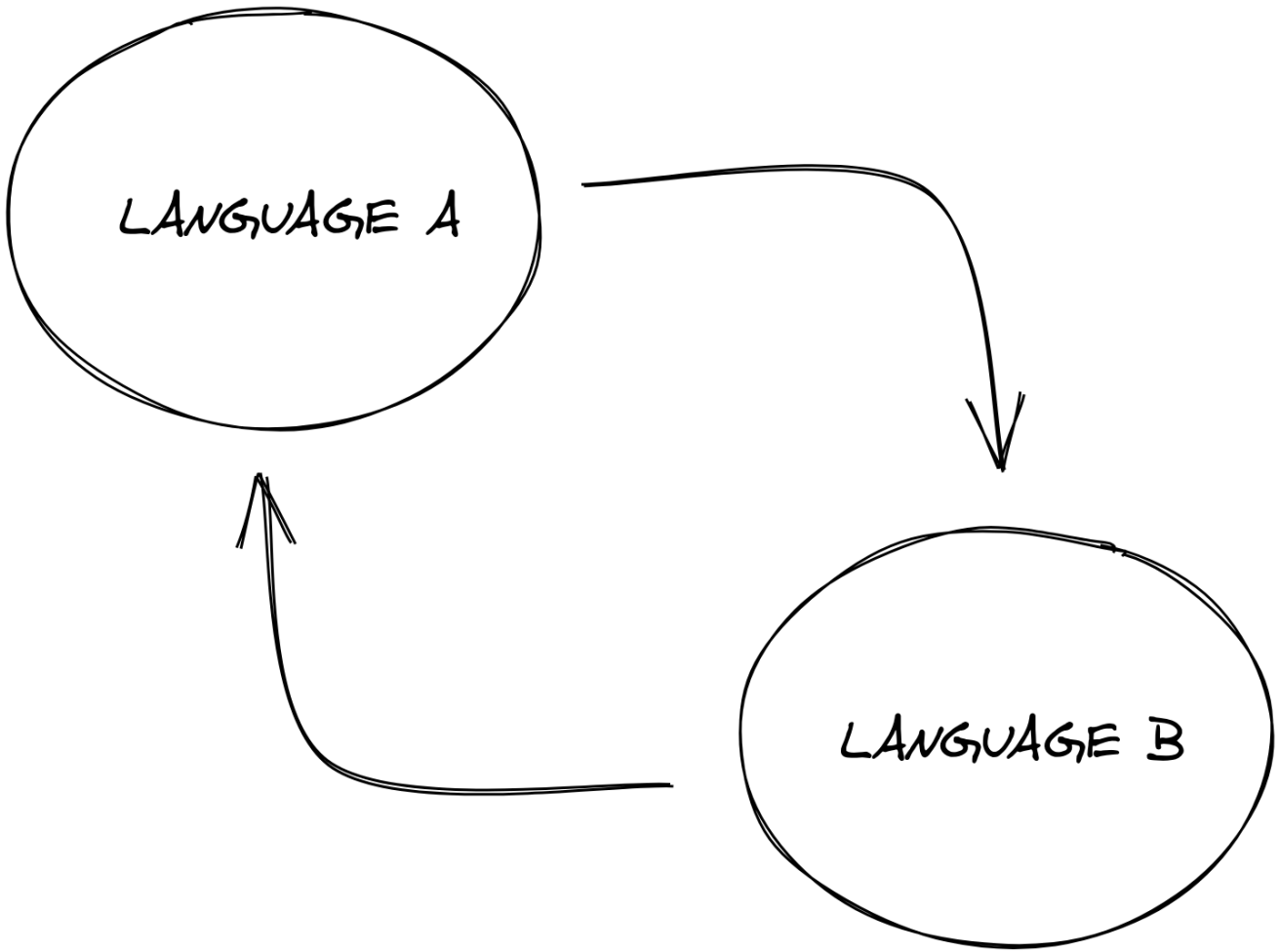
Adversarial Attacks on Cross-lingual Models

