

LIE DETECTION - OPINION DATASET

Cognitive, Behavioural and Social Data

Graziana Capurso

Alessia d'Addario



Agata Garbin

Alice Ronzoni

LIE DETECTION

- Lie detection involves the process of determining whether a given communication is true or false.
- Research has shown that it is possible to study ad hoc strategies to detect lies.
- In the last years, studies have been conducted using Machine Learning and Deep Learning algorithms combined with Natural Language Processing (NLP)

PREVIOUS WORK

The starting point of this paper is the work of Loconte et al.[1], a study of a fine-tuned open-source LLM named FLAN-T5 developed by Google, used in the lie detection problem

