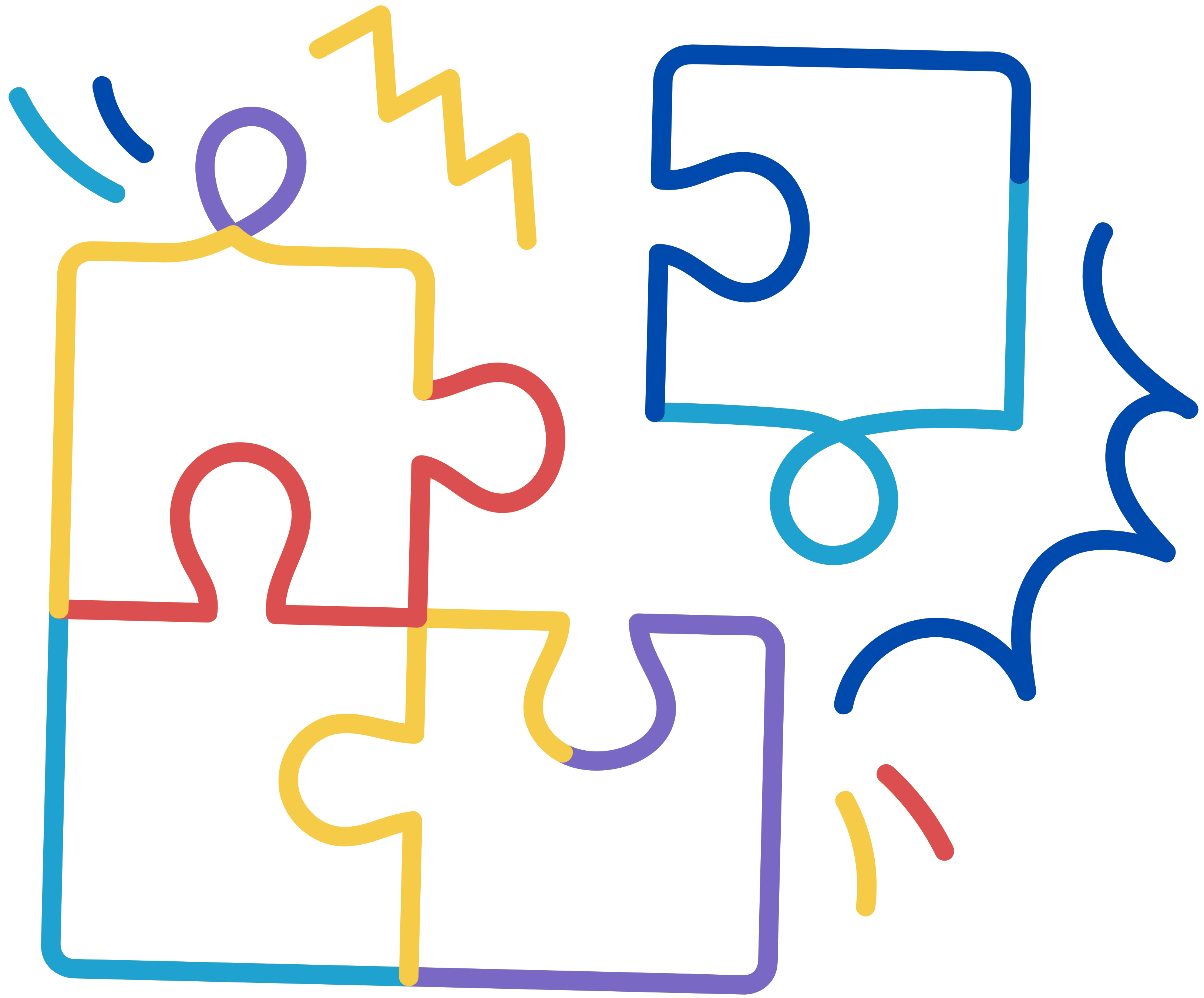
Hanoi University of Science and Technology
School of Information and Communication Technology