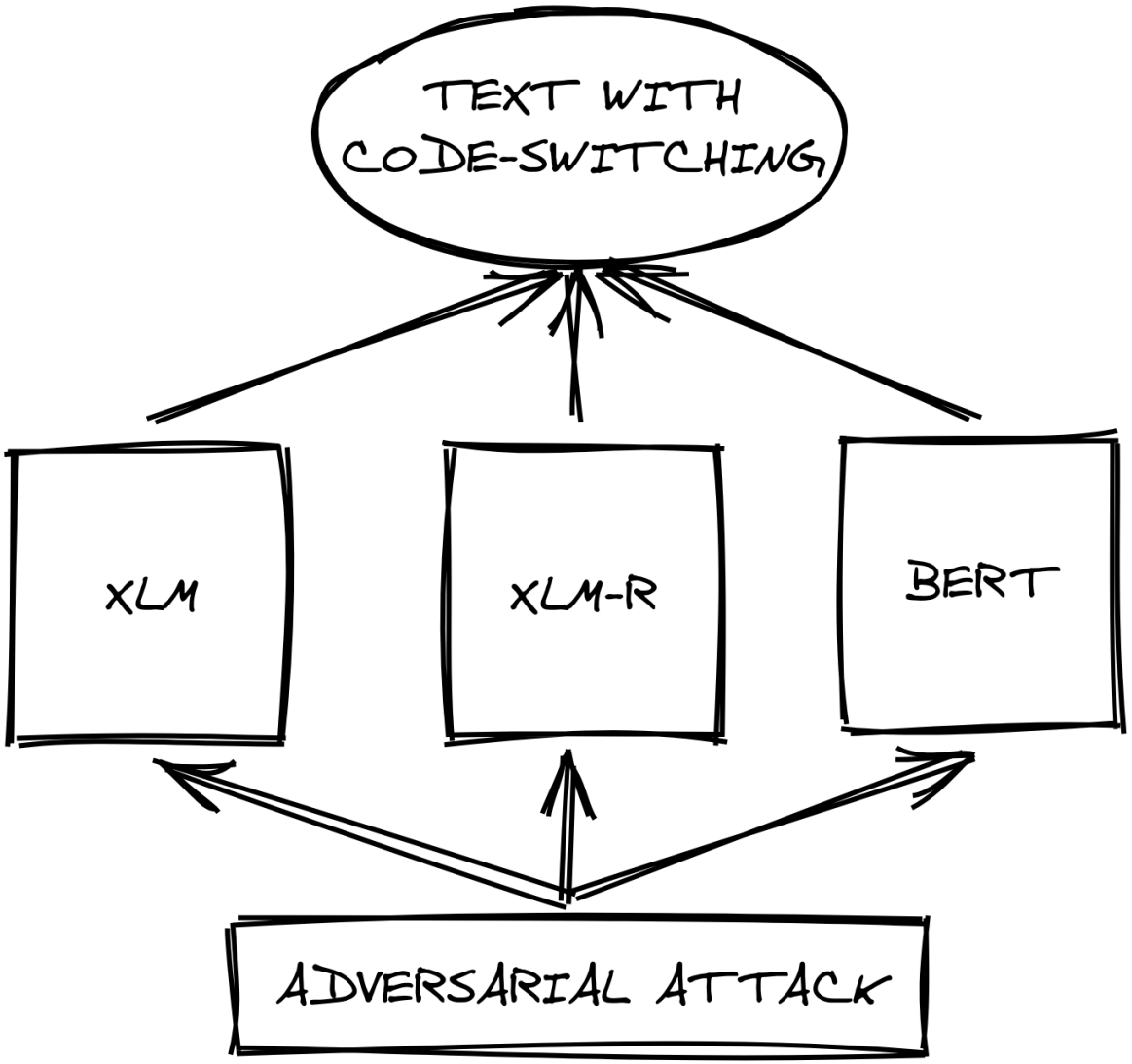
Birshert Alexey

March 26, 2021

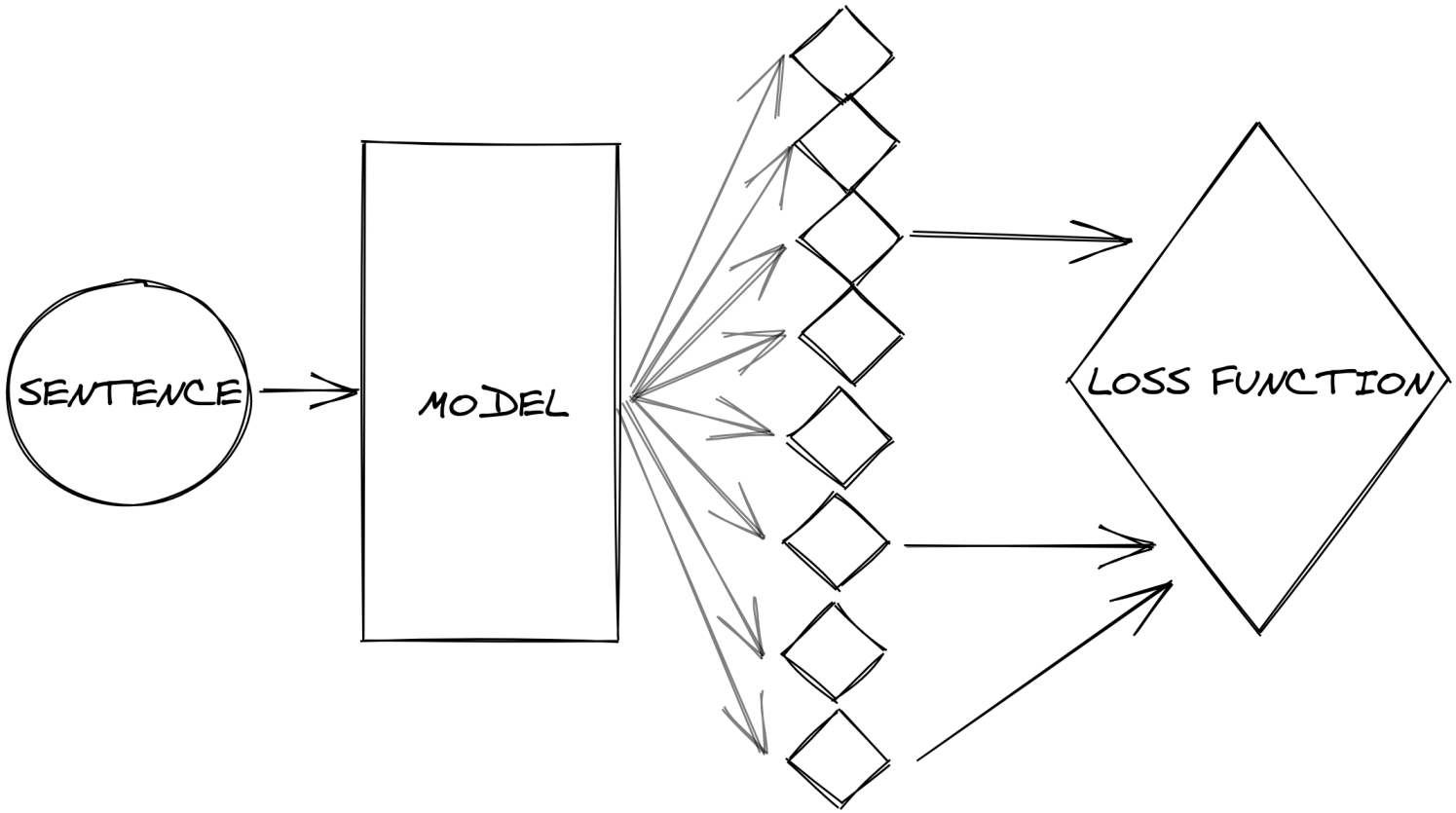
Code-switching



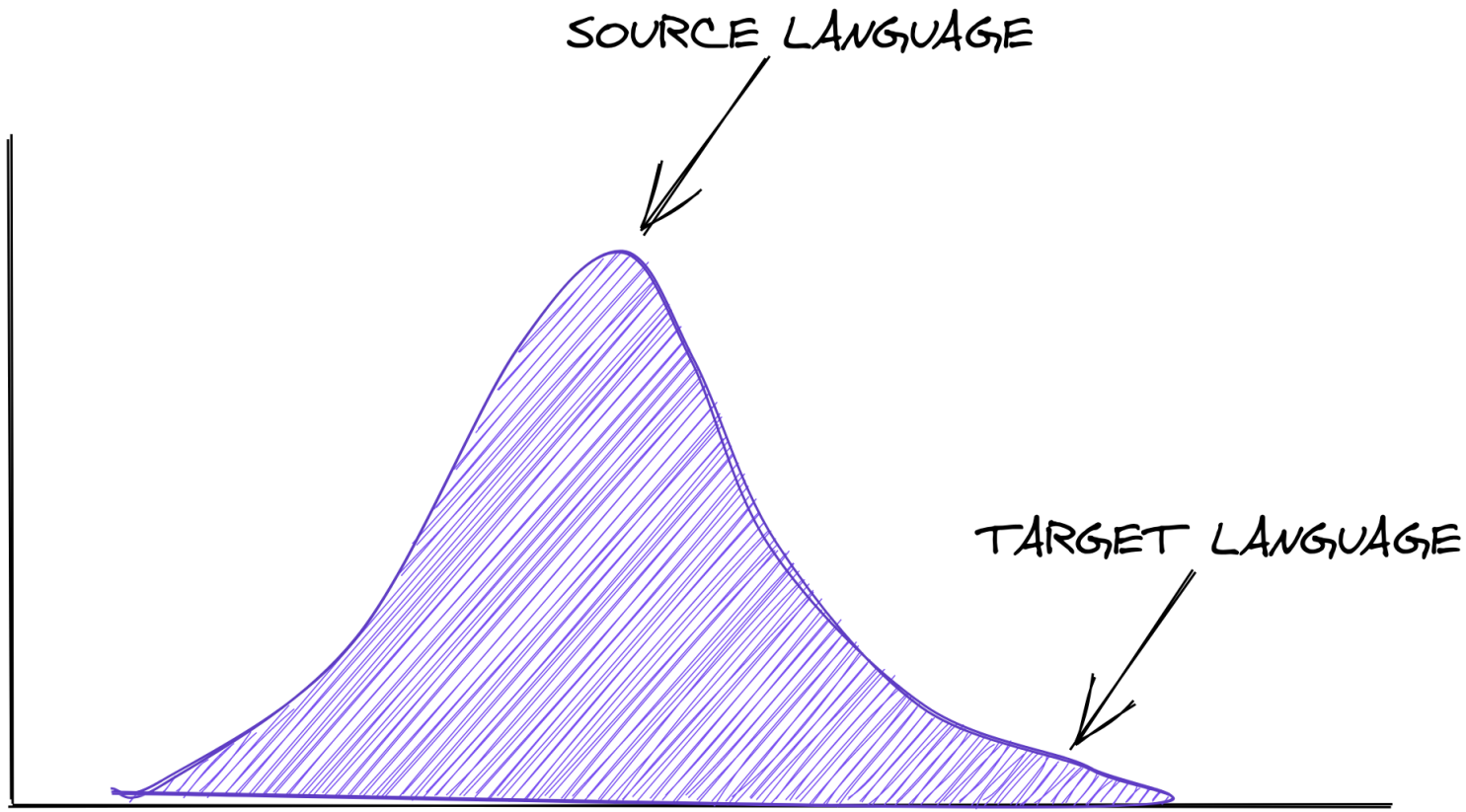
Texts with code-switching generation



