention of reaching any agreement. Sunday's offensive was thus an attempt on their part to show themselves to be a force to be reckoned with.

In the opinion of observers, however, events in San Salvador practically rule out all hopes of success for future talks between the government and representatives of the insurgents. Moreover, the fighting in the Salvadoran capital, a Reuter correspondent believes, dealt the fragile peace settlement process in Central America a considerable blow.

Farabundo Marti National Liberation Front (FMLN) detachments have conducted the largest military operation in the entire history of the Salvadoran conflict in the country's capital. An offensive was launched simultaneously throughout the country's territory; however, San Salvador's city blocks, which were the scene of fierce fighting between FMLN detachments and government troops, were the epicenter of Sunday's events.

The intense firefight involving artillery and helicopters lasted approximately 18 hours and encompassed virtually the entire capital. Alfredo C... the losses

A state of... in force in... was broad... to the Pre... these wor... Internatio... and the in... represent... bombard

dialogue with them solely out of propaganda considerations, with no intention of reaching any agreement. Sunday's offensive was thus an attempt on their part to show themselves to be a force to be reckoned with.

In the opinion of observers, however, events in San Salvador practically rule out all hopes of success for future talks between the government and representatives of the insurgents. Moreover, the fighting in the Salvadoran capital, a Reuter correspondent believes, dealt the fragile peace settlement process in Central America a considerable blow.

## Datasets

- **MUC 3/4:** A naturally occurring corpus of newswire reports on terrorism in Latin America. The dataset has 1300/200/200 training/validation/testing documents.
- **WikiEvents** (Li et al., 2021) Another Document-level event dataset. It has 200/20/20 training/validation/testing inputs and has wider ranges of incident types.

ident  
reports of  
curfew is  
speech  
according  
for that  
vernment  
tillery  
cided on

```
[
  {
    "incident_type": "attack",
    "PerpInd": [
      [ "detachments" aka "fmln detachments" ]
    ],
    "PerpOrg": [
      ["farabundo marti national liberation front" aka "fmln" ]
    ],
    "Target": [
      ["president alfredo cristiani's official
        and personal residences" ]
    ],
    "Victim": [
      ["alfredo cristiani"]
    ],
    "Weapon": []
  },

```

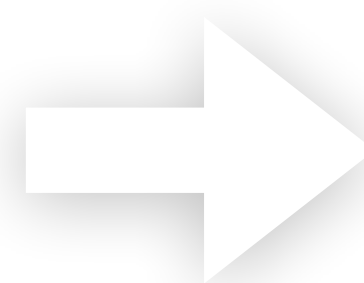
```
{
  "incident_type": "bombing",
  "PerpInd": [],
  "PerpOrg": [],
  "Target": [
    ["left-wing trade union's headquarters"
      aka "trade union's headquarters"]
  ],
  "Victim": [],
  "Weapon": [
    ["artillery"]
  ]
}
```

[...] **FMLN detachments** have conducted the largest military operation in the entire history of the Salvadoran conflict in the country's capital. An offensive was launched [...]

According to Reuter, attempts were made to storm **President Alfredo Cristiani's official and personal residences**; however, it is reported that the president was not hurt.[...]

The third round of these talks should have been held recently in Caracas, but opposition representatives refused to take part in them after a **left-wing trade union's headquarters** was subjected to **artillery** bombardment resulting in the deaths of at least 10 people. According to the insurgents, [...]

[...] FMLN detachments have conducted the largest military operation in the entire history of the Salvadoran conflict in the country's capital. An offensive was launched [...] According to Reuter, attempts were made to storm **President Alfredo Cristiani's official and personal residences**; however, it is reported that the president was not hurt.[...] The third round of these talks should have been held recently in Caracas, but opposition representatives refused to take part in them after **a left-wing trade union's headquarters** was subjected to **artillery bombardment** resulting in the deaths of at least 10 people. According to the insurgents, [...]



**Event type:** **Template #1**  
**Attack**/Bombing/Kidnap  
**Perpetrator:**  
1 • FMLN detachments  
**Target:**  
1 • President Alfredo Cristiani official and

...

**Event type:** **Template #2**  
Attack/**Bombing**/Kidnap  
**Perpetrator:**  
• N/A  
**Target:**  
1 • left-wing trade union's headquarters

...

[...] FMLN detachments have conducted the largest military operation in the entire history of the Salvadoran conflict in the country's capital. An offensive was launched [...] According to Reuter, attempts were made to storm **President Alfredo Cristiani's official and personal residences**; however, it is reported that the president was not hurt.[...] The third round of these talks should have been held recently in Caracas, but opposition representatives refused to take part in them after **a left-wing trade union's headquarters** was subjected to **artillery bombardment** resulting in the deaths of at least 10 people. According to the insurgents, [...]



**Event type:** **Template #1**

Attack/Bombing/Kidnap

**Perpetrator:** **Victim:**

1 • FMLN detachments

**Target:**

1 • President Alfredo Cristiani official and  
...

**Template ?**

**Event type:**

Attack/Bombing/Kidnap

**Perpetrator:**

• N/A

**Target:**

1 • opposition representatives  
...

