LingML: Linguistic-Informed Machine Learning for Enhanced Fake News Detection

Jasraj Singh, Nanyang Technological University (NTU), Singapore
Fang Liu, Singapore University of Social Sciences (SUSS), Singapore
Hong Xu, Nanyang Technological University (NTU), Singapore
Bee Chin Ng, Nanyang Technological University (NTU), Singapore
Wei Zhang, Singapore Institute of Technology (SIT), Singapore

The Fake News Problem

- **Digital Connectivity**: Social media has transformed how information is created and shared, but also accelerates the spread of misinformation and fake news.
- **Unverified Information**: During major events such as the COVID-19 pandemic, social media can become a breeding ground for unverified claims and conspiracy theories.
- **High Stakes & Uncertainty**: People can become more susceptible to misleading narratives when events involve high stakes and significant uncertainty.
- **Consequences**: Uncontrolled fake news can shape public opinion, erode trust, and undermine the value of timely and reliable information.

Existing Methods

Machine learning (ML): a great potential for fighting against the fake news proliferation on social media platforms.

- A **classification problem** to distinguish fake and non-fake news and assign a true label to fake news and false otherwise.
- **Natural language processing (NLP)**, a main category of ML, has been applied to detect fake news because of its strength in analyzing text-based data like news.
- NLP models have become **larger and larger** to achieve improved performance and several **large language models (LLM)** have been proposed in recent years.
- Some LLM have also been specialized for certain topics, e.g., CT-BERT for COVID-19

ML Limitations

- **Accuracy**: Detection error rates have decreased with increasingly advanced models, but further improvements are becoming harder and increasingly marginal.
- **Interpretability**: During Increasing model complexity leads to poor interpretability, making the detection process difficult to understand and turning models into “black boxes.”
- **Generalizability**: Models may overfit to their training data, while rapidly evolving social-media language, such as new buzzwords and acronyms, can reduce their effectiveness on unseen data.

