

EMNLP 2026 Submission Template: A Skeleton for Empirical NLP Papers

First Author¹ **Second Author**^{1,2} **Third Author**²

¹Department of Computer Science, University Name

²Industry Research Lab

{first,second}@university.edu, third@lab.com

Abstract

This document is a skeleton for papers submitted to the 2026 Conference on Empirical Methods in Natural Language Processing (EMNLP 2026). The abstract should be a single paragraph that states the problem, summarizes the proposed method, and reports the key quantitative findings. Keep it under 200 words so that it fits comfortably in the first column.

1 Introduction

Introduce the task, explain why it matters, and identify the gap in prior work that this paper fills. EMNLP long papers are limited to eight pages of content and short papers to four pages; references, the Limitations section and appendices do not count toward the limit.

Close the introduction with a concise list of contributions:

- A new method for [task] that [key idea].
- A benchmark or dataset of [size] [units] covering [languages or domains].
- Empirical results showing [improvement] over strong baselines, with analysis of [phenomenon].

2 Related Work

Situate the paper relative to prior work on [topic A] (Devlin et al., 2019), [topic B] (Vaswani et al., 2017), and [topic C]. Use `\citet{}` for textual citations, e.g. Brown et al. (2020) show that ..., and `\citep{}` for parenthetical ones.

3 Method

Describe the model or approach. Given an input sequence $x = (x_1, \dots, x_n)$, we model the output y as

$$P(y \mid x) = \prod_{t=1}^{|y|} P(y_t \mid y_{<t}, x; \theta), \quad (1)$$

where θ are the model parameters. Training minimizes the negative log-likelihood

$$\mathcal{L}(\theta) = - \sum_{(x,y) \in \mathcal{D}} \log P(y \mid x; \theta). \quad (2)$$

3.1 Architecture

Give the architectural details needed to reproduce the work: layer counts, hidden sizes, tokenization, and any task-specific components.

3.2 Training Objective

Explain any auxiliary losses, curriculum, or data-selection strategy.

4 Experiments

Datasets. List the datasets with sizes and splits, and cite their sources.

Baselines. Describe the systems you compare against and how they were tuned.

Implementation details. Report hyperparameters, hardware, training time, and the number of runs used for each reported number.

5 Results

Table 1 reports the main results. Report means and standard deviations over multiple seeds where possible, and mark statistically significant improvements.

Model	Acc.	F1	BLEU
Baseline A (Devlin et al., 2019)	84.2	83.1	27.4
Baseline B (Brown et al., 2020)	86.0	85.3	29.1
Ours (base)	87.4	86.6	30.2
Ours (large)	89.1	88.4	31.8

Table 1: Main results on the test set. Best results in bold.

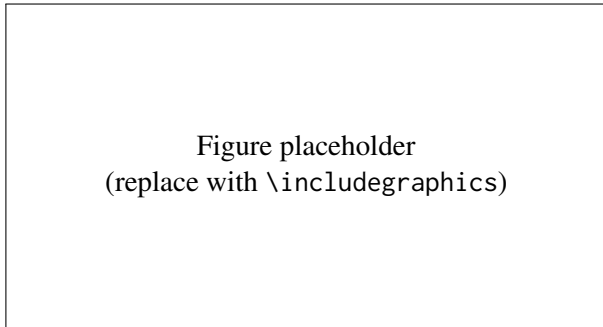


Figure 1: Example figure. Use vector graphics (PDF) where possible and make sure text is legible when printed.

5.1 Analysis

Include ablations, error analysis, and qualitative examples that explain *why* the method works.

6 Conclusion

Summarize the findings, restate the main contribution, and outline directions for future work.

Limitations

EMNLP requires every paper to include a Limitations section. Discuss the scope of the claims, the languages and domains evaluated, computational requirements, and any assumptions that may not hold in practice. This section does not count toward the page limit.

Ethics Statement

Optionally discuss potential risks, intended use, data licensing, annotator compensation, and any steps taken to mitigate harm. This section does not count toward the page limit.

Acknowledgments

Acknowledgments appear only in the camera-ready version. Remove or anonymize them for the review submission.

References

- Tom Brown, Benjamin Mann, Nick Ryder, et al. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, et al. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*.

A Additional Results

Appendices follow the references and may contain extra experiments, prompts, and dataset details.