**Template #2**

**Event type:**

Attack/Bombing/Kidnap

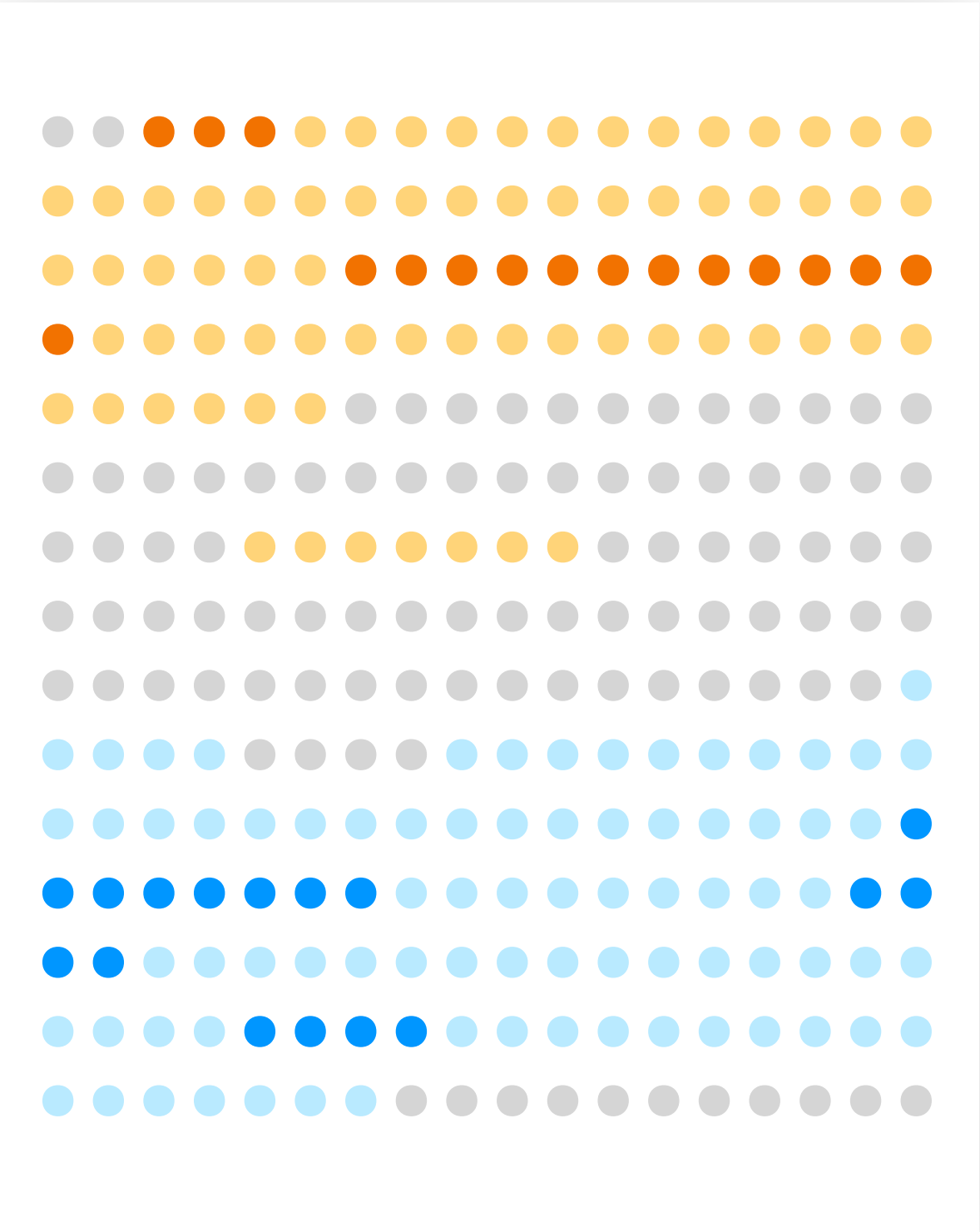
**Perpetrator:**

• N/A

**Target:**

1 • left-wing trade union's headquarters  
...

[...] FMLN detachments have conducted the largest military operation in the entire history of the Salvadoran conflict in the country's capital. An offensive was launched [...] According to Reuter, attempts were made to storm **President Alfredo Cristiani's official and personal residences**; however, it is reported that the president was not hurt.[...] The third round of these talks should have been held recently in Caracas, but opposition representatives refused to take part in them after **a left-wing trade union's headquarters** was subjected to **artillery bombardment** resulting in the deaths of at least 10 people. According to the insurgents, [...]