Fake News Detection

Course Code: IT3191E
Class Code: 156818
Machine Learning and Data Mining

Instructor: Ph.D. Nguyen Duc Anh



Group 1

Group Member

Ho Bao Thu - 20226003

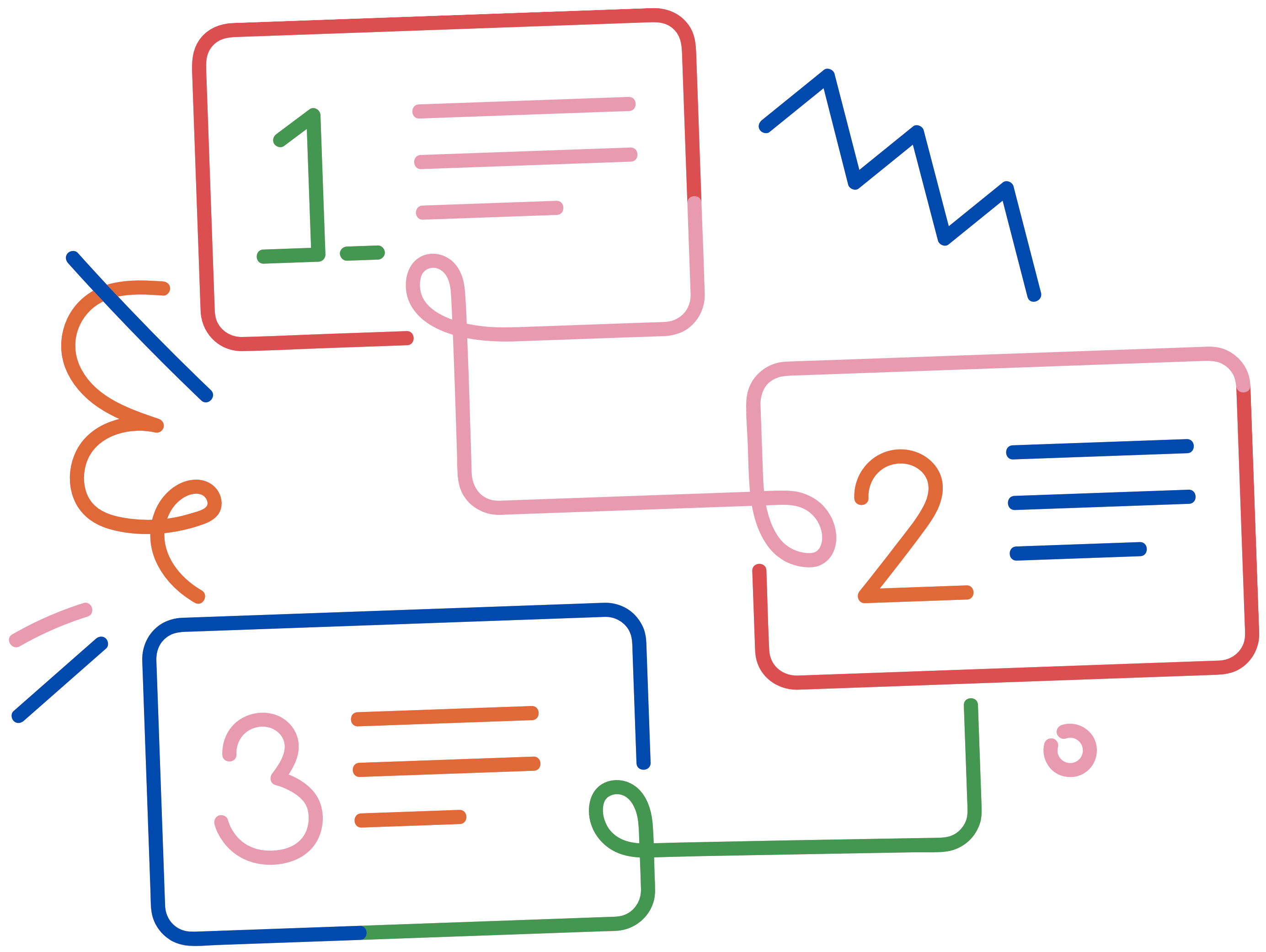
Tran Kim Cuong - 20226017

Nguyen Dinh Duong - 20225966

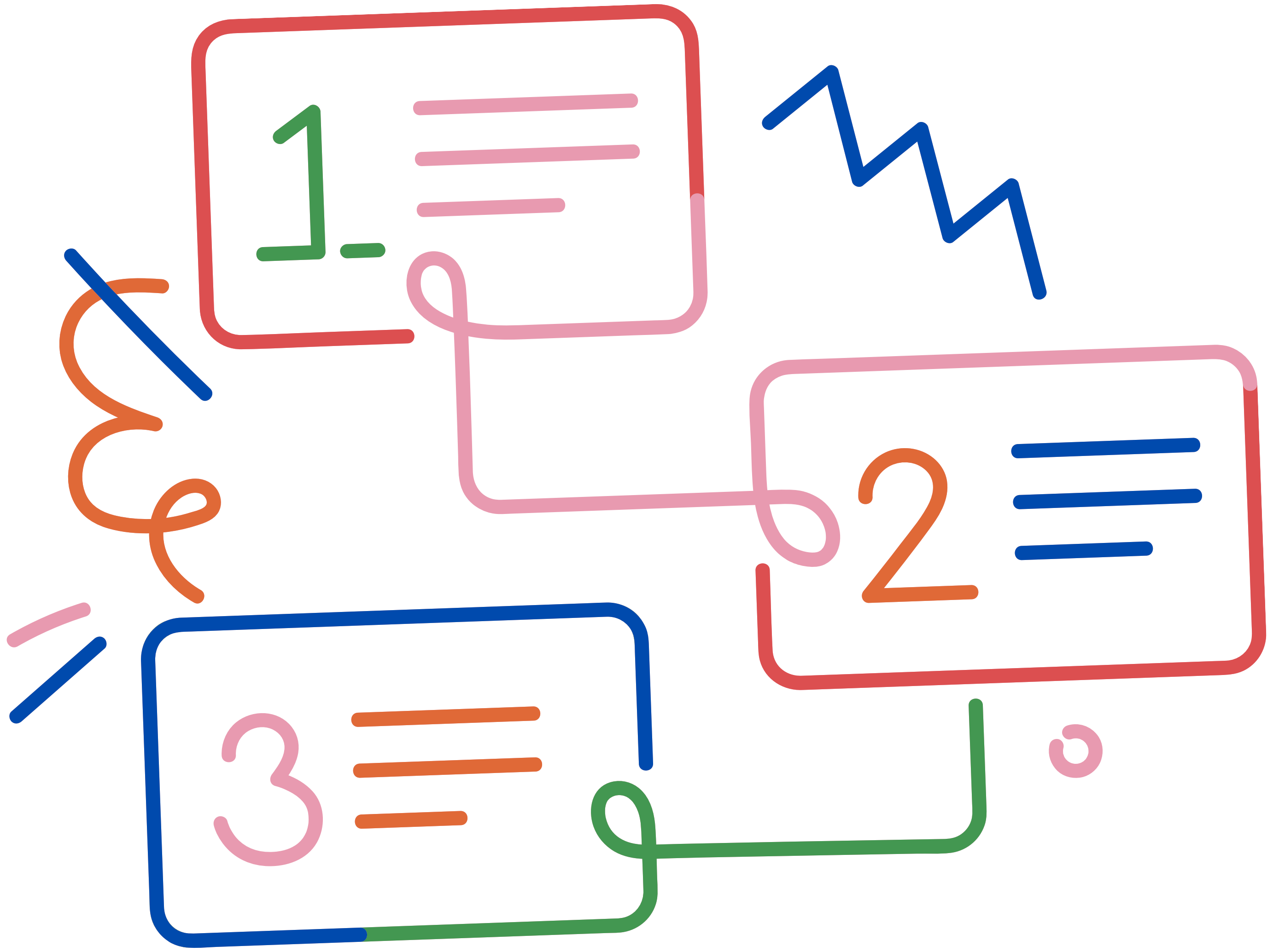
Nguyen My Duyen - 20225967

Ha Viet Khanh - 20225979

Dang Van Nhan - 20225990



Content



01. Introduction
02. Related Work
03. Dataset Survey & Selection
04. Methodology
05. Experiments
06. Conclusions

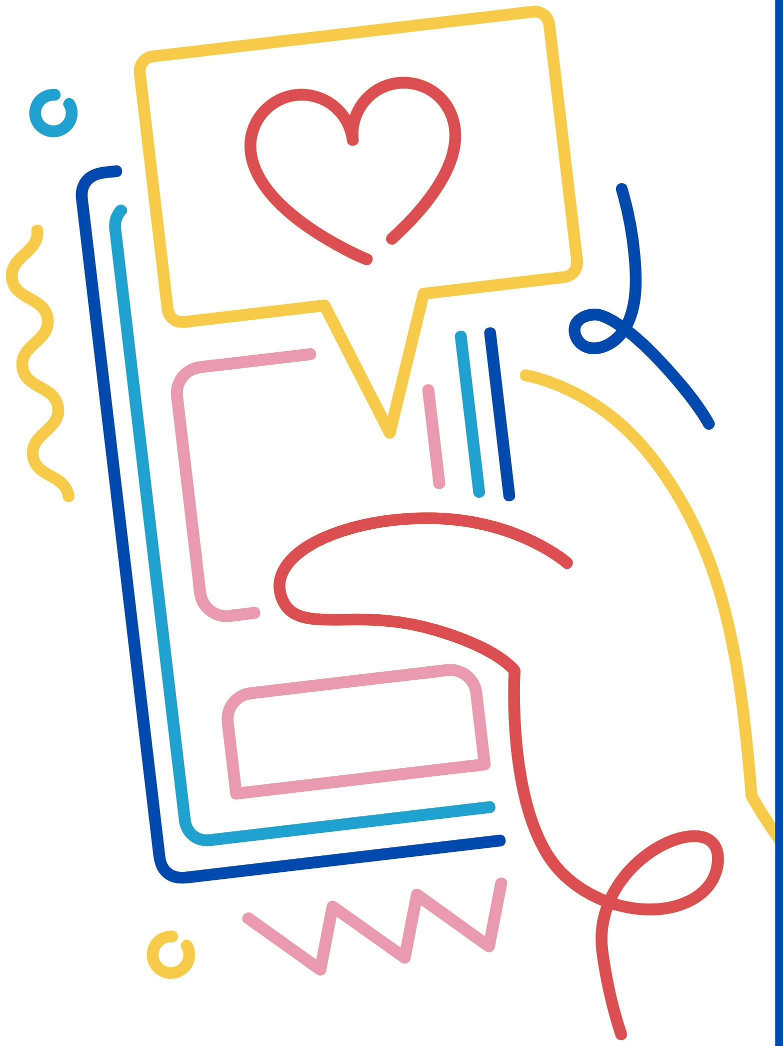
INTRODUCTION

REAL or FAKE?