*Can we leverage upon **domain knowledge?***

Linguistic-informed ML (LingML) for fake news detection

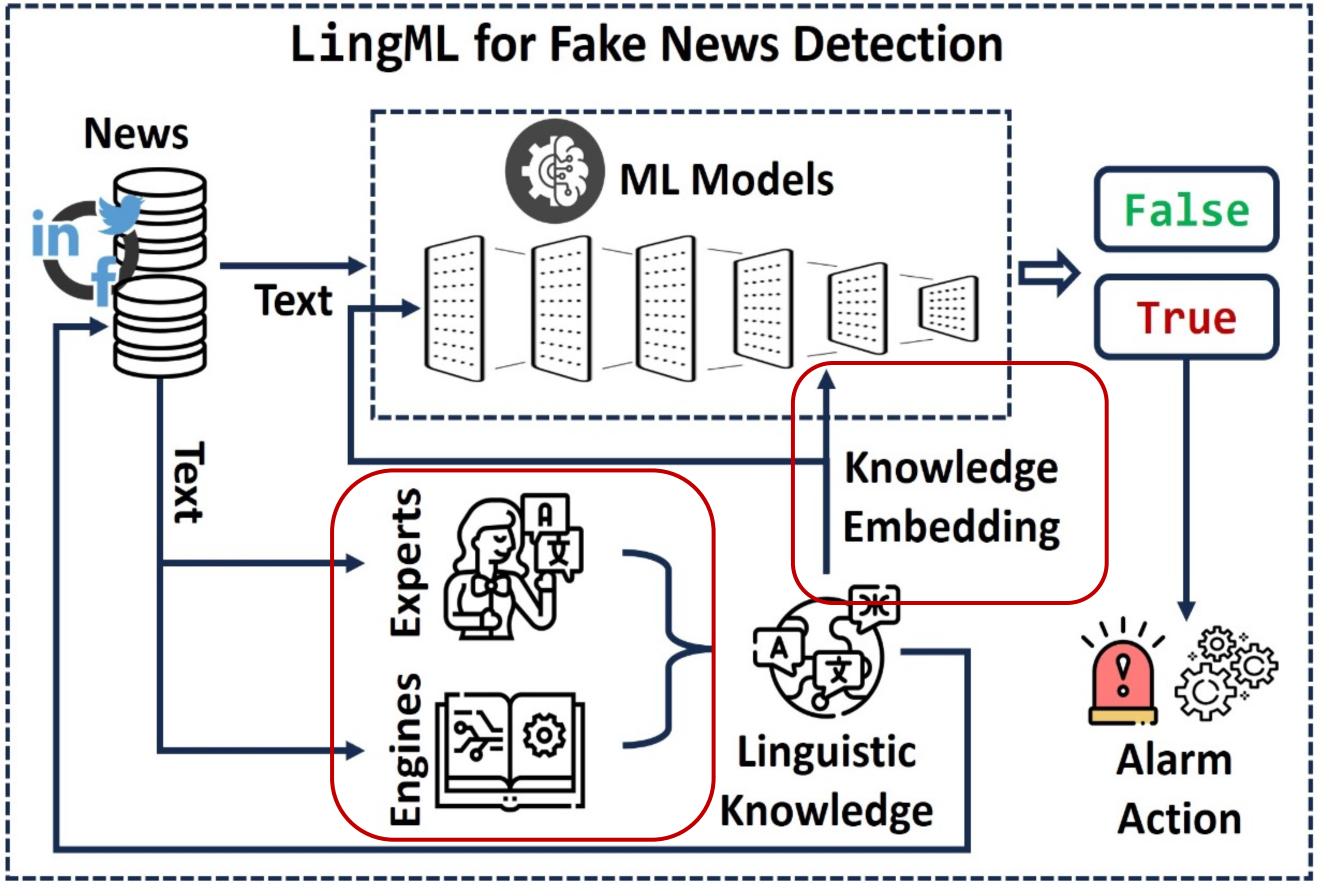

Fig. 1. The system architecture of the LingML for fake news detection. As an ML-based system, the news is the input and the output is true if it is fake news and false otherwise. The key technical challenge in the system is knowledge representation and utilization. LingML extracts linguistic knowledge from the input news and produces ML digestible linguistic features. The knowledge can be integrated into LingML in different ways and stages for achieving the optimal detection performance and the system reacts with risk mitigation measures once a fake news is detected.

Importance of Linguistic Knowledge

Feature	Description & Examples
feeling	feel, hard, cool, felt, etc.
assent	yeah, yes, okay, ok, etc.
perception	in, out, up, there, etc.
discrepancy (discrep)	would, can, want, could, etc.
certitude	really, actually, of course, real, etc.
causation (cause)	how, because, make, why, etc.
words per sentence (wps)	average words per sentence
space	in, out, up, there, etc.
auditory	sound, heard, hear, music, etc.
all-or-none (allnone)	no. of words, e.g., all, never, always
motion	go, come, went, came, etc.
negative tone (tone_neg)	bad, wrong, too much, hate, etc.
swear words (swear)	shit, fuck, fucking, damn, etc.
positive tone (tone_pos)	good, well, new, love, etc.
exclamation points (exclam)	no. of exclamation points
question mark (qmark)	no. of question marks
netspeak	unofficial words, e.g., :), u, lol, haha
conversation	yeah, oh, yes, okay, etc.

Emotion & Tone (5 features)

Cognition & Perception (6 features)

Social & Style (7 features)

Why Linguistic Features Matter?

These features capture psychological patterns of deception, making them highly effective for fake news detection while remaining interpretable.

*Understanding the **18 linguistic features** that distinguish fake news from real news.*

LingM Framework: Linguistic-Only ML

Training ML models with 18 linguistic features only - no raw text processing

- **Input:** 18 normalized linguistic features (z-score standardized) extracted from news text
- **Models:** 5 basic ML algorithms - SVM, Decision Tree, Random Forest, XGBOOST, Neural Network
- **Advantage:** Highly interpretable and explainable - no black-box processing
- **Performance:** 18-22% error rate with basic models, demonstrating strong linguistic signal

Even with only linguistic features, basic ML models achieve reasonable accuracy, highlighting the effectiveness of linguistic knowledge for fake news detection.

LingL Framework: Linguistics + LLM Fusion

Two-stage fusion architecture combining linguistic knowledge with transformer-based LLMs.