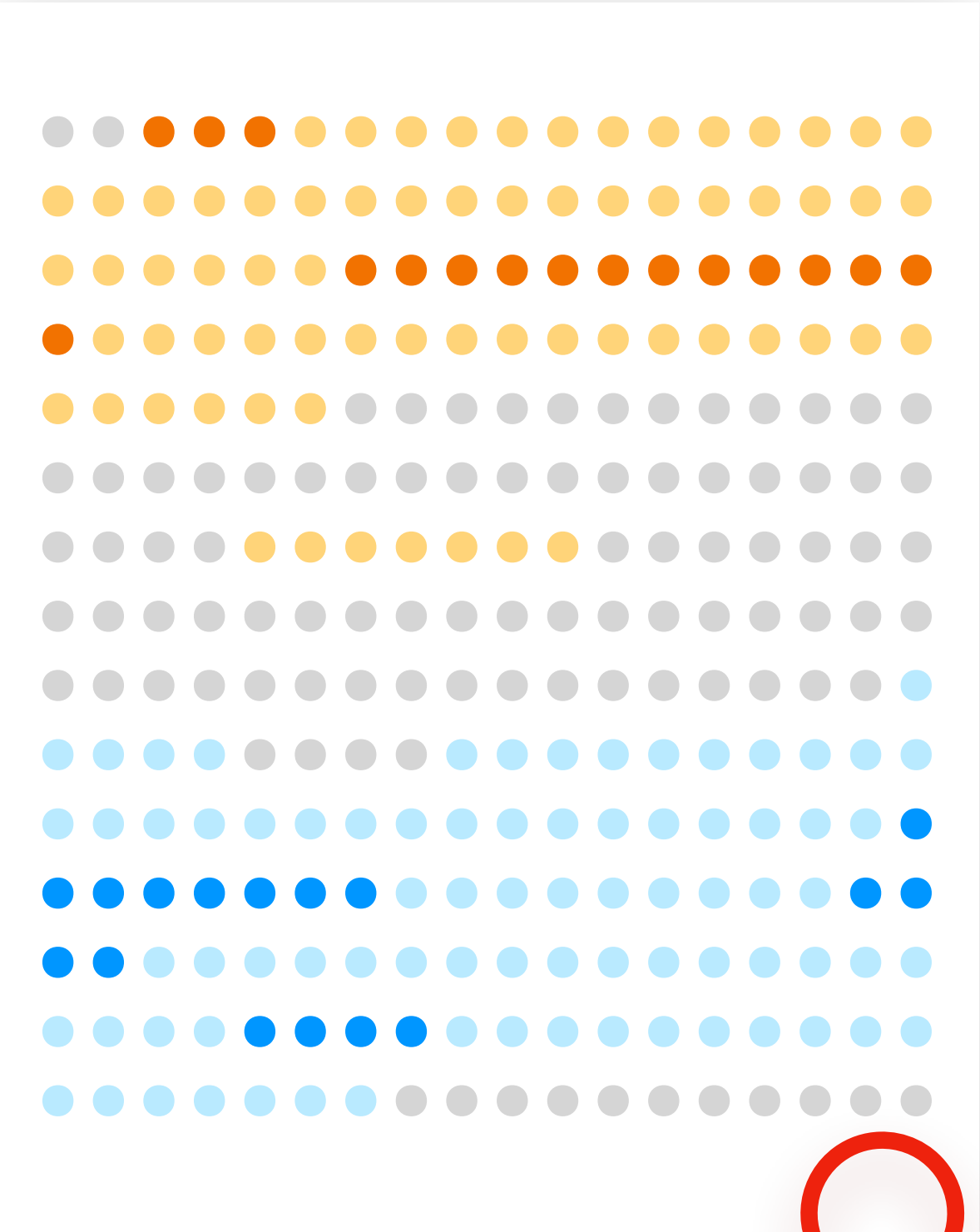


**Event type:** **Template #1**  
Attack/Bombing/Kidnap  
**Perpetrator:**  
1 • FMLN detachments  
**Target:**  
1 • President Alfredo Cristiani official and  
...

**Event type:** **Template #2**  
Attack/Bombing/Kidnap  
**Perpetrator:**  
• N/A  
**Target:**  
1 • left-wing trade union's headquarters  
...



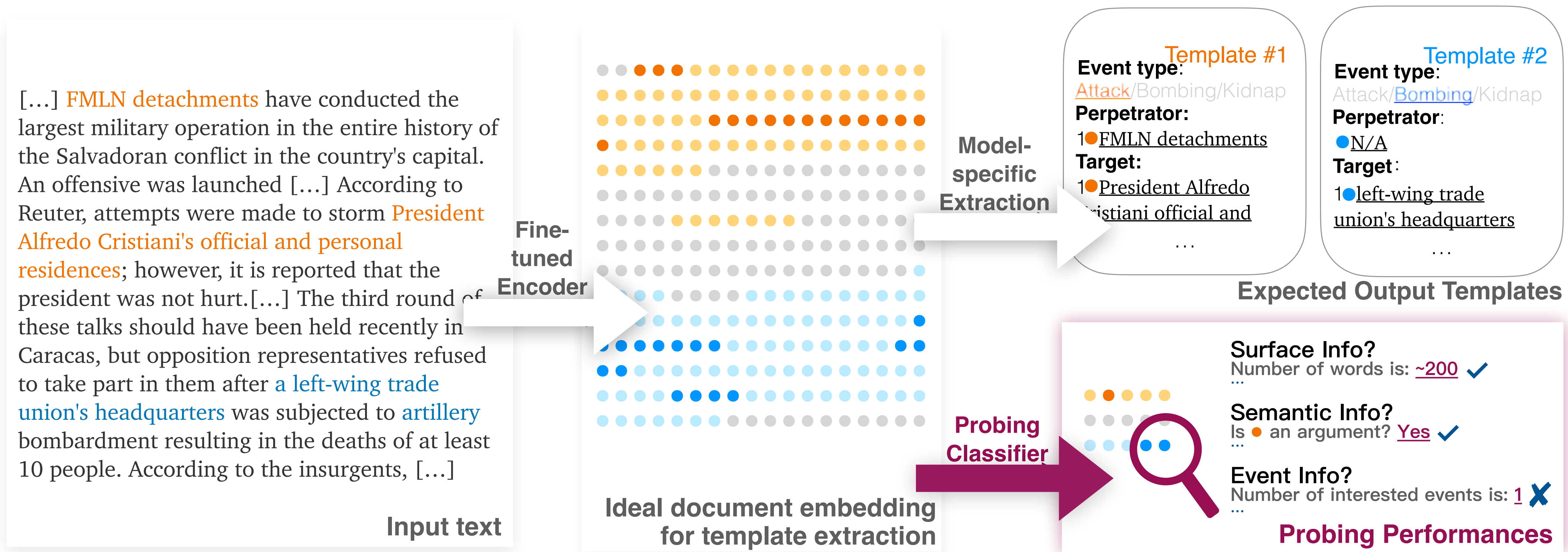
[...] FMLN detachments have conducted the largest military operation in the entire history of the Salvadoran conflict in the country's capital. An offensive was launched [...] According to Reuter, attempts were made to storm **President Alfredo Cristiani's official and personal residences**; however, it is reported that the president was not hurt.[...] The third round of these talks should have been held recently in Caracas, but opposition representatives refused to take part in them after **a left-wing trade union's headquarters** was subjected to **artillery bombardment** resulting in the deaths of at least 10 people. According to the insurgents, [...]