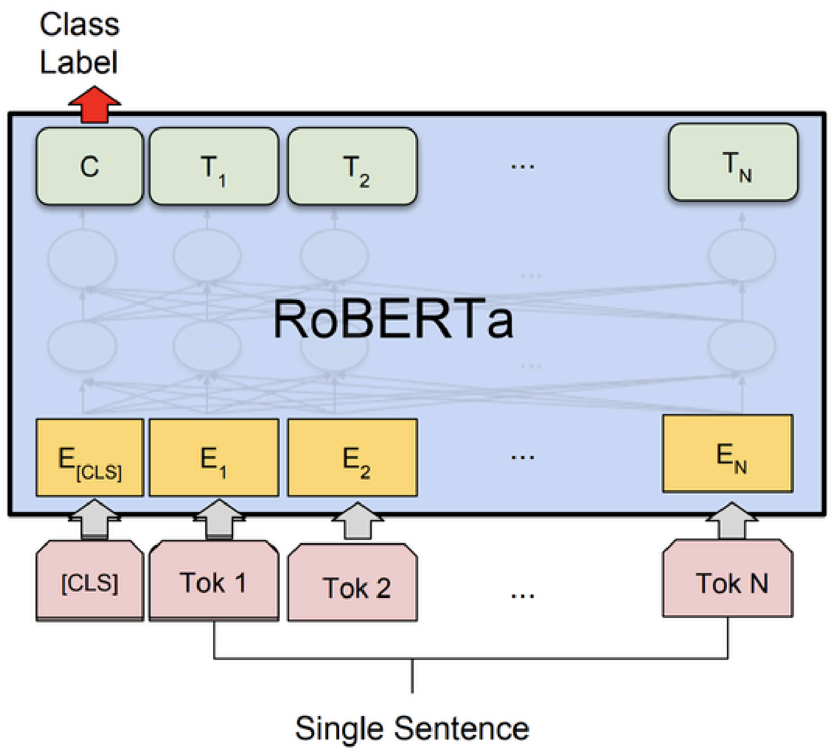
Social media evolution

News' spreading instantly

High risk of fake information



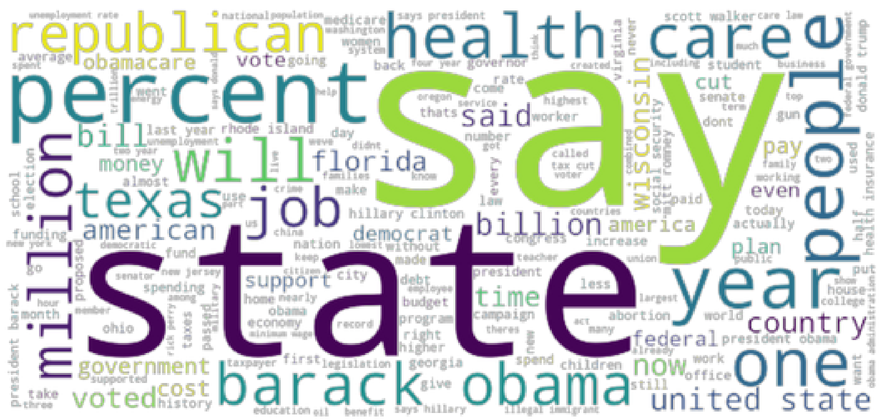
Introduction

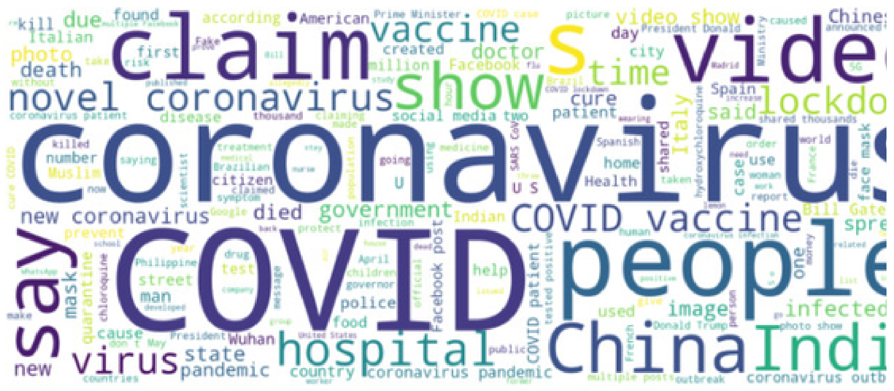


“Advanced” BERT

FakeNewsNet

LIAR





COVID-19
Fake News

Datasets

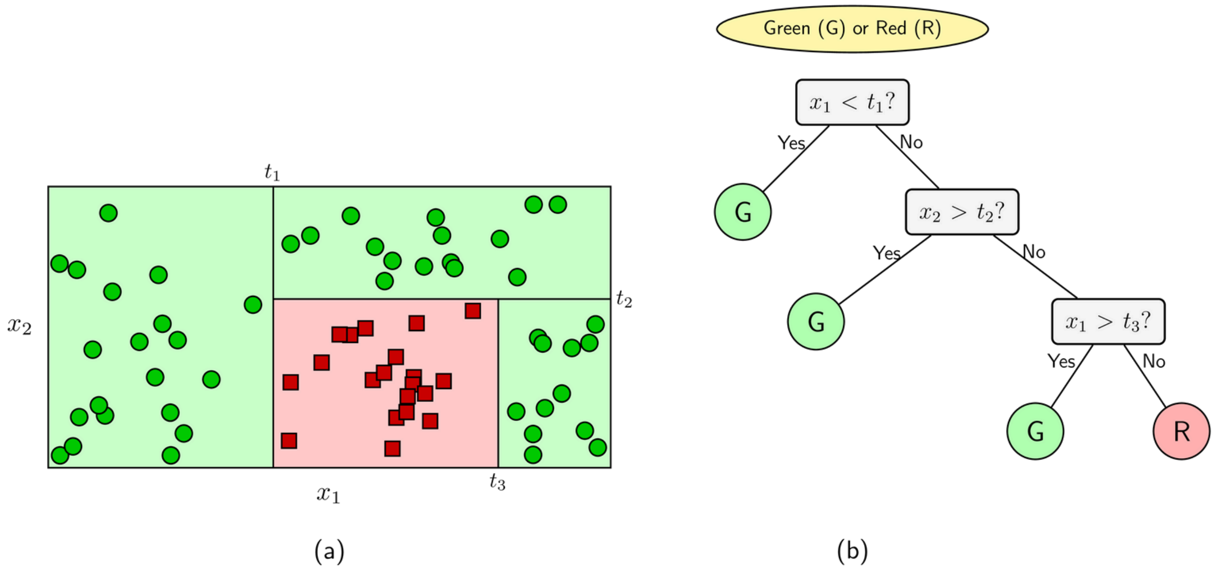
2 RELATED WORK

Deep Learning methods

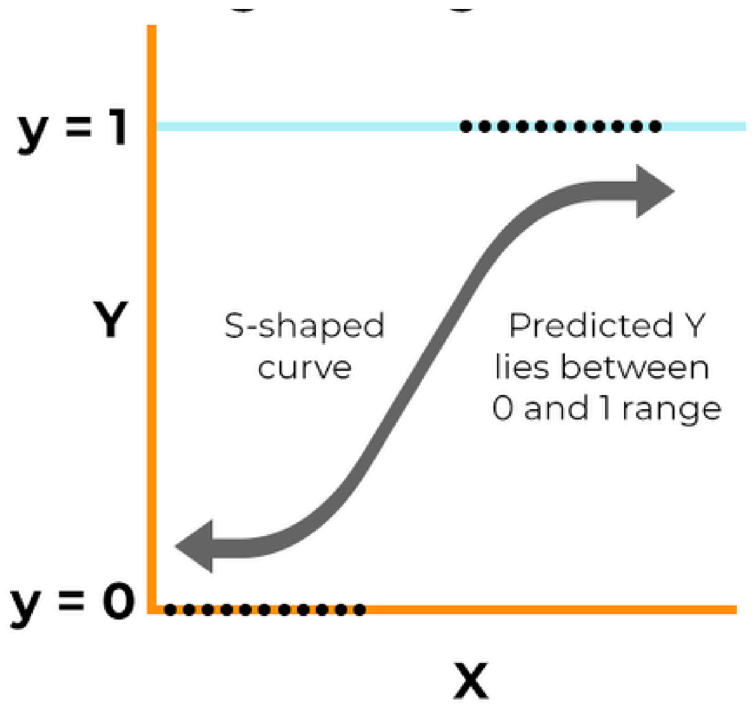
Traditional: ML to Neural Networks

Advanced: Transformer-based

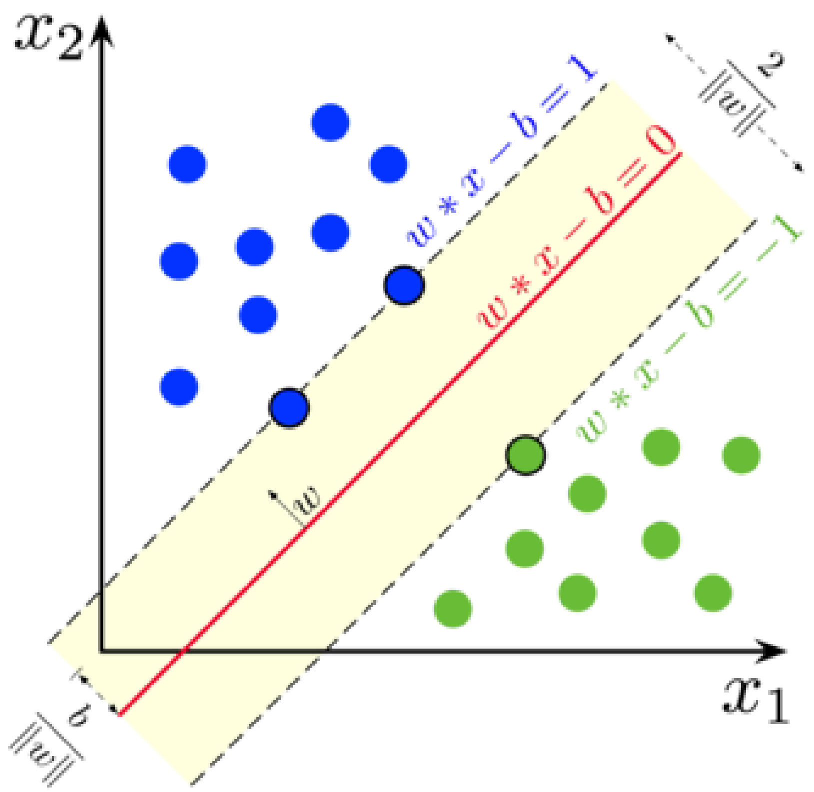
Traditional Methods



Decision Tree



Logistic Regression

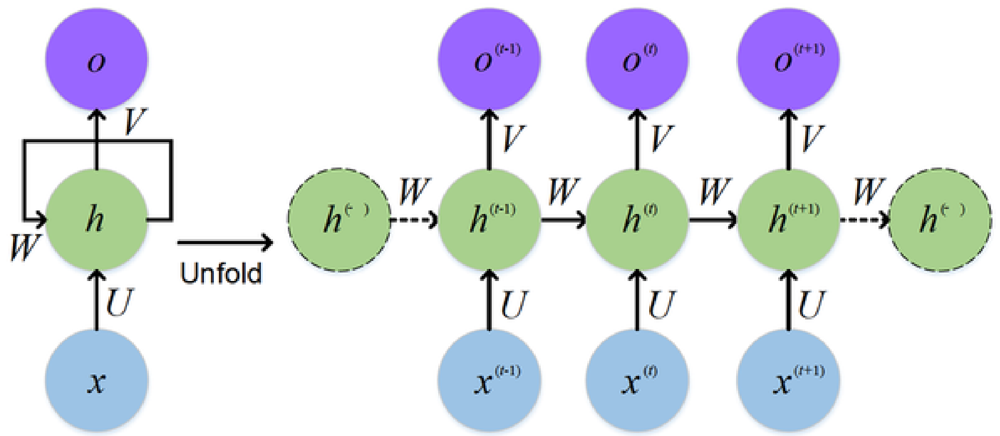


SVM

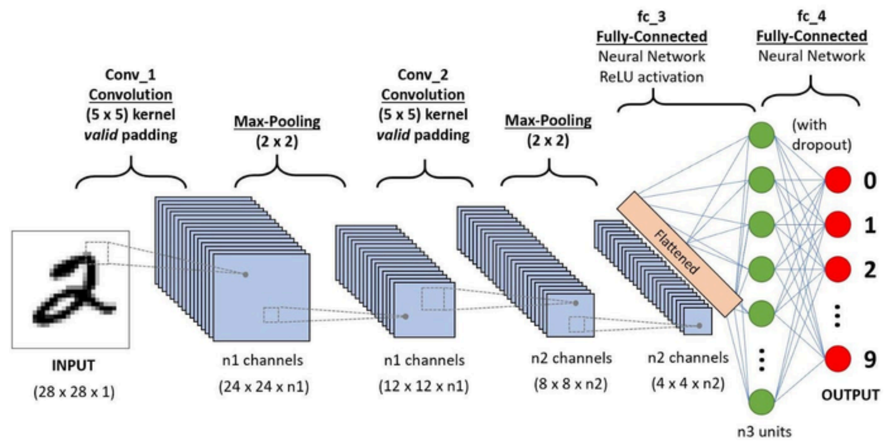
Rely on shallow features, failing to capture deep semantics, generalize to new fake-news tactics.

Traditional Methods

RNN

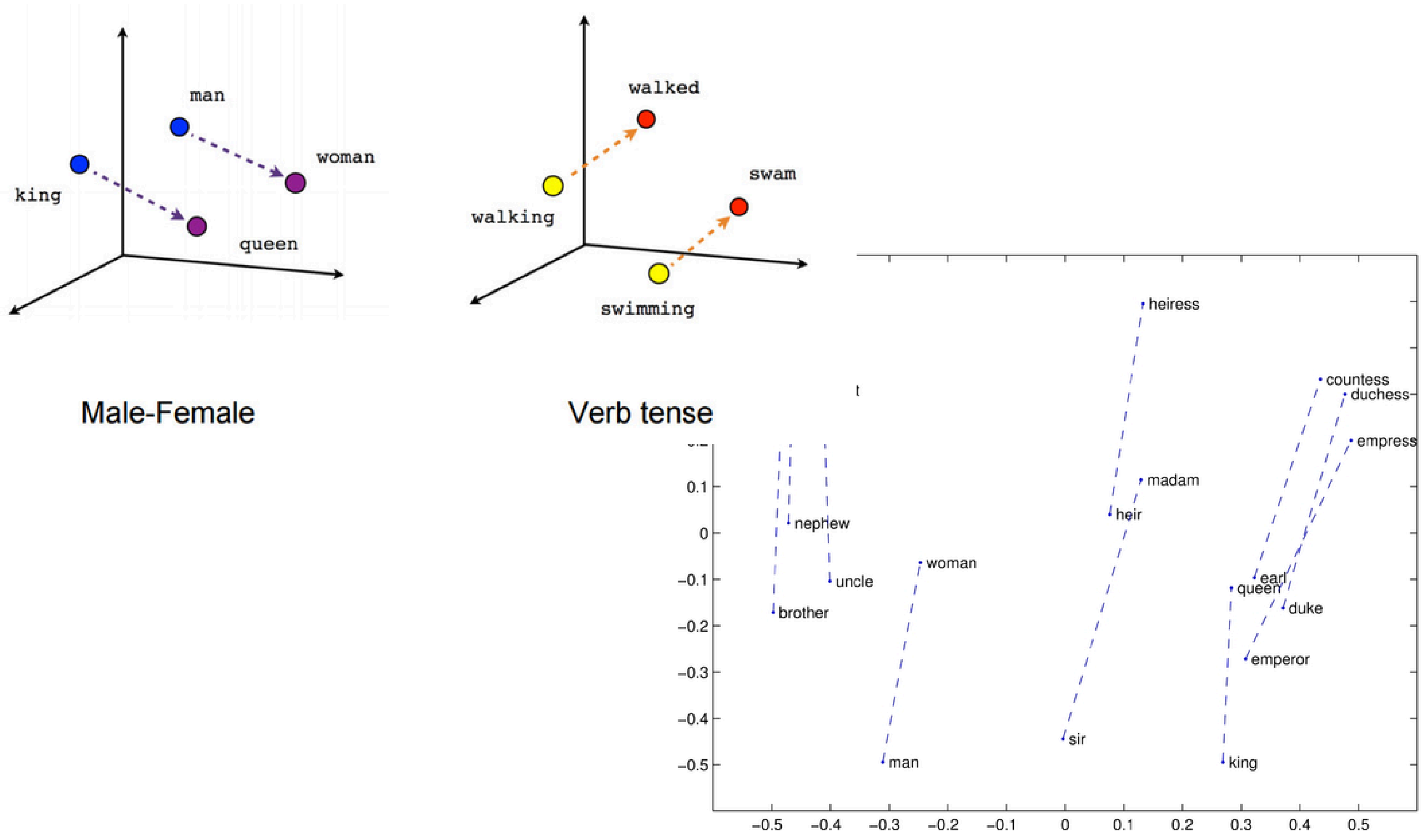


CNN



LSTM, Bi-LSTM, ...