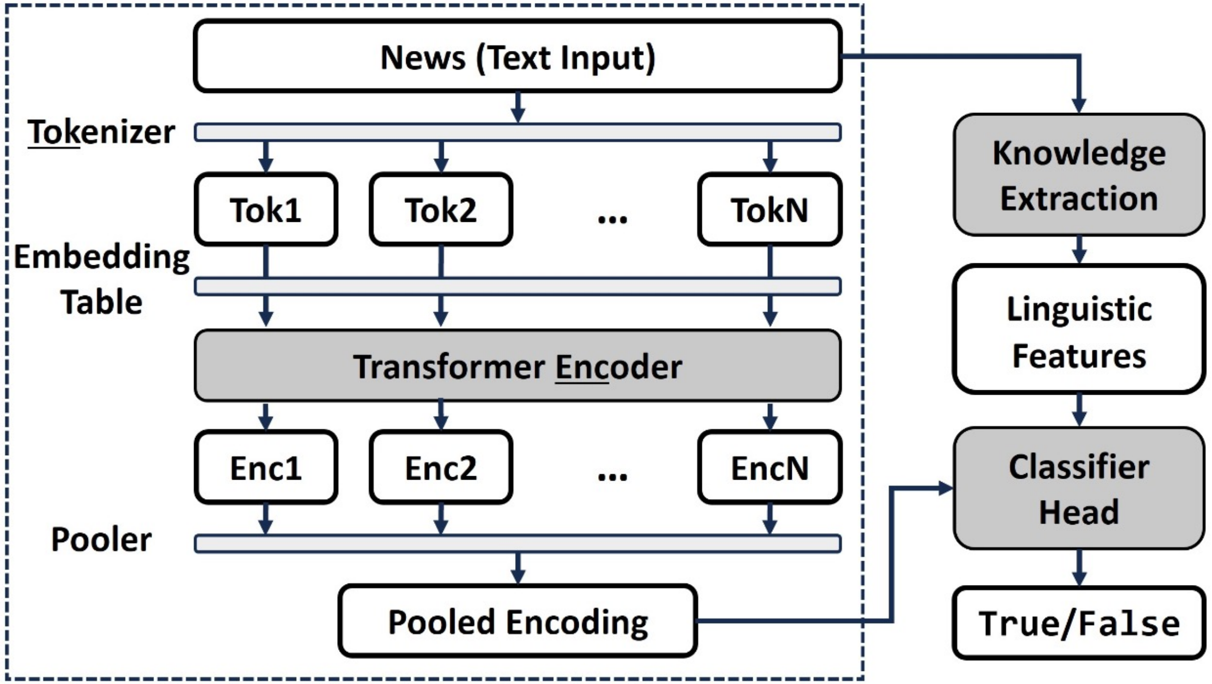

Fig. 2. An illustration of LingL's system architecture with news as input. Linguistic knowledge is extracted from news and represented as numerical features, which are concatenated with the pooled encoding of the news produced by a LLM. The concatenated vector is directed to a classifier to produce the detection outcome of the news.

Performance Analysis

- Dataset Configuration:
  - **Source**: COVID-19 fake news dataset with 6,000 news samples from social media (Facebook, Twitter, Instagram) and official channels
  - **Split**: 60/20/20 for training, validation, and testing respectively
  - **Verification**: Manually checked and updated by linguistic experts for label reliability
  - **Preprocessing**: Uncasing, tokenization ([URL], [MENTION], [HASHTAG], [COVID]), emoji conversion

Original news: A British company has launched a household cleaner which destroys #COVID19 in just 60 seconds. Follow live #coronavirus updates 👇 https://t.co/NyZLPJyfxM

Processed news: a british company has launched a household cleaner which destroys [COVID] in just 60 seconds. follow live [COVID] updates [EMOJI] backhand index pointing down [EMOJI] [URL]

LingM Performance & Feature Importance

- **Effective:** Basic ML models achieve 20% error using linguistic features only.
- **Stable**: Linguistic features provide consistent performance across runs.
- Demonstrate that linguistic knowledge alone provides strong discriminative information.

Table 2. LingM's performance (error rate) with five small-scale ML models trained with hand-designed linguistic features.

LingM	Error Rate		
	Mean	Best	SD
SVM	22.53%	22.53%	0.000
DT	21.51%	21.50%	0.019
RF	19.11%	18.79%	0.231
XGBOOST	**18.32%**	**18.32%**	0.000
NN	20.24%	19.67%	0.332

LingM Performance & Feature Importance

- **Feature Importance:** Contributions vary from 1.8% to 15.9%.
- **“Feeling” Dominates:** 15.9% importance, 54% higher than the second-ranked feature.
- **Top-5 Features:** Contribute >50% of total importance.