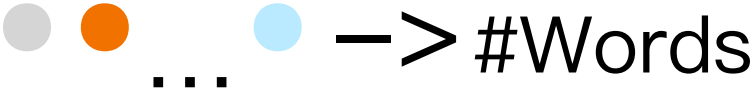
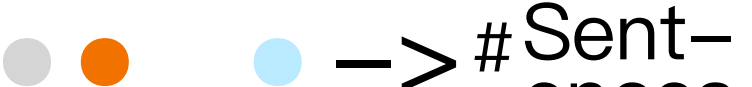

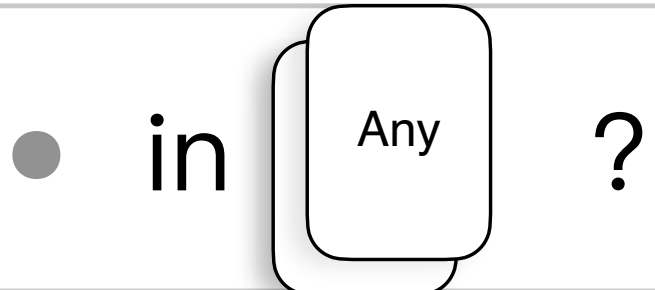
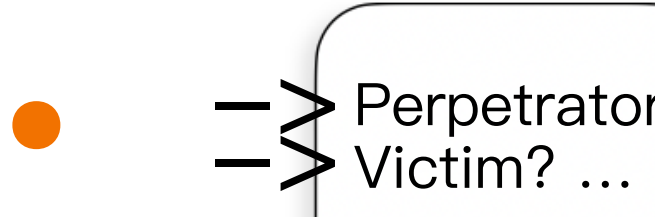
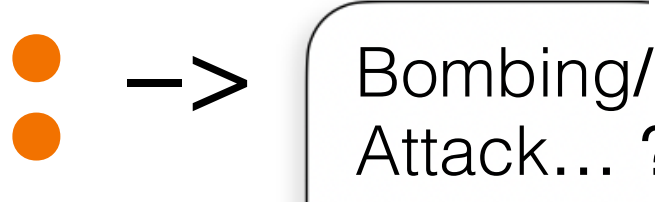

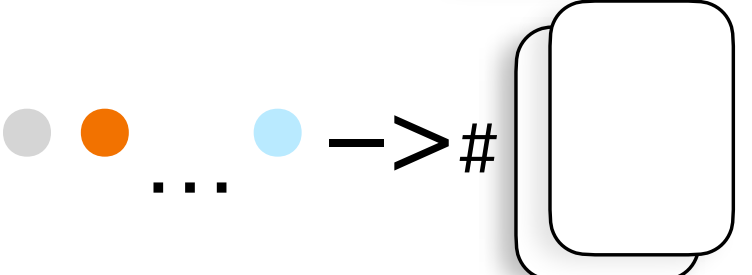
**Event type:** **Template #1**  
Attack/Bombing/Kidnap  
**Perpetrator:**  
1 • FMLN detachments  
**Target:**  
1 • President Alfredo Cristiani official and  
...

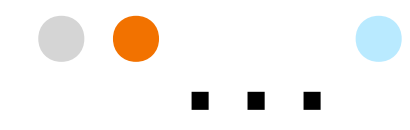
**Event type:** **Template #2**  
Attack/Bombing/Kidnap  
**Perpetrator:**  
• N/A  
**Target:**  
1 • left-wing trade union's headquarters  
...

**Probing!**  
(Tenney et al., 2019)

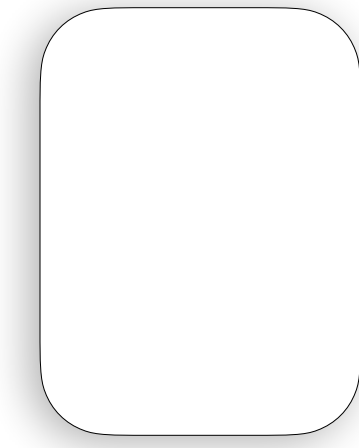


**Figure 1: Overview of an ideal event extraction example and probing.** Ideally, after contextualization by a fine-tuned encoder on the IE task, the per-token embedding can capture richer semantic and related event information, thereby facilitating an easier model-specific extraction process. Our probing tasks test how different frameworks and conditions (e.g. IE training, coherence information access) affect information captured by the embeddings.