use



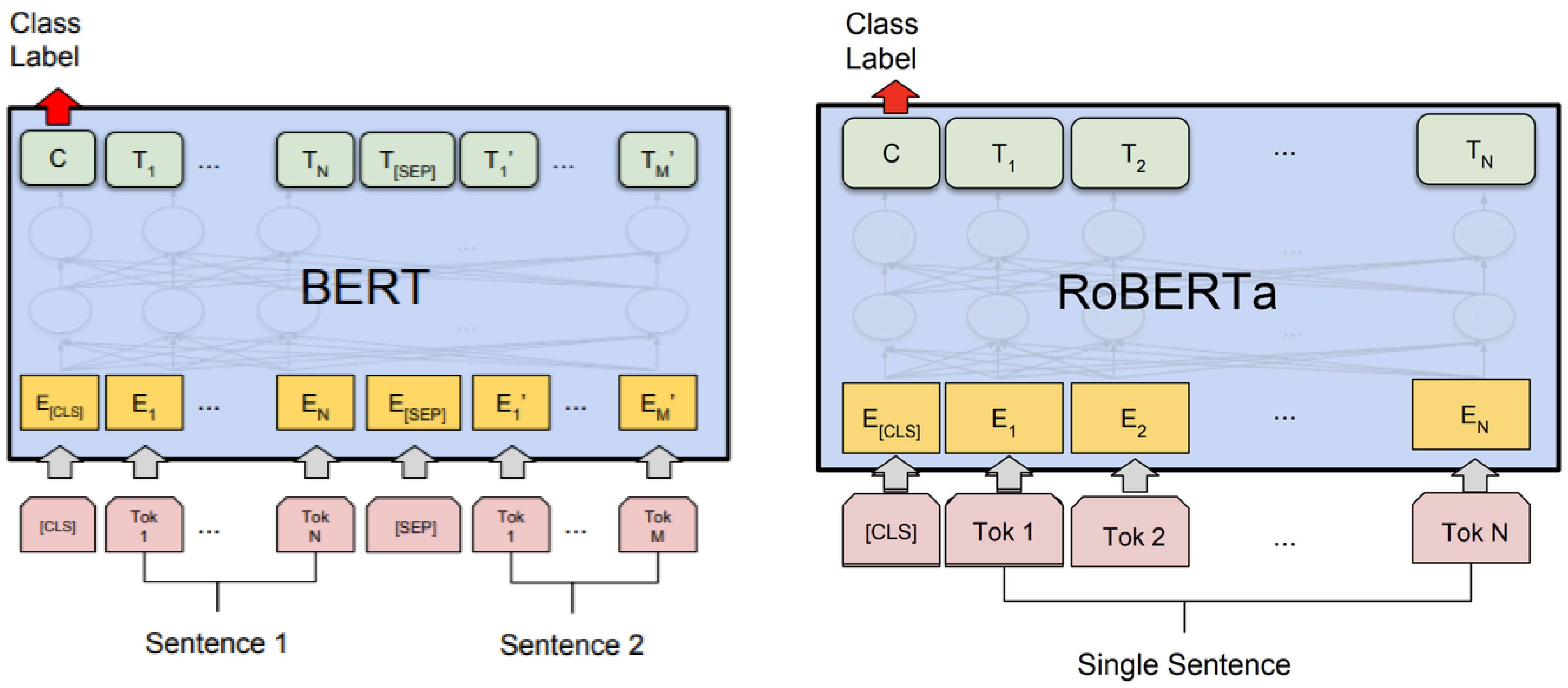
Embedding (Word2Vec, GloVe,...)

Limited to text, poor at generalizing.

Transformer-based

“Attention Is All You Need”*

Especially



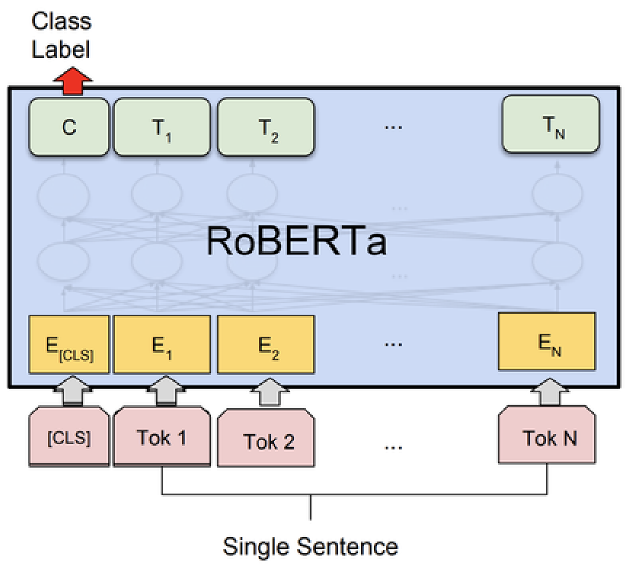
Capture long-term dependencies
Rich contextual information in text!

*Source: Attention Is All You Need (in *Advances in Neutral Information Processing Systems*). 2017

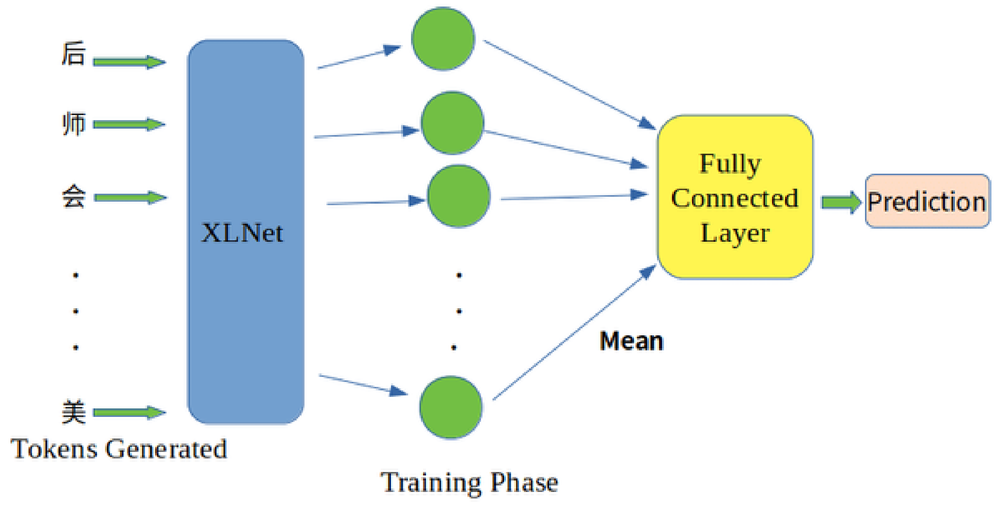
Advanced Methods

Lack of comparison!

RoBERTa



XLNet



Evaluate generalizability, and develop lightweight, efficient models for practical applications.

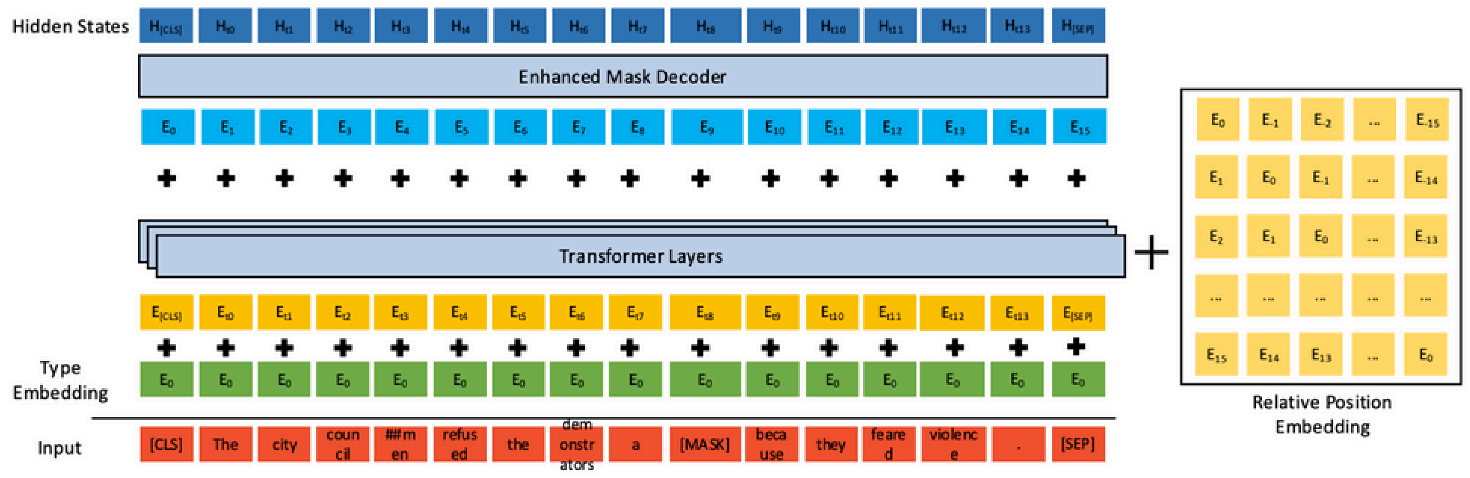
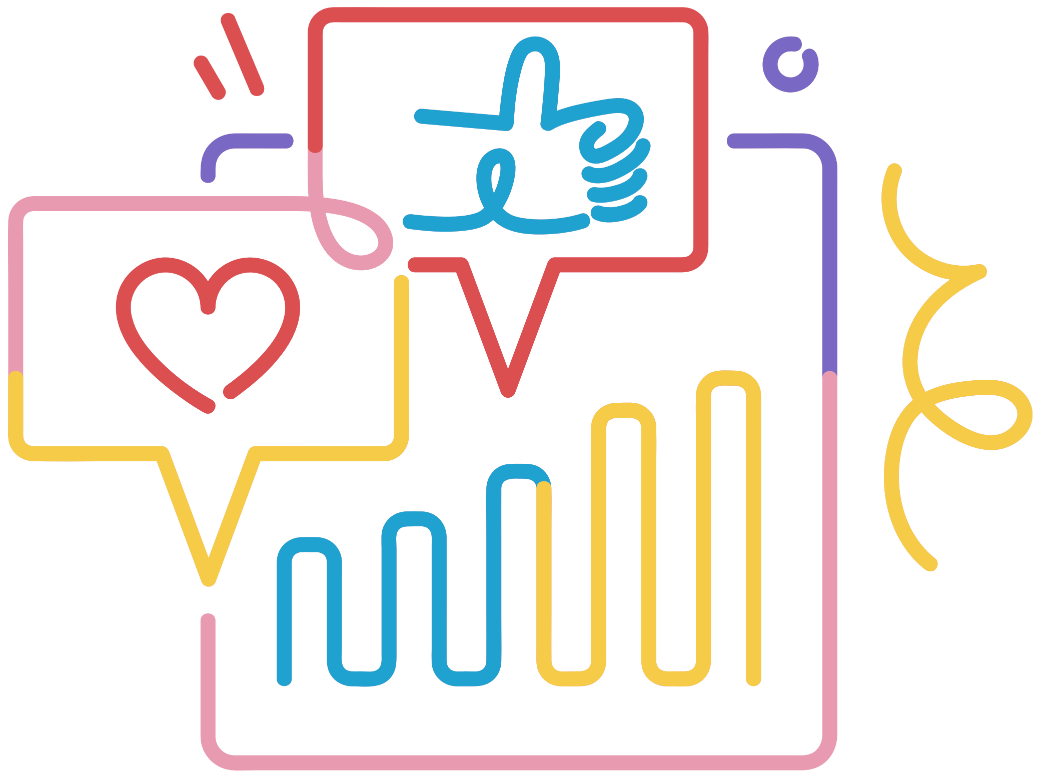


Fig 1. The model architecture of DeBERTa

DeBERTa

DATASET
SURVEY & SELECTION



3

Overview

We selected only **public, near-balanced datasets** to ensure reproducibility.