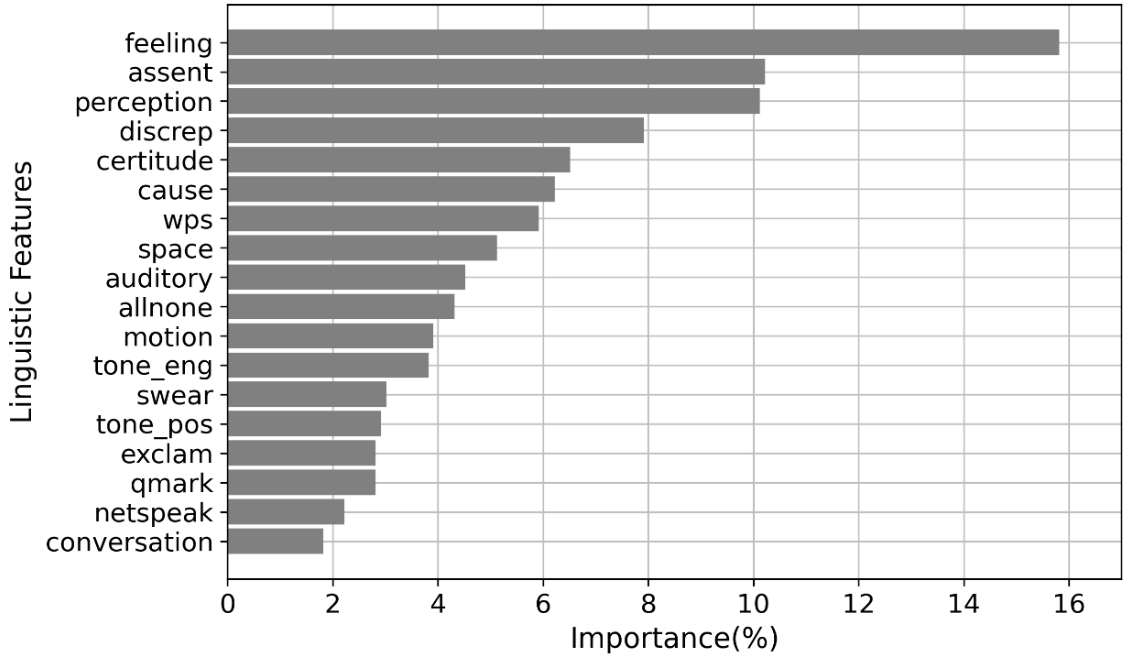

Fig. 4. The ranking of the features based on their importance (as percentage) to `LingM`. The summation of all importance/percentages is 100%. Feature abbreviation is used for space saving and the detail can be found in Table 1.

LingM Performance & Feature Importance

Algorithm 1: DT with with linguistic feeling feature

Input: a news t

$x_1 \leftarrow$ the value of the linguistic `feeling` feature of t;

if $x_1 <= -0.174$ **then**

  return true; ▷ this is a fake news

else

  return false; ▷ this is not a fake news

end

- **Even ONE Feature Works:** Using only “feeling” + Decision Tree achieves 28.4% error.
- **Highly Interpretable:** A simple if–else rule can already distinguish fake from real news.
- Linguistic knowledge provides a strong signal for fake news detection. So, how about **linguistic knowledge** integration with **advanced LLM**?

LingL Performance Analysis

- **Strong LLM Baselines:** All tested LLMs achieve <4% error, with CT-BERT best at 2.11%.
- **Consistent Improvement:** LingL reduces error for all 11 LLM backbones.
- **Best Performance:** CT-BERT + LingL achieves 1.81% mean error and 1.68% best error.
- **Average Gain:** Linguistic integration provides 11.8% average improvement across all LLMs.
- **Largest Gain:** XLM + LingL achieves the highest relative improvement of 17.9%.

Table 3. LLM and LingL's classification error rates with different base models. The last column is calculated as the improvement in mean error rate. All values except SD are reported as percentages (%).

Model Name	LLM	LingL			Improv.
	Mean	Mean	Best	SD	
ALBERT [7]	3.37	2.99	2.80	0.11	11.4
BERT [6]	3.21	2.75	2.57	0.10	14.5
BERTweet [11]	2.87	2.61	2.48	0.13	9.1
CT-BERT [10]	2.11	**1.81**	**1.68**	0.08	14.0
DistilBERT [15]	3.08	2.78	2.66	0.06	10.0
Longformer [2]	3.40	2.89	2.76	0.08	15.1
RoBERTa [9]	3.20	2.80	2.38	0.22	12.3
Twitter-RoBERTa [1]	2.94	2.72	2.52	0.11	7.6
XLM [5]	3.30	2.71	2.48	0.14	**17.9**
XLM-RoBERTa [4]	2.87	2.57	2.38	0.12	10.4
XLNet [19]	3.36	3.09	2.94	0.12	7.8

Insights & Implications

Bigger Models Are Not the Only Path

- **Resource Intensive**: BERTweet uses 80 GB / 850M tweets, while Twitter-RoBERTa uses an additional 58M tweets for domain adaptation.
- **Alternative Path**: LingML improves performance by adding explicit linguistic knowledge, without significant model or data overhead.
- **Beyond Accuracy**: Linguistic knowledge also supports better interpretability and generalizability.
- **Key Insight**: <u>Domain knowledge</u> can complement model scaling for more effective fake news detection.

Thank you! Q&A