Category	Illustration	Task	Task Full Name
Surface	 $\rightarrow \#Words$	WordCt	Word Count
	 $\rightarrow \#Sentences$	SentCt	Sentence Count
Semantic	 a.k.a. ?	Coref	Are Coreferent
	 in <span style="border: 1px solid black; border-radius: 10px; padding: 2px;">Any</span> ?	IsArg	Is an Argument
	 $\rightarrow$ <span style="border-left: 1px solid black; border-right: 1px solid black; padding: 0 5px;">Perpetrator? Victim? ...</span>	ArgTyp	Argument Type
	 $\rightarrow$ <span style="border: 1px solid black; border-radius: 10px; padding: 2px;">Bombing/ Attack... ?</span>	EvntTyp <sub>2</sub>	Event Type
Event	 both in <span style="border: 1px solid black; border-radius: 10px; padding: 2px;">Any</span> ?	CoEvnt	Co-Event
	 $\rightarrow \#$ <span style="border: 1px solid black; border-radius: 10px; padding: 2px;"></span>	EvntCt	Event Count



Full Text Embeddings



Gold Event

 Randomly Sampled token embedding

 Randomly Sampled Argument embedding

**Models** ▼

- TANL
- GTT
- DyGIE++

**Dataset** ▼

- MUC
- WikiEvent
- Adaptable to others

**IE Training** ▼

- 5 Epochs
- 10 Epochs
- 15 Epochs

**Embedding** ▼

- Full Text One-Pass
- Sentence Concat'ed

**Encoder Layers** ▼

- BERT - 0
- ...
- GTT-BERT - 12

**Probing Tasks** ▼

- Surface Info
- Semantic Info
- Event Info

**Probing Model** ▼

- Attention + SentEval NN
- +Multi-Head Attn.

**Input Length** ▼

- Short (<209 words)
- Medium (210~420)
- Long (420+)

# Information Extraction Models

- **DyGIE++** (Wadden et al., 2019) is a discriminative multi-task framework. It achieves IE by enumerating and scoring sections (spans) of encoded text and using the relations between spans to detect triggers and events.
- **GTT** (Du et al., 2021) is a sequence-to-sequence event-extraction model that perform the task end-to-end, without using labeled triggers. It is trained to decode a serialized template, with tuned decoder constraints.
- **TANL** (Paolini et al., 2021) is a sequence-to-sequence multi-task model that "translates" input text to augmented languages. For IE, the in-text augmented parts identify triggers and roles. It uses a two-stage approach for event extraction by first extracting trigger then finding arguments for each trigger predicted.