Datasets like Verification Corpus, MisInfoText, BuzzFace, FacebookHoax, r/Fakeddit, and CREDBANK were excluded due to access issues or severe label imbalance.

COVID-19 Fake News

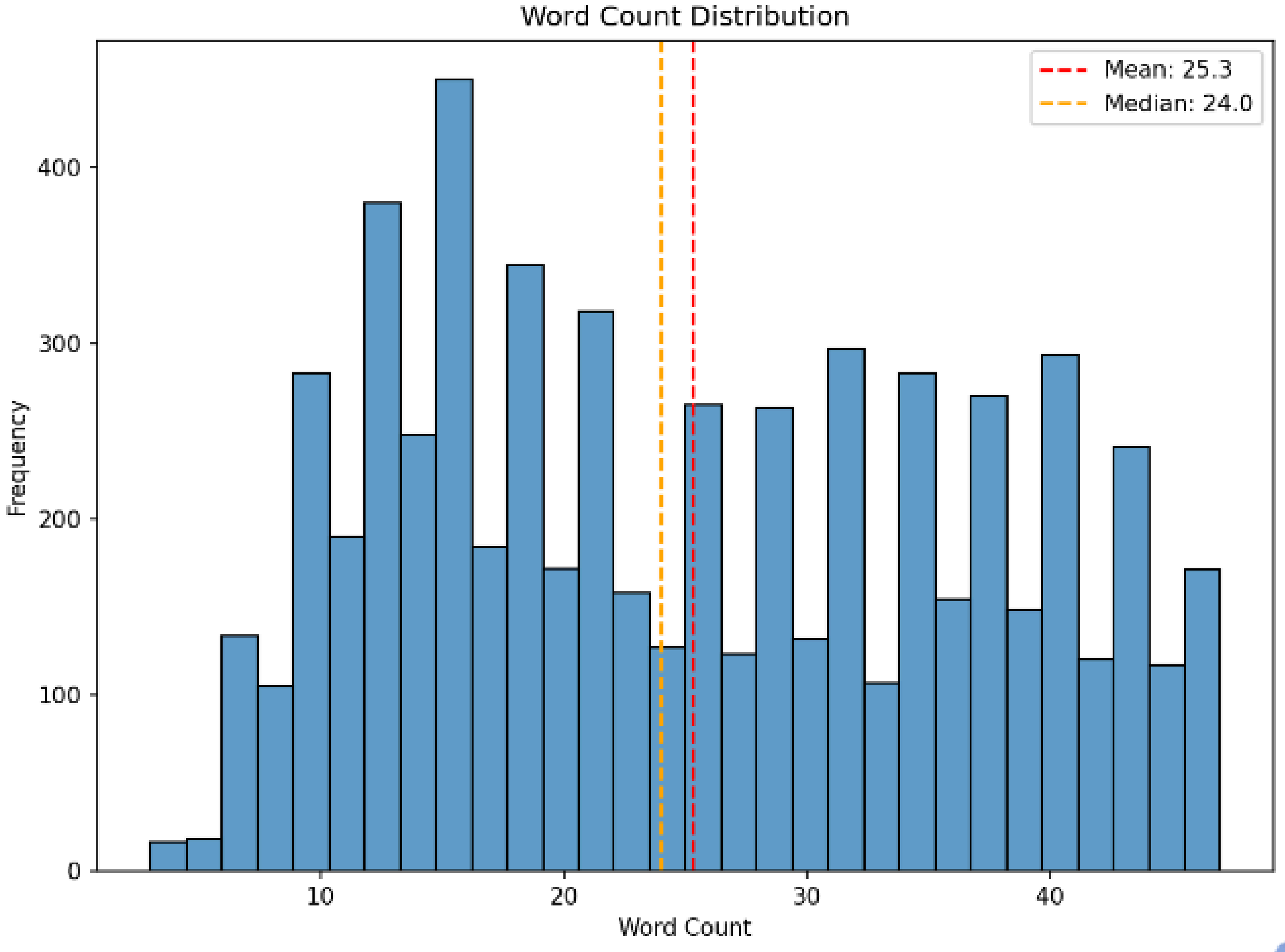
FakeNewsNet

LIAR

COVID-19 Fake News

- **10,700** labeled samples
- Labels are binary: “**real**” or “**fake**”
- Designed to **support misinformation detection** in health crises.

Subset	Samples	Real	Fake
Train	6,420	3,360	3,060
Validation	2,140	1,120	1,020
Test	2,140	1,120	1,020
Total	**10,700**	**5,600**	**5,100**



FakeNewsNet

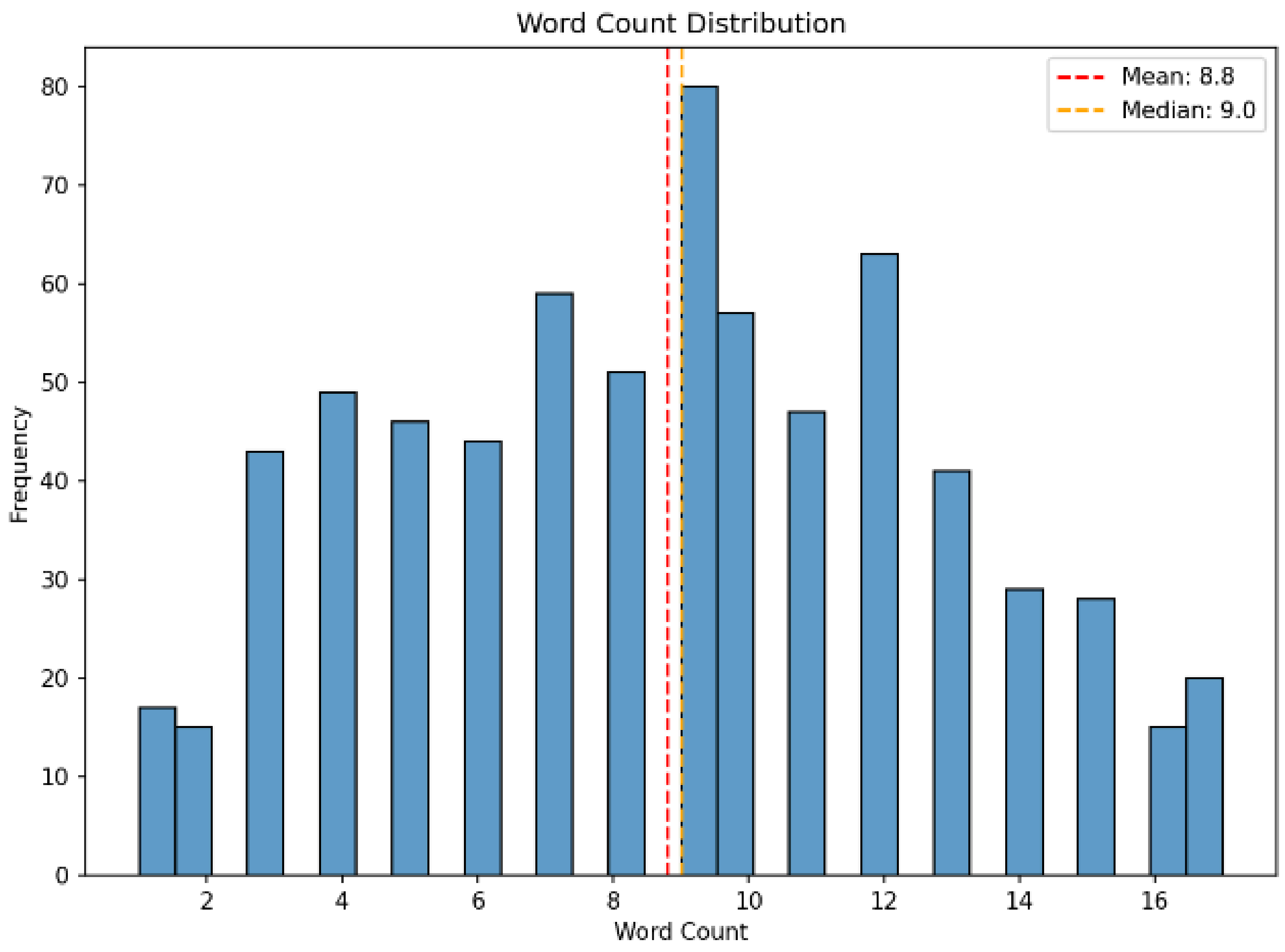
- FakeNewsNet is covering two domains:
  - **GossipCop** (entertainment & celebrities)
  - **PolitiFact** (political claims)

- Each sample has id, news_url, title, and tweet_ids.
- Supports text and social media analysis with consistent, non-null titles.