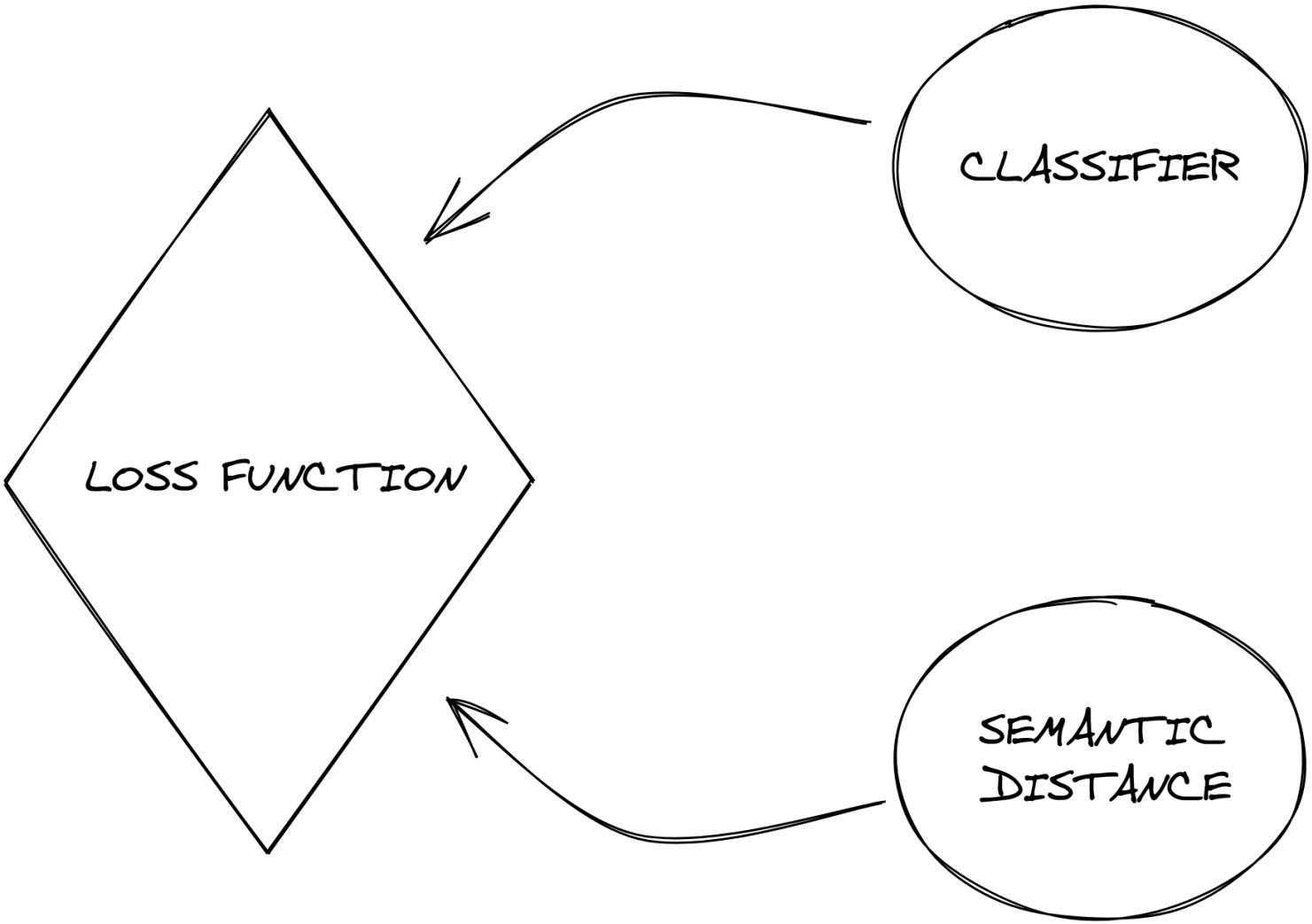
Training procces



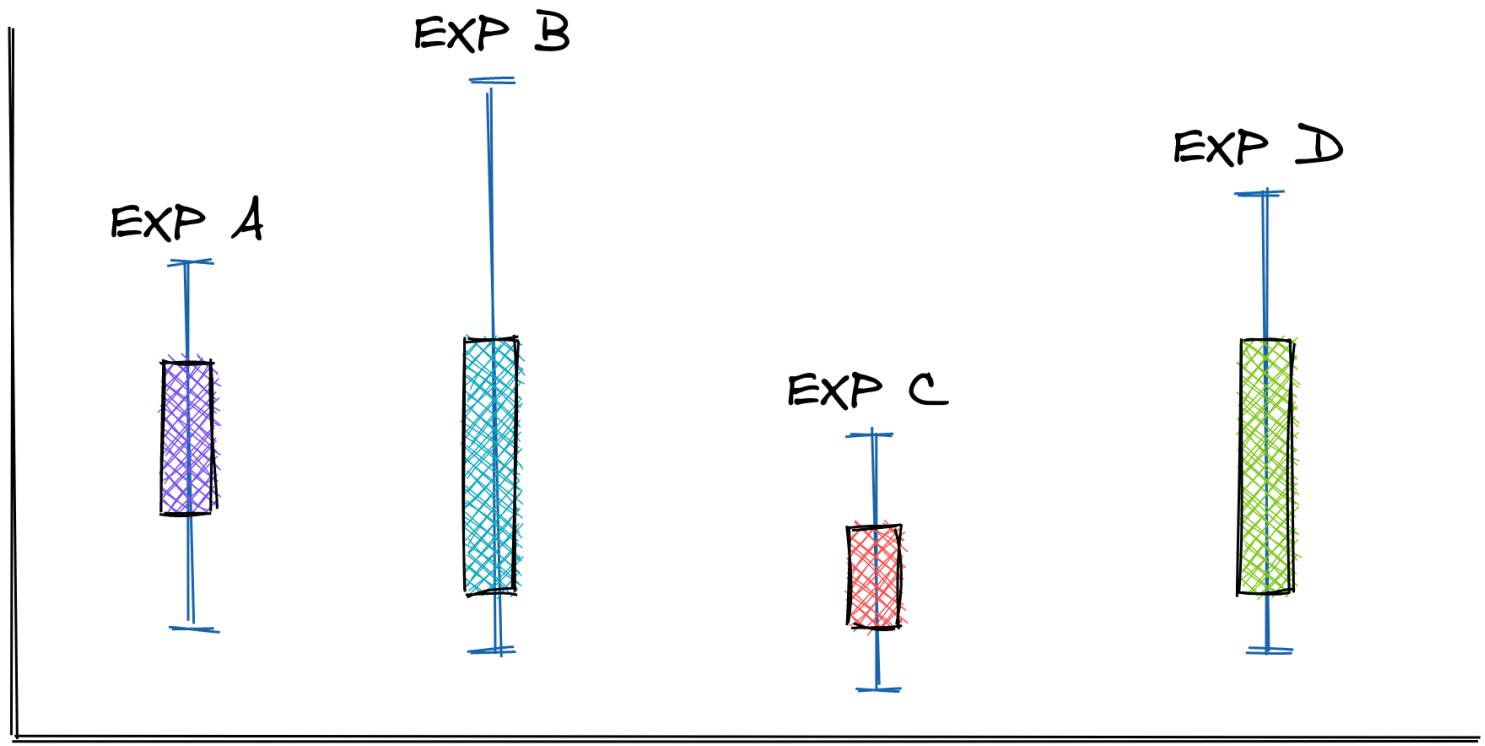
Training procces



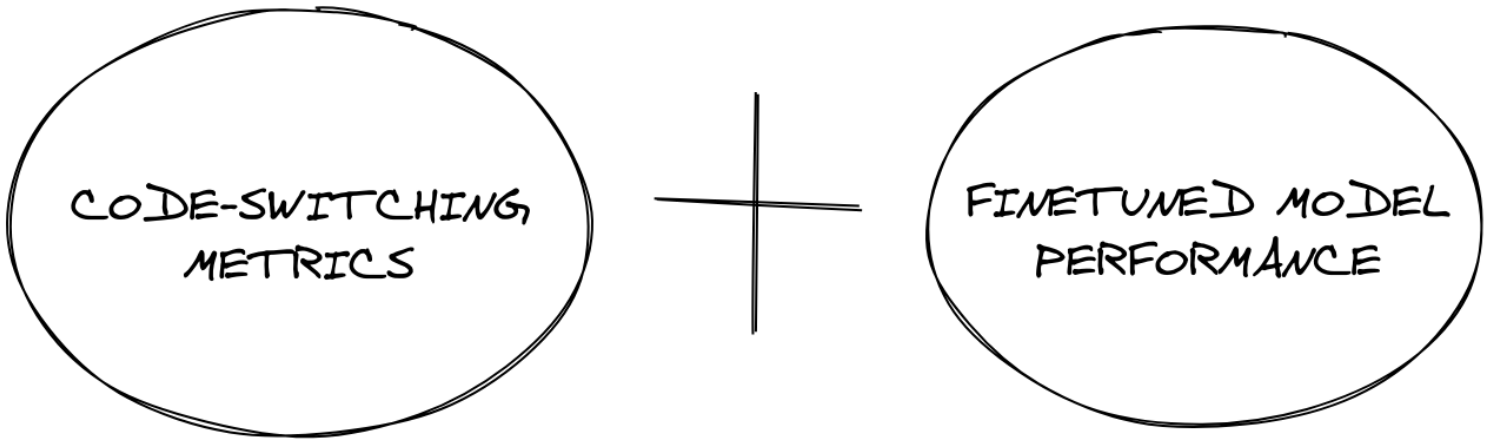
Training procces