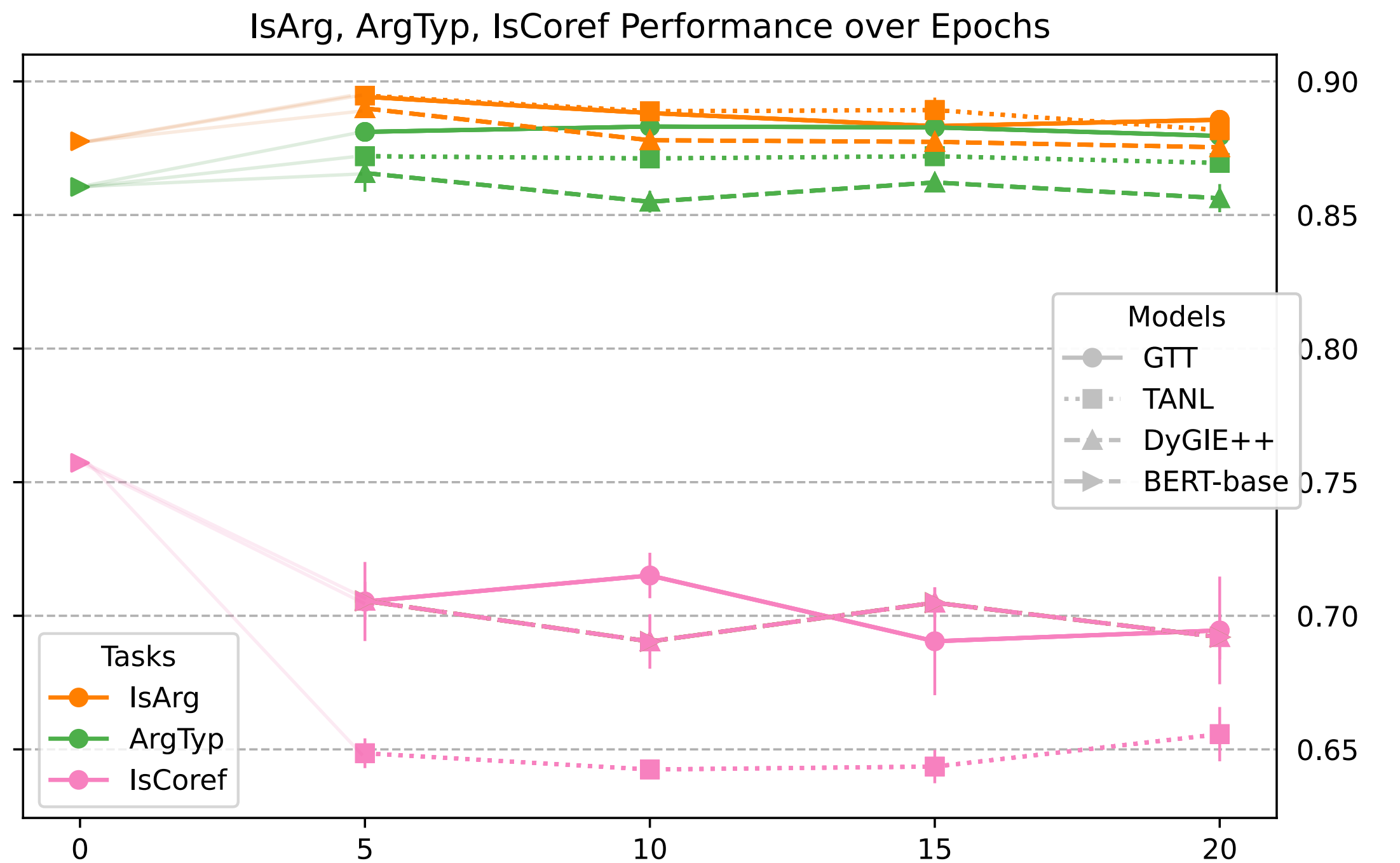
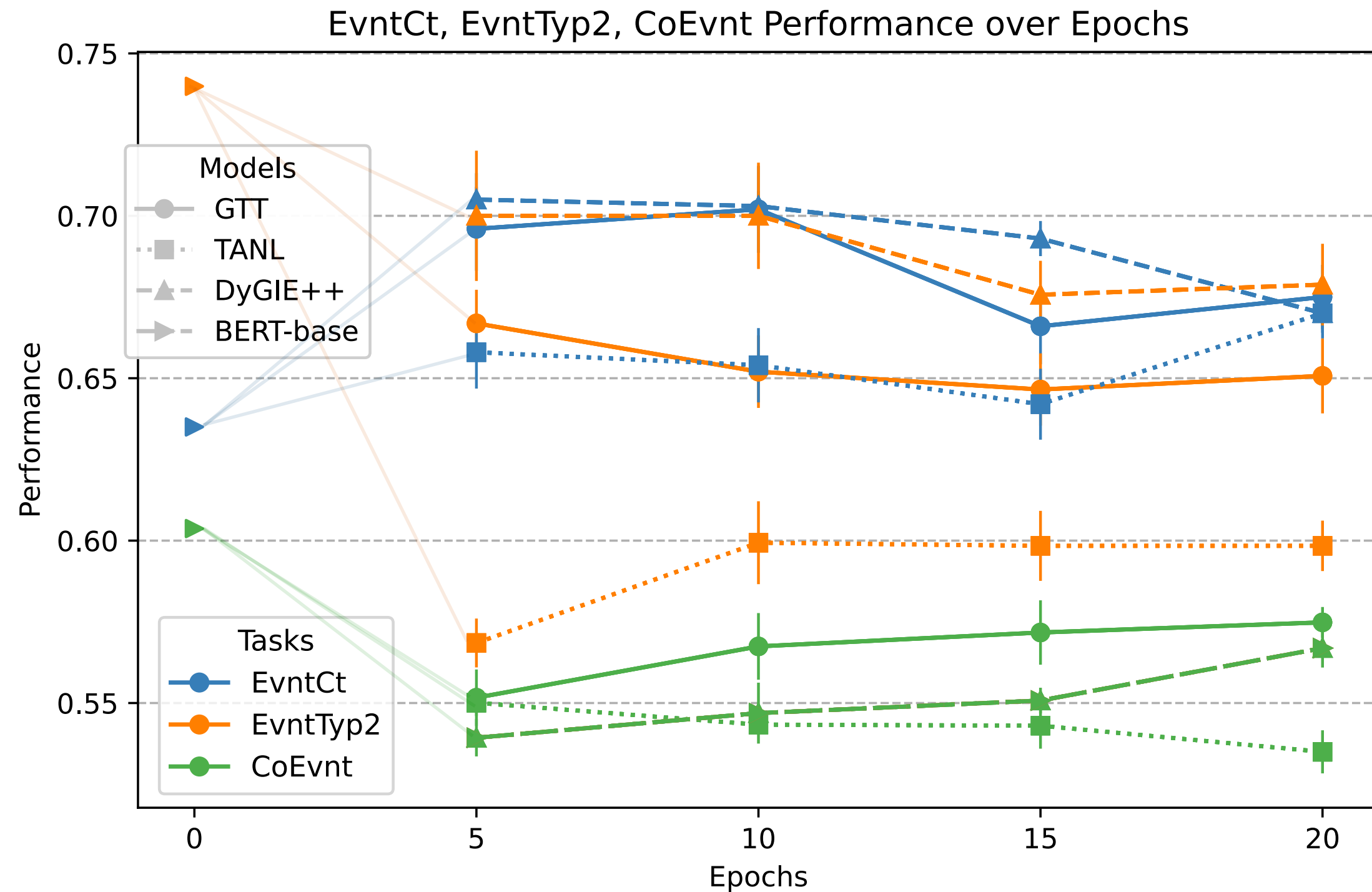
# Probing Performances

## with Different IE Frameworks and Embedding Method

Model (IE-F1)	Input	WordCt	SentCt	IsArg	ArgTyp	Coref	EvntTyp <sub>2</sub>	CoEvnt	EvntCt	Avg
<b>DyGIE++</b> (41.9)	FullText	58.6	<u>47.0</u>	87.1	83.8	64.7	<b><u>60.5</u></b>	73.6	67.2	67.8
	SentCat	<u>57.4</u>	58.9	87.5	85.6	69.2	56.7	<u>67.9</u>	67.0	68.8
<b>GTT</b> (49.0)	FullText	58.6	46.3	<u>88.3</u>	<b><u>88.5</u></b>	66.7	60.4	66.4	<b><u>68.3</u></b>	67.9
	SentCat	55.8	<b><u>58.9</u></b>	<b><u>88.6</u></b>	<u>88.0</u>	<u>69.5</u>	<u>57.5</u>	65.07	<u>67.5</u>	68.8
<b>TANL</b> (33.2)	FullText	54.2	43.3	88.2	86.8	66.6	57.8	60.0	65.8	65.3
	SentCat	34.3	40.8	88.2	87.0	65.6	53.5	59.8	67.0	62.0
<b>BERT</b> <sub>base</sub>	FullText	<b><u>65.5</u></b>	45.0	87.8	86.1	<b><u>75.7</u></b>	60.4	<b><u>74.0</u></b>	63.5	69.7

**Table 1: Probing Task Test Average Accuracy.** IE frameworks trained for 20 epochs on MUC, and we run probing tasks on the input representations. We compare the 5-trial averaged test accuracy on full-text embeddings and concatenation of sentence embeddings from the same encoder to the untrained BERT baseline. IE-F1 refers to the model’s F1 score on MUC test. Underlined data are the best in same embedding method, while bold, overall.