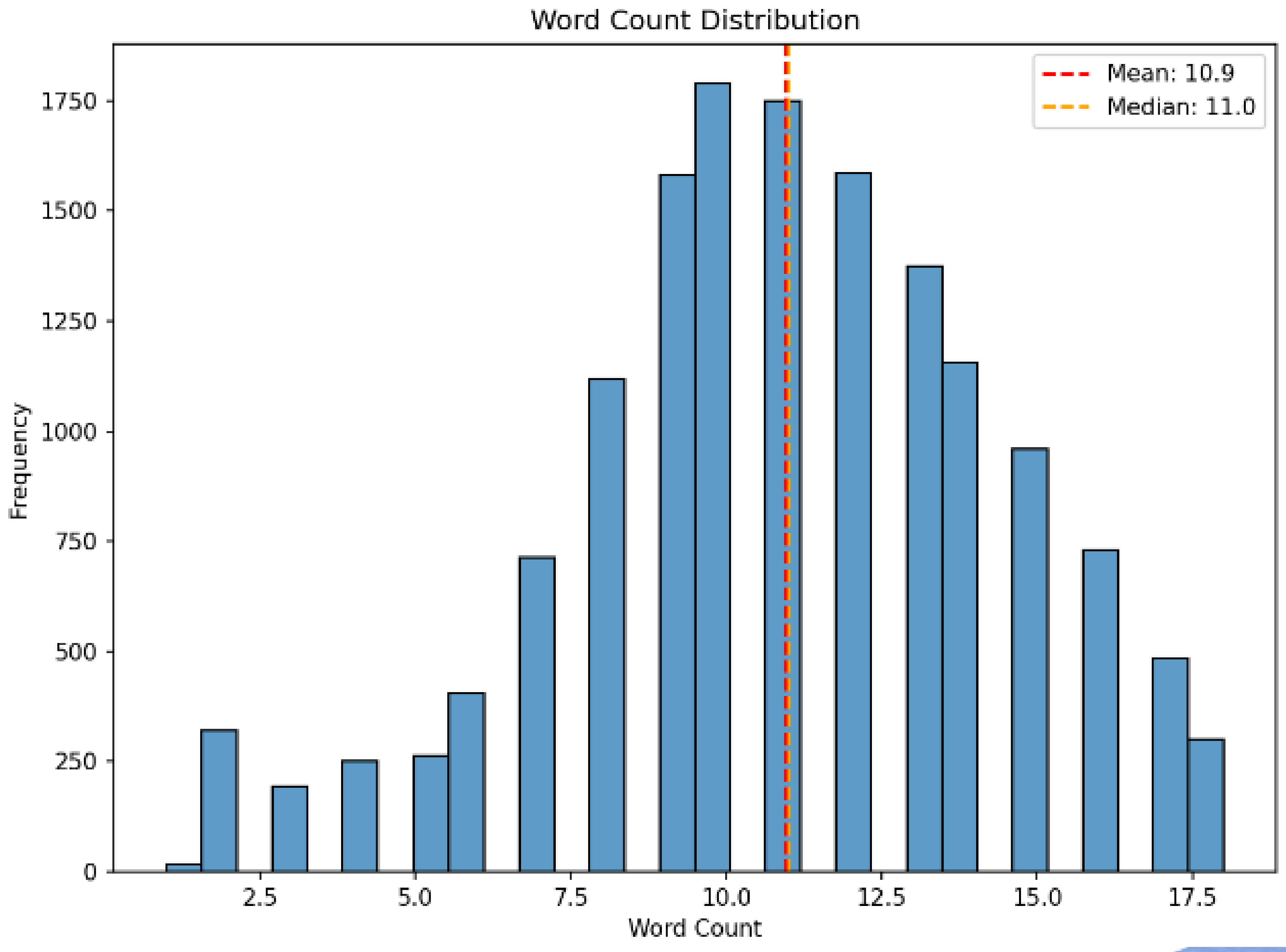
Subset	Samples	Real	Fake	Domain
Train	15,498	11,772	6,602	GossipCop
Validation	3,321	2,523	864	
Test	3,321	2,522	818	
Train	739	437	302	PolitiFact
Validation	158	93	65	
Test	159	94	65	
Total	**23,196**	**16,441**	**6,755**	**Both**

FakeNewsNet

PolitiFact



GossipCop

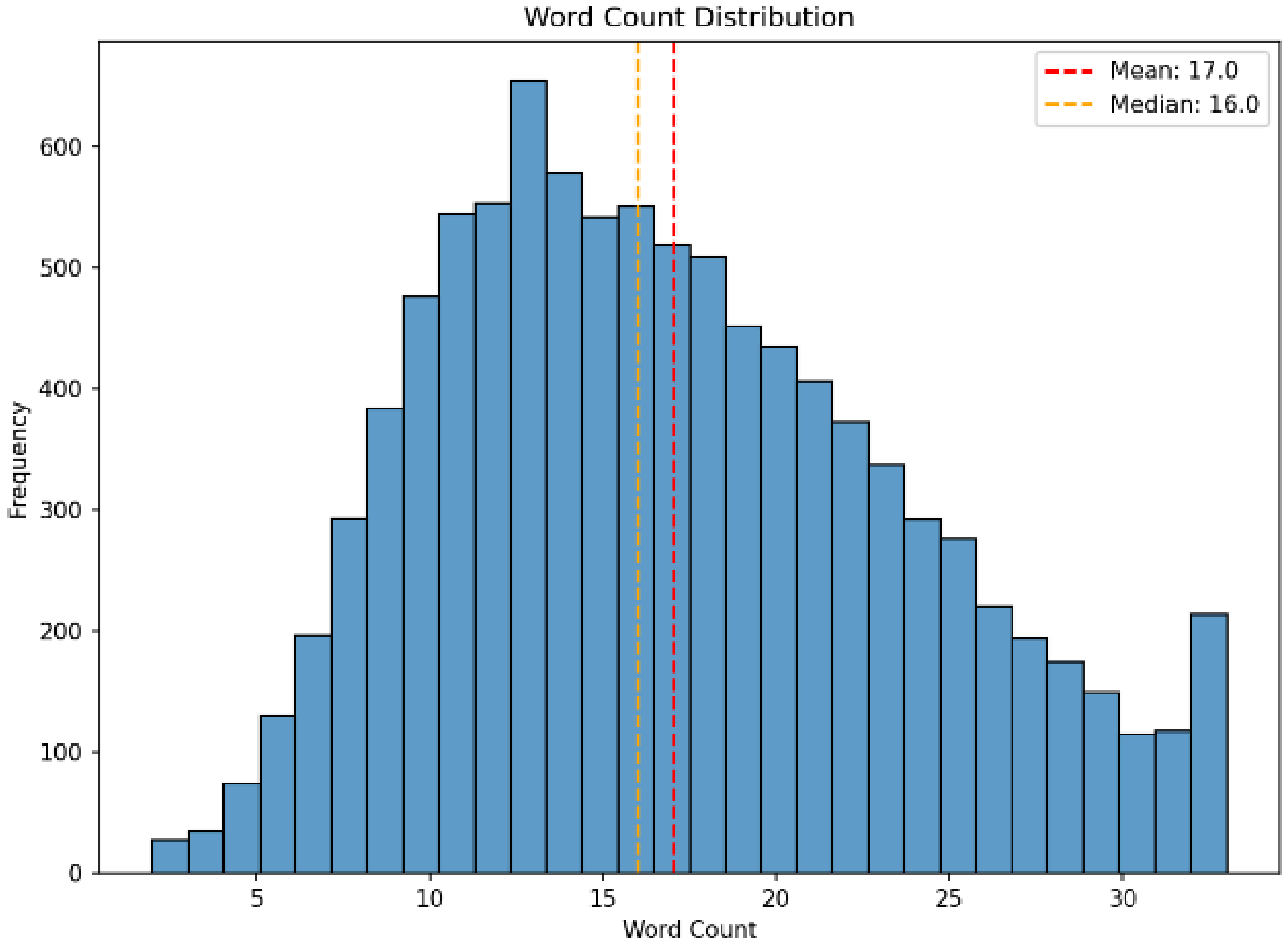


LIAR

- 13,791 fact-checked **political** statements
- Labels are grouped into two classes:
  - Real: true, mostly-true
  - Fake: half-true, barely-true, false, pants-fire

- **Only the statement field is used**; additional metadata includes speaker, topic, party, and context.

Subset	Samples	Real	Fake
Train	10,240	3,638	6,602
Validation	1,284	420	864
Test	1,267	449	818
Total	**12,791**	**4,507**	**8,284**



4

METHODOLOGY

BERT: Architecture

BERT's **model architecture** is a multi-layer bidirectional Transformer encoder based on the original implementation

- **Multi-head Attention:**

Each attention head learns distinct patterns and enhances the ability to capture complex dependencies.

- **Residual Connection** (Skip Connection):

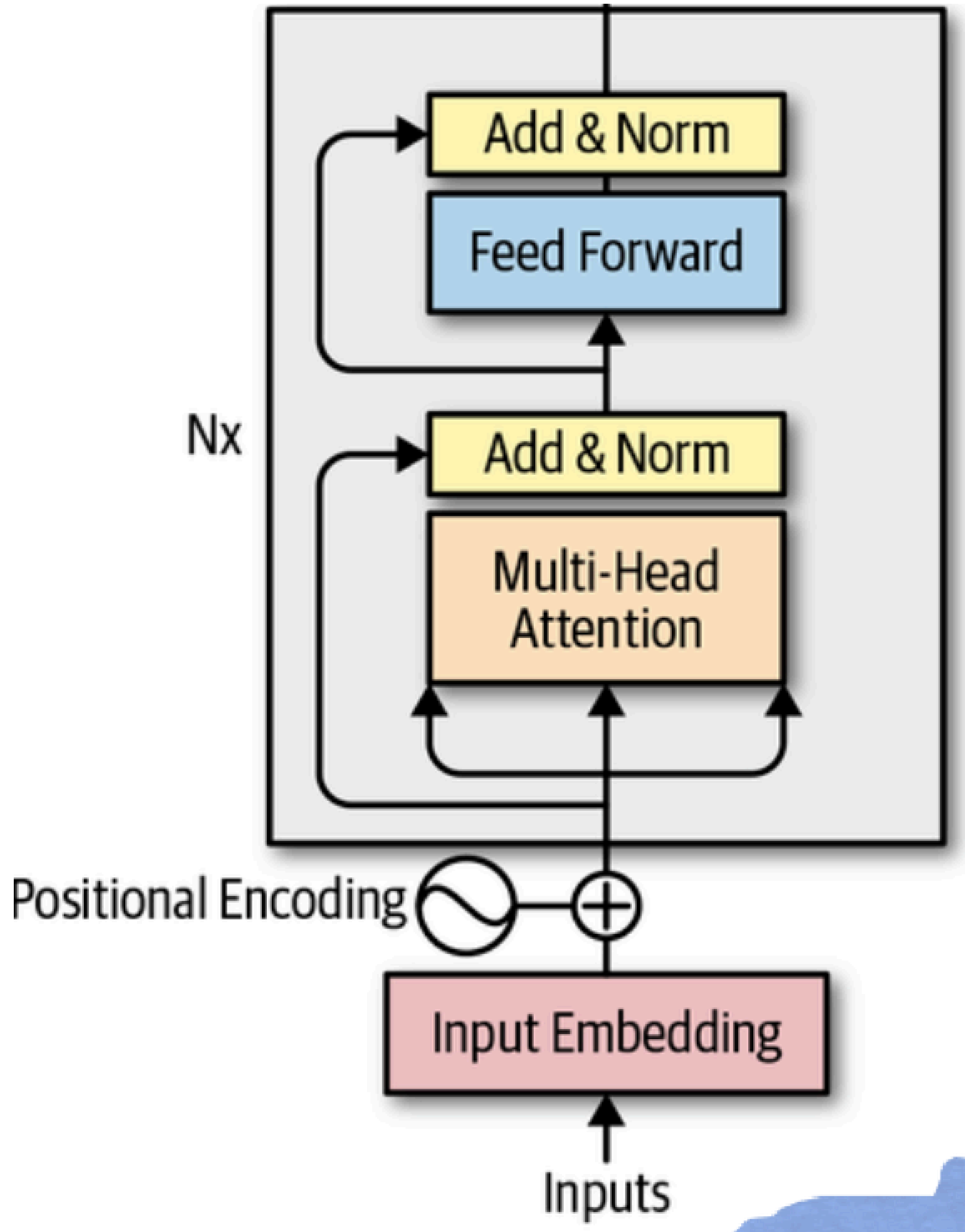
Preserves the original input by adding it to the output of sub-layers and improves training stability and convergence in deep networks.