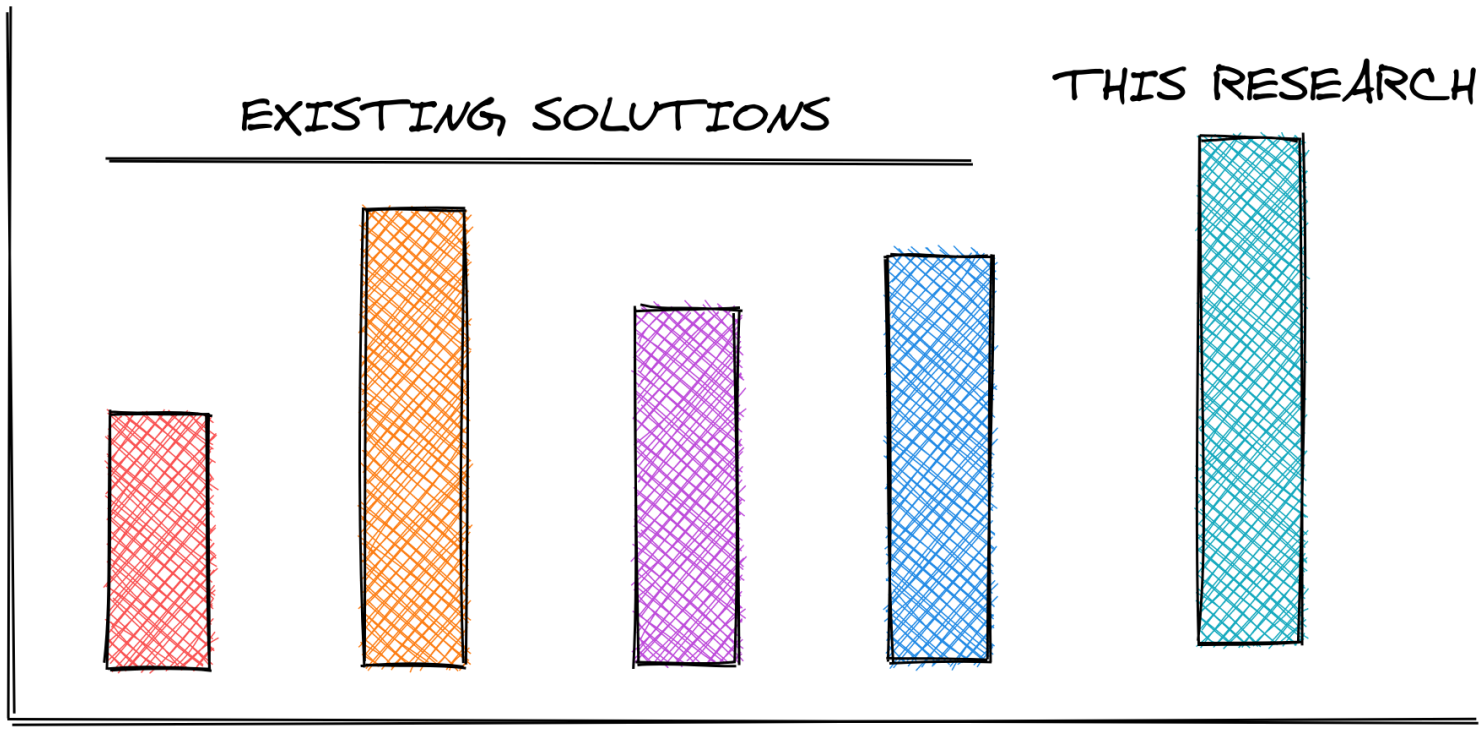


Experiments

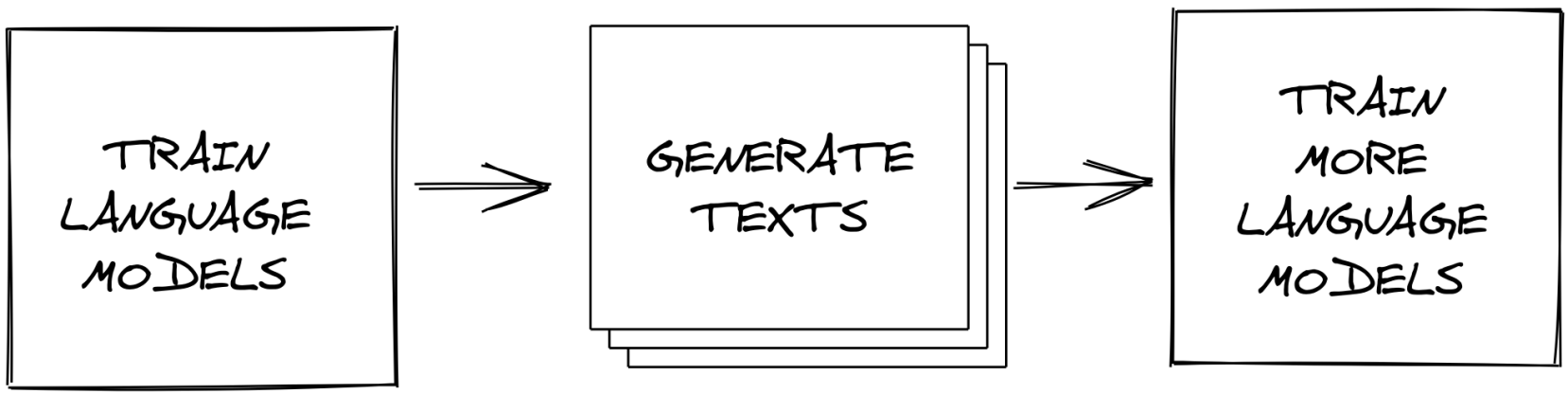


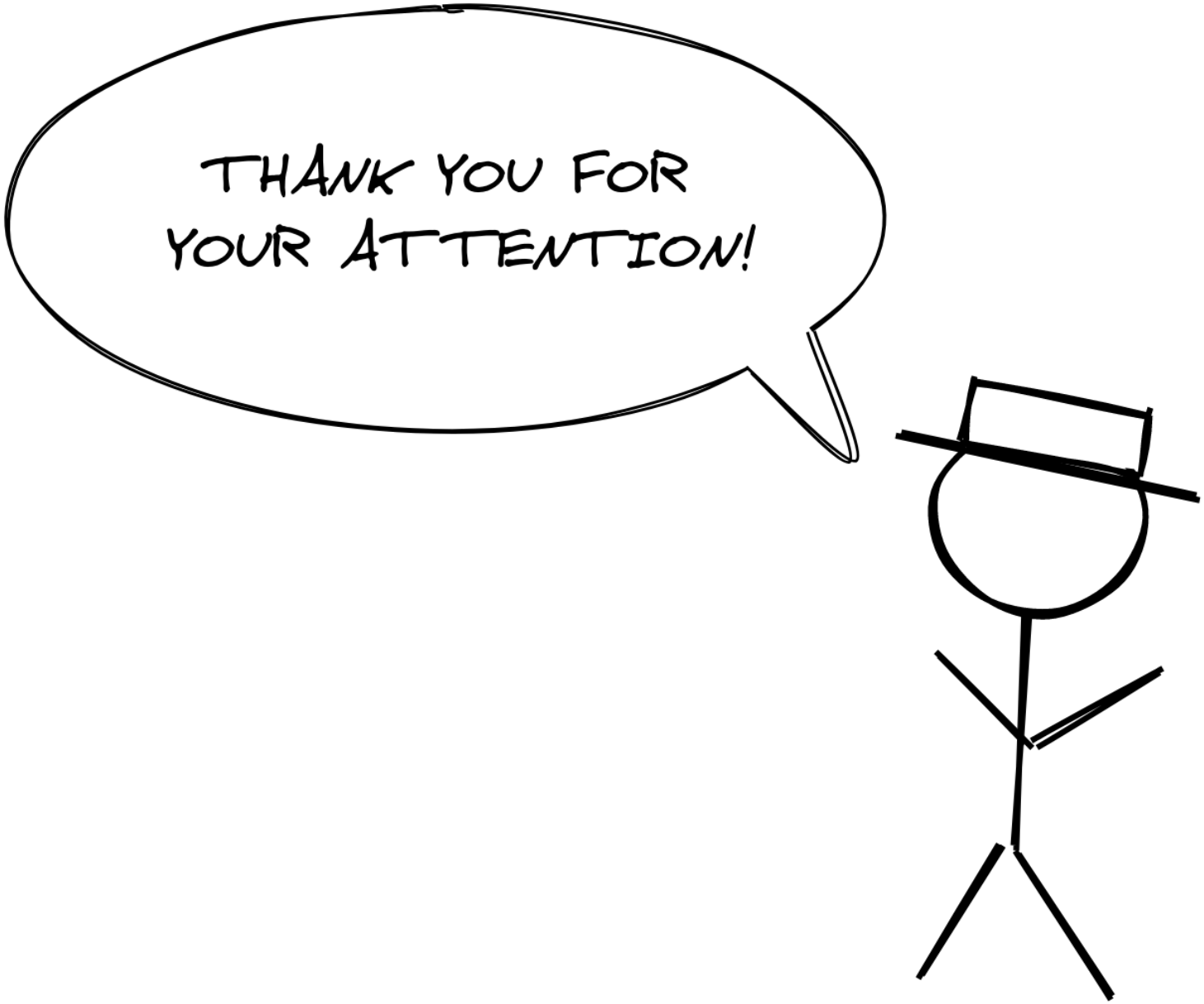
Evaluation




EXISTING SOLUTIONS
THIS RESEARCH

Future plans




THANK YOU FOR
YOUR ATTENTION!

1. "Attention Is All You Need." arXiv preprint arXiv:1706.03762v5.
2. "Differentiable Language Model Adversarial Attacks on Categorical Sequence Classifiers." arXiv preprint arXiv:2006.11078.
3. "A Semi-supervised Approach to Generate the Code-Mixed Text using Pre-trained Encoder and Transfer Learning." Findings of the Association for Computational Linguistics: EMNLP 2020, pages 2267–2280
4. "DialoGLUE: A Natural Language Understanding Benchmark for Task-Oriented Dialogue." arXiv preprint arXiv:2009.13570v2.
5. Guillaume Lample, Alexis Conneau. 2019. "Cross-lingual Language Model Pretraining." arXiv preprint arXiv:1901.07291.
6. "Unsupervised Cross-lingual Representation Learning at Scale." arXiv preprint arXiv:1911.02116v2.