# Probing Performances over IE Framework Training Epochs



# EventCt Probing Performances over Different Input Lengths

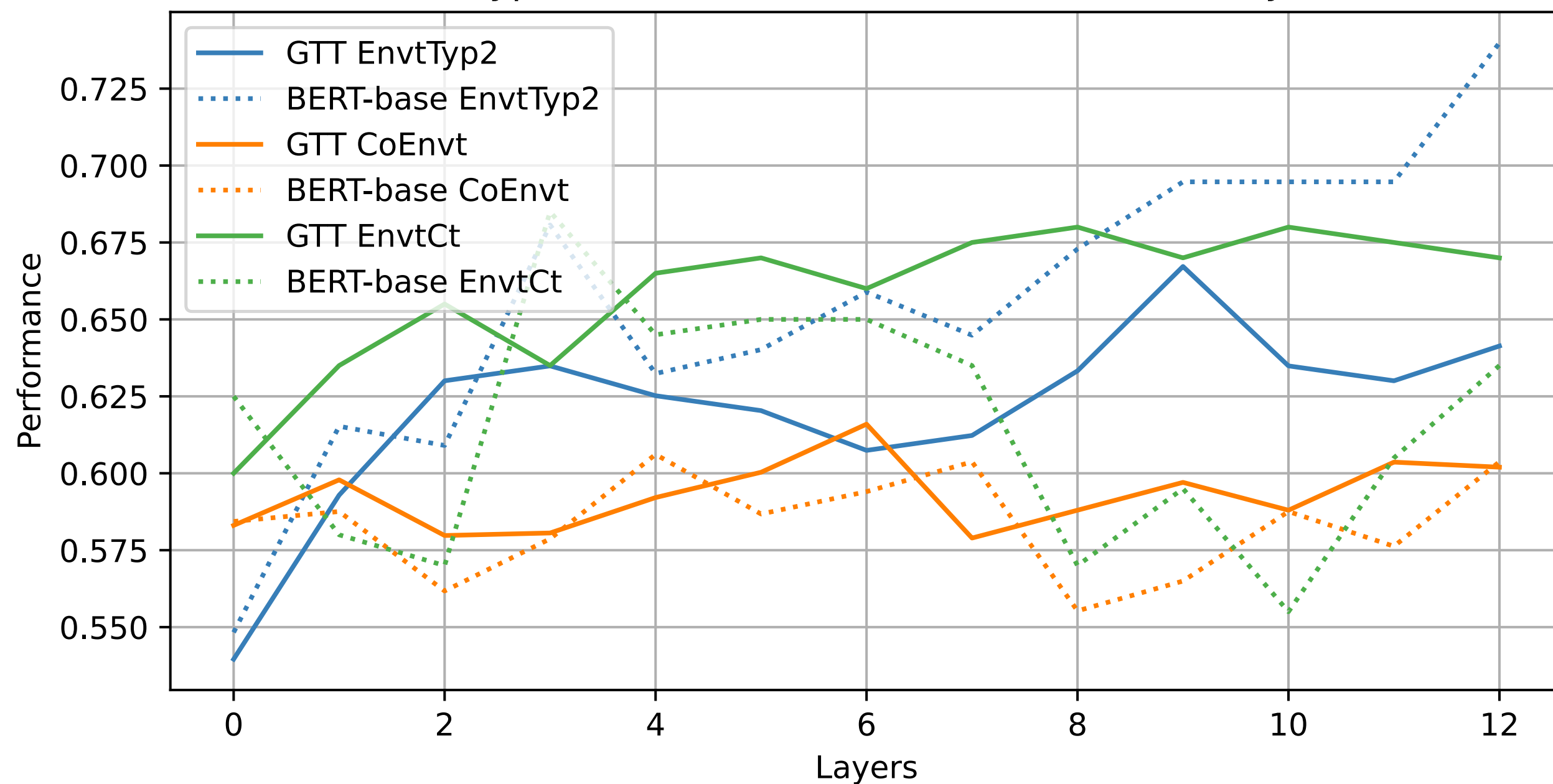
<b>Model</b>	<b>FullText Best</b>	<b>FullText Avg</b>	<b>Sent Best</b>	<b>Sent Avg</b>
WordCount: $\leq 209$				
DyGIE++	68.5	67.1	<b>69.7</b>	<b>68.8</b>
GTT	70.3	<b>68.7</b>	<b>72.1</b>	68.0
TANL	<b>71.8</b>	<b>70.2</b>	66.3	64.2
WordCount: 210-420				
DyGIE++	67.0	<b>65.7</b>	<b>67.6</b>	64.7
GTT	<b>67.6</b>	<b>67.0</b>	66.4	64.7
TANL	<b>64.8</b>	<b>62.0</b>	63.6	60.8
WordCount: $\geq 431$				
DyGIE++	70.6	70.2	<b>74.2</b>	<b>72.1</b>
GTT	69.1	68.7	<b>71.5</b>	<b>70.2</b>
TANL	67.3	65.2	<b>69.7</b>	<b>68.3</b>

**Table 2: EvntCt Probing Test Accuracy (%)** 5 random seed averaged. When WordCount  $\geq 431$ , both FullText and SentCt models had the same performance.

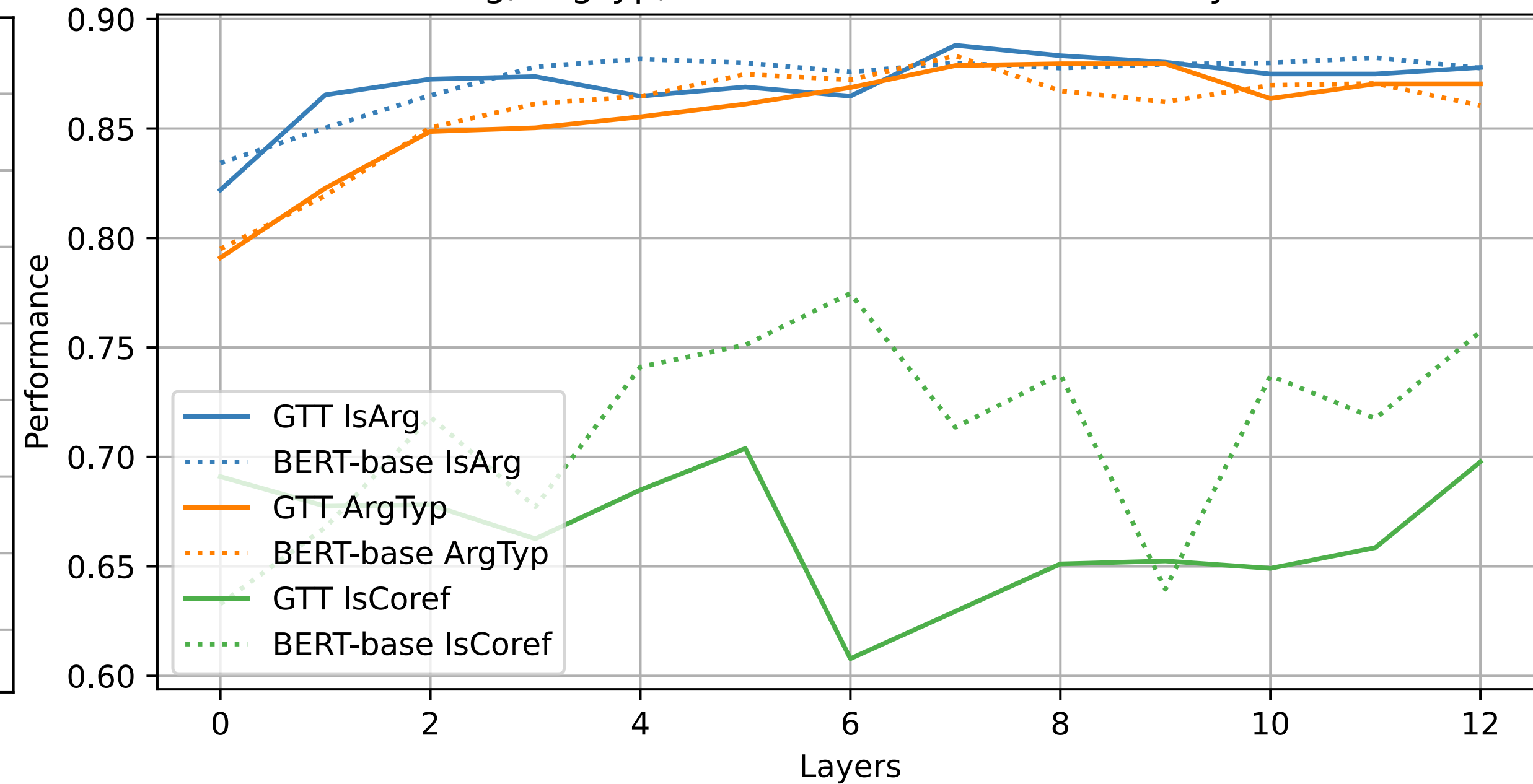


# Probing Performances over BERT Layers

EnvTyp2, CoEnvt, EnvTct Performance over Layers



IsArg, ArgTyp, IsCoref Performance over Layers



**Probe acc.** on event & semantic information over BERT layers as-is and from GTT trained over 18 epochs.

# Take Away

- Trained encoding improves on capabilities like event detection and argument labeling, but IE training compromises encoder's ability to encode coreference and event typing information.
- Current models marginally outperformed the baseline in capturing event information at best, uncovered by comparisons of IE frameworks.
- Encoder models struggle with document length and cross-sentence discourse, as concatenation of sentence embeddings yielded stronger probing performances.

# References

And thank you for listening!

- 1991. Third Message Understanding Conference (MUC-3): Proceedings of a Conference Held in San Diego, California, May 21-23, 1991.
- Aliva Das, Xinya Du, Barry Wang, Kejian Shi, Jiayuan Gu, Thomas Porter, and Claire Cardie. 2022. Automatic error analysis for document-level information extraction. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 3960–3975, Dublin, Ireland. Association for Computational Linguistics.
- Xinya Du, Alexander Rush, and Claire Cardie. 2021. Template filling with generative transformers. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 909–914, Online. Association for Computational Linguistics.
- Giovanni Paolini, Ben Athiwaratkun, Jason Krone, Jie Ma, Alessandro Achille, Rishita Anubhai, Cicero Nogueira dos Santos, Bing Xiang, and Stefano Soatto. 2021. Structured prediction as translation between augmented natural languages.
- Ian Tenney, Patrick Xia, Berlin Chen, Alex Wang, Adam Poliak, R Thomas McCoy, Najoung Kim, Benjamin Van Durme, Samuel R. Bowman, Dipanjan Das, and Ellie Pavlick. 2019b. What do you learn from context? probing for sentence structure in contextualized word representations.
- David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. 2019. Entity, relation, and event extraction with contextualized span representations. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 57845789, Hong Kong, China. Association for Computational Linguistics.