- **Add & LayerNorm:**

Uses the LayerNorm operation to normalize across hidden dimensions and stabilizes learning process

- **Feed-Forward Network (FF):**

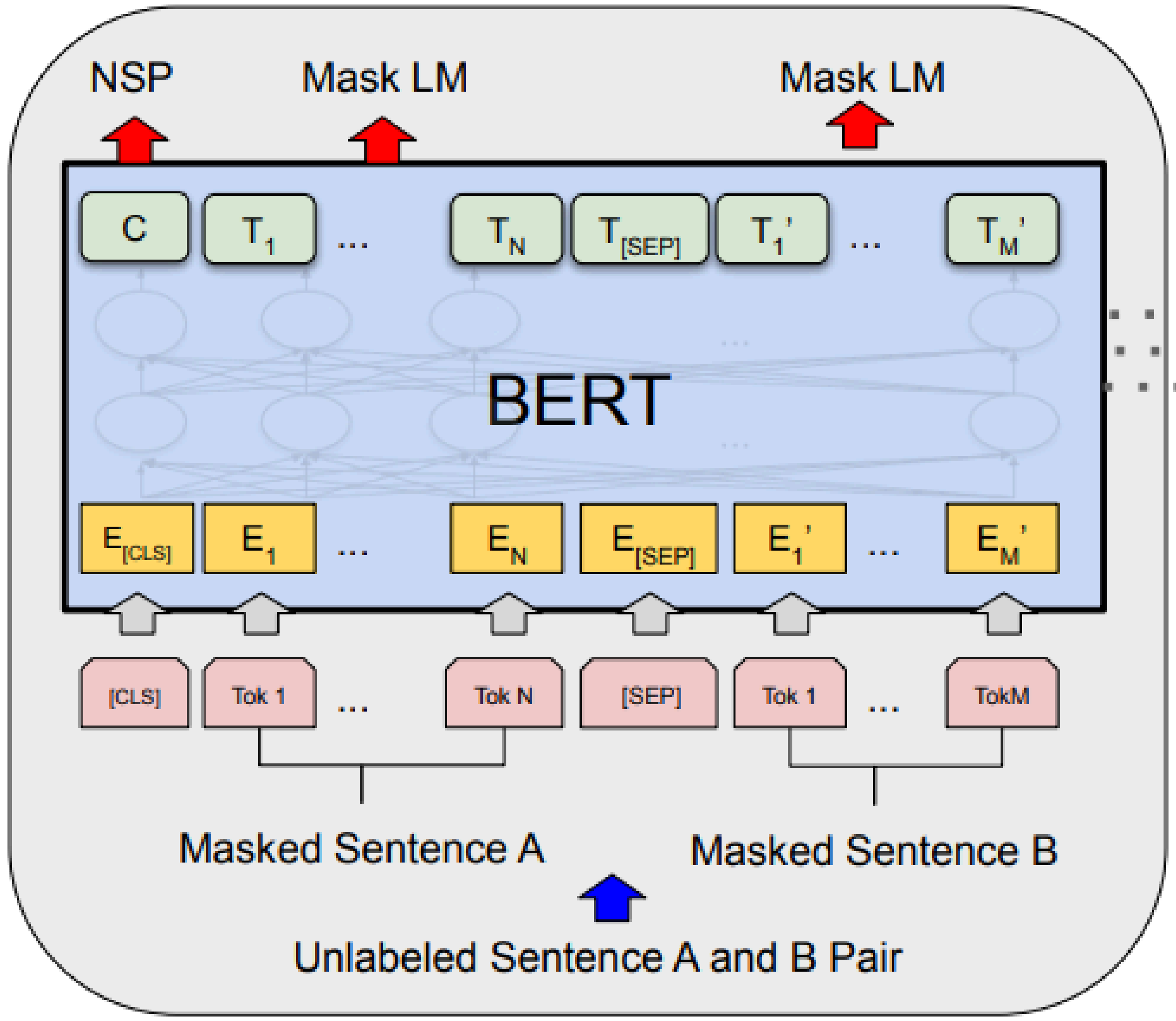
Introduces non-linearity (via ReLU or GELU) to enhance representation power.



BERT: Pretraining

Train BERT on unlabeled text data so it can learn general-purpose contextual language representations. Randomly mask 15% of the tokens in the input sequence. Two objectives:

- **Masked Language Modeling (MLM)**
- **Next Sentence Prediction (NSP)**



Pre-training

RoBERTa

Architecture: Similar to BERT

Enhancement:

- **Dynamic Masking**
- **Elimination** of **Next Sentence Prediction (NSP)** Task
- **Larger** Batch Sizes and **Higher** Learning Rates
- **Larger** Training Data and **Extended** Training Time
    - DataBookCorpus + English Wikipedia (similar to BERT)
    - CC-News + OpenWebText + Stories

DeBERTa

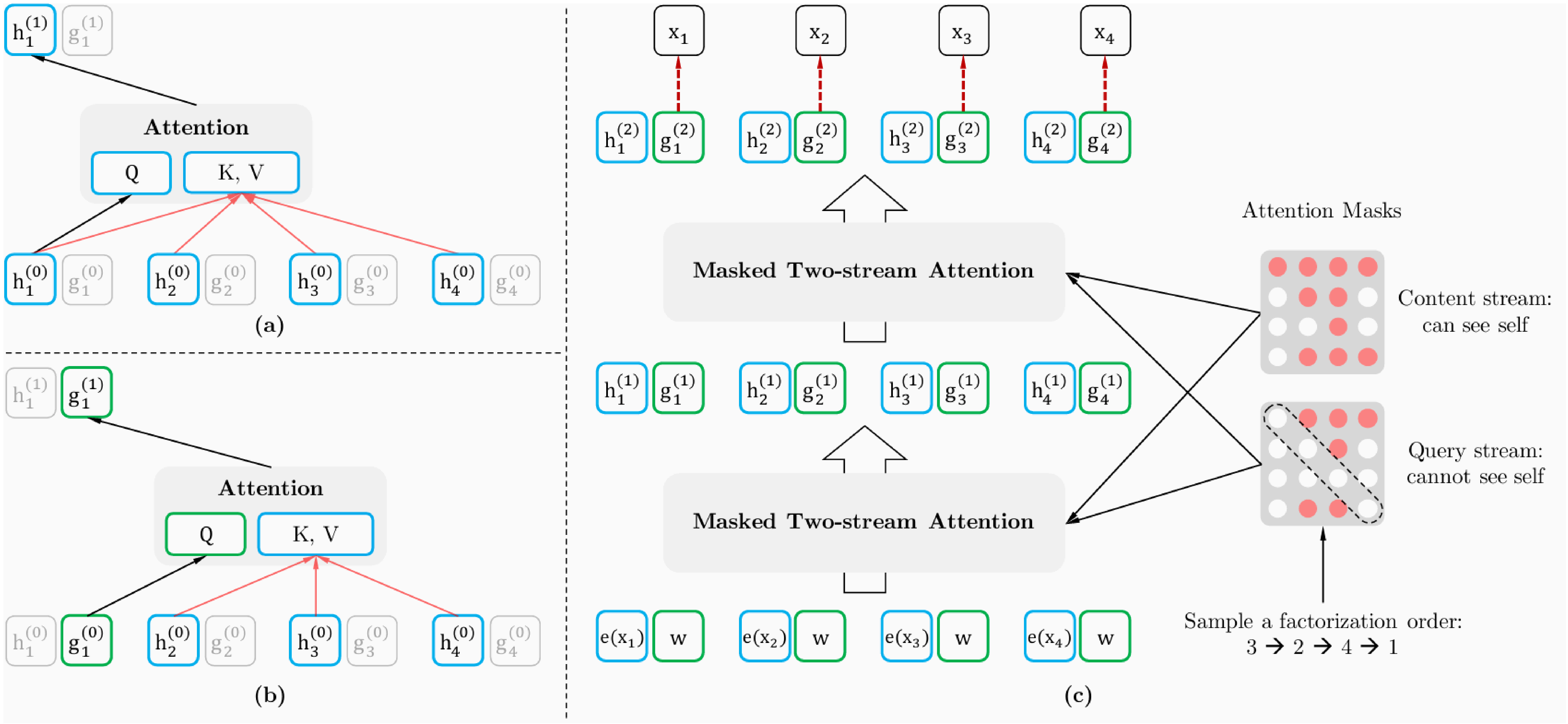
Architecture: Similar to BERT, RoBERTa

Enhancement:

- **Disentangled Attention Mechanism**
- **Enhanced Mask Decoder (EMD)**
- **Virtual Adversarial Training (VAT)**
- **Efficient** Pretraining

XLNet

Architecture: the overall structure of each Transformer layer remains unchanged; only the attention mechanism inside each head is modified



XLNet

Two-Stream Self-Attention

- **Content Stream:** The content stream behaves like standard Transformer attention, incorporating both context and the current token.
- **Query Stream:** the query stream uses only the position and prior context, explicitly excluding the current token's content.

Permutation Order

Instead of predicting tokens in a fixed **left-to-right order** (as in GPT), it randomly permutes the factorization order during training. This allows each token to be predicted given a context that may include tokens both to its left and right.

Partial Prediction

Limits training to only predicting the last few tokens in each permutation order—those with the richest context.

Reported Comparision

SQuAD2.0	EM	F1	SQuAD1.1	EM	F1
Dev set results (single model)					
BERT [10]	78.98	81.77	BERT† [10]	84.1	90.9
RoBERTa [21]	86.5	89.4	RoBERTa [21]	88.9	94.6
XLNet	**87.9**	**90.6**	XLNet	**89.7**	**95.1**
Test set results on leaderboard (single model, as of Dec 14, 2019)					
BERT [10]	80.005	83.061	BERT [10]	85.083	91.835
RoBERTa [21]	86.820	89.795	BERT* [10]	87.433	93.294
XLNet	**87.926**	**90.689**	XLNet	**89.898**‡	**95.080**‡

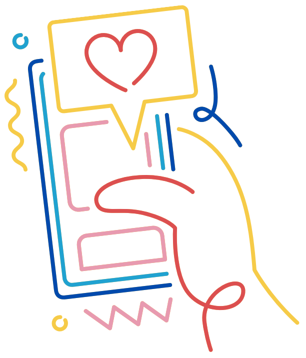
Table 3: Results on SQuAD, a reading comprehension dataset. † marks our runs with the official code. $*$ indicates ensembles. ‡: We are not able to obtain the test results of our latest model on SQuAD1.1 from the organizers after submitting our result for more than one month, and thus report the results of an older version for the SQuAD1.1 test set.

Summary

	XLNet	RoBERTa	DeBERTa
Computation cost	Highest — due to permutations, masking, and two-stream attention	Smallest	Higher — due to disentangled attention & enhanced decoder
Speed	Faster - Transformer-XL caching	Slowest	Fastest
Dependency among masked tokens	Yes	No	No
Long-term memory	Strongest	Weakest	Moderate

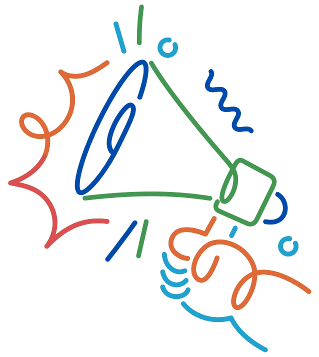
5 EXPERIMENTS

Training Settings



General

- Maximum sequence length: **128** token with Tail Truncation
- **Dynamic Padding**
- Learning rate: **2e-5**, early stopping
- Batch size: **64**, with shuffle if combined domains
- Warmup steps: **10%** of total training steps
- Optimizer: **AdamW**, with **weight decay (0.01)**



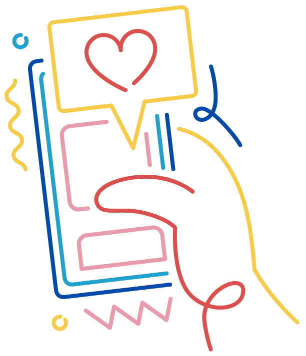
Model

- XLNet (“xlnet-base-cased”)
- RoBERTa (“roberta-base”)
- DeBERTa (“deberta-v3-base”)

Tokenizer

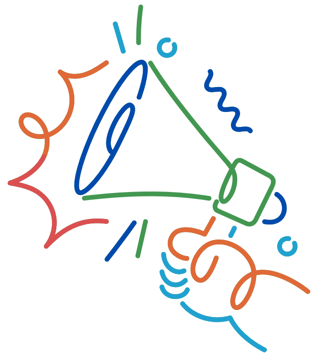
- XLNetTokenizer
- RobertaTokenizer
- DebertaV2Tokenizer

Evaluation Metrics



Label Mapping:

- **0** as **Real**
- **1** as **Fake**

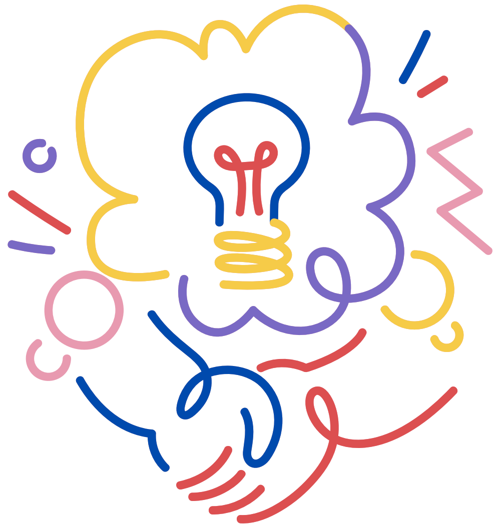


- **Accuracy**
- **Precision**
- **Recall**
- **F1-Score**

Approaches

Domain-Specific Fine Tuning (DSFT)

Each model was fine-tuned separately on individual domain-specific datasets to establish a performance baseline before investigating PDFT



Pooled-Domain Fine Tuning (PDFT)

We combined all domain-specific datasets into a single dataset . Each model was then trained on this combined dataset to encourage learning of more generalized representations across heterogeneous data.

Domain-Matched Ensemble (DME)

We combines the predictions of three models, each trained on the combined dataset. This allows us to leverage the complementary strengths of different model architectures.

DSFT on COVID-19

Metric	RoBERTa	XLNet	DeBERTa	SOTA*
Accuracy	97.85	96.1	97.35	98.70
Precision	98.80	96.0	99.32	98.91
Recall	96.67	95.8	95.10	98.33
F1-score	97.72	95.9	97.16	98.62

CT-BERT lacks cross-domain generalization. It is pre-trained solely on COVID-19 tweets, making it less effective on datasets like LIAR or PolitiFact due to domain differences. Additionally, its best performance requires adversarial training and heated softmax, which are computationally expensive and not reproducible in our setup.

*Source: Exploiting CT-BERT and Ensembling Learning for COVID-19 Fake News Detection. Dominican Republic, 2021

DSFT on LIAR

Metric	RoBERTa	XLNet	DeBERTa	SOTA*
Accuracy	60.93	68.35	65.14	70
Precision	75.84	72.54	70.47	-
Recall	57.95	82.03	79.02	-
F1-score	65.70	76.99	74.50	63.7

The BERT-Base + CNN model outperforms the baseline by integrating additional metadata such as emotion, specific features, and sentiment. This enriched input allows the CNN to capture nuanced contextual signals, leading to improved accuracy (70%) and F1-score (0.637).

*Source: Sentimental LIAR: Extended Corpus and Deep Learning Models for Fake Claim Classification. 2020

DSFT on PolitiFact

Metric	RoBERTa	XLNet	DeBERTa	SOTA*
Accuracy	89.94	84.9	85.14	86.16
Precision	92.98	91.8	77.14	85.91
Recall	81.54	69.2	90.00	86.16
F1-score	86.89	78.9	83.08	78.56

RoBERTa achieves state-of-the-art results on both GossipCop and PolitiFact, showing strong domain adaptability. Minor performance differences across runs, despite identical settings, are attributed to random initialization and GPU-related stochasticity.

*Source: Comparative Analysis of Graph Neural Networks and Transformers for Robust Fake News Detection. 2024

DSFT on GossipCop

Metric	RoBERTa	XLNet	DeBERTa	SOTA*
Accuracy	86.36	86.3	86.61	86.16
Precision	71.68	73.9	76.27	85.91
Recall	71.59	66.3	64.30	86.16
F1-score	71.63	69.9	69.77	78.56

RoBERTa achieves state-of-the-art results on both GossipCop and PolitiFact, showing strong domain adaptability. Minor performance differences across runs, despite identical settings, are attributed to random initialization and GPU-related stochasticity.

*Source: Comparative Analysis of Graph Neural Networks and Transformers for Robust Fake News Detection. 2024

PDFT Results (1)

Dataset	Metrics	XLNet	RoBERTa	DeBERTa
COVID-19	**Accuracy**	94.16	96.78	97.06
	Precision	94.88	98.27	97.9
	Recall	92.75	94.9	95.88
	F1-score	93.8	96.56	96.88
LIAR	**Accuracy**	65.75	64.77	64.4
	Precision	70.47	75.18	73.68
	Recall	80.81	65.53	69.8
	F1-score	75.28	70.02	71.69

PDFT Results (2)

Dataset	Metric	XLNet	RoBERTa	DeBERTa
PolitiFact	Accuracy	79.25	84.91	81.76
	Precision	74.24	83.61	83.33
	Recall	75.38	78.46	69.23
	F1-score	74.81	80.95	75.63
GossipCop	Accuracy	85.94	86.78	86.51
	Precision	76.43	76.24	75.47
	Recall	60.08	65.46	65.08
	F1-score	67.27	70.44	69.89

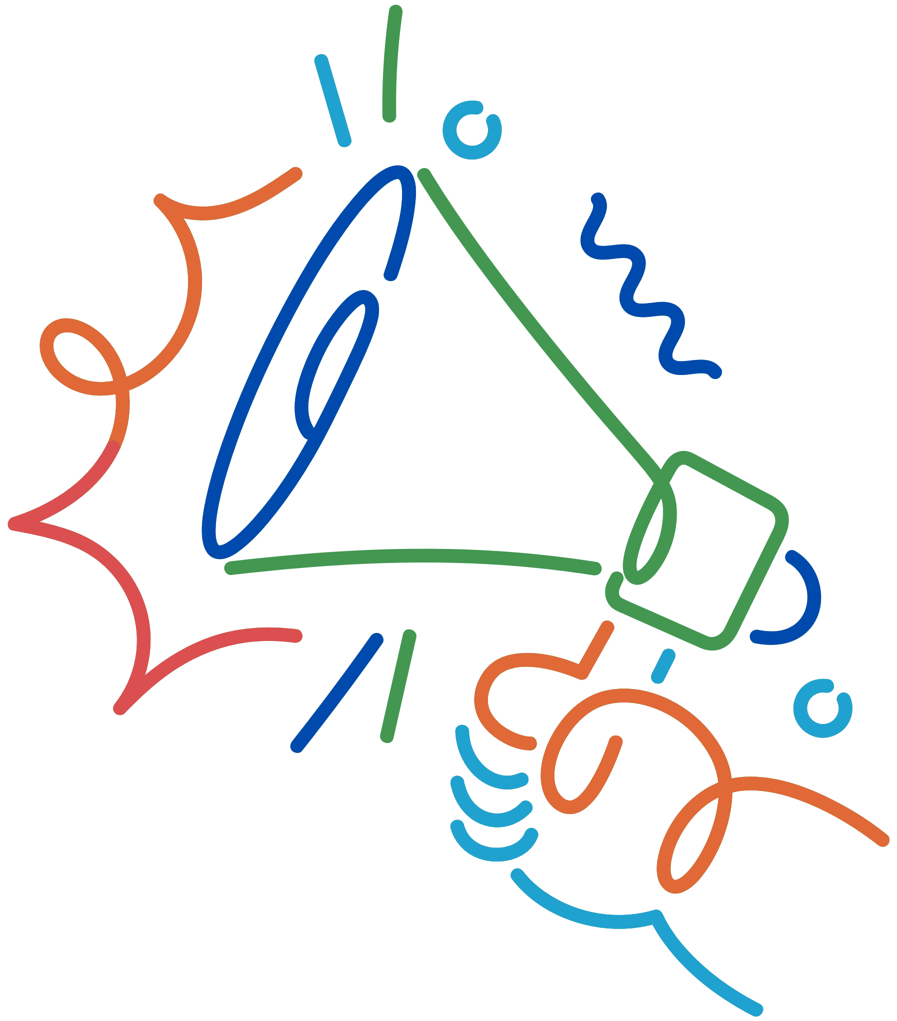
DME Results (1)

	Metric	DME Model
COVID-19	Accuracy	97.48
	Precision	98.59
	Recall	96.08
	F1-score	97.32
LIAR	Accuracy	67.40
	Precision	75.49
	Recall	75.31
	F1-score	74.89

DME Results (2)

Dataset	Metric	DME Model
PolitiFact	Accuracy	86.16
	Precision	89.09
	Recall	75.38
	F1-score	81.67
GossipCop	Accuracy	87.53
	Precision	80.51
	Recall	63.58
	F1-score	71.05

6 CONCLUSIONS



- **Domain-specific models** achieve the **best performance** within their **respective domains.**

- **Pool-domain models** offer **generalization** but are prone to **conflicting signals** during evaluation.

- **Domain-matched ensemble models** improve **accuracy**, especially on **complex** data.

- Model choice depends on the **deployment goal:**
    - **Domain-specific** → for **high precision**
    - **PDFT/DME** → for **cross-domain scalability**

- **Future work:** Unify expertise via adaptive routing and knowledge distillation.

Thank you for your attention!

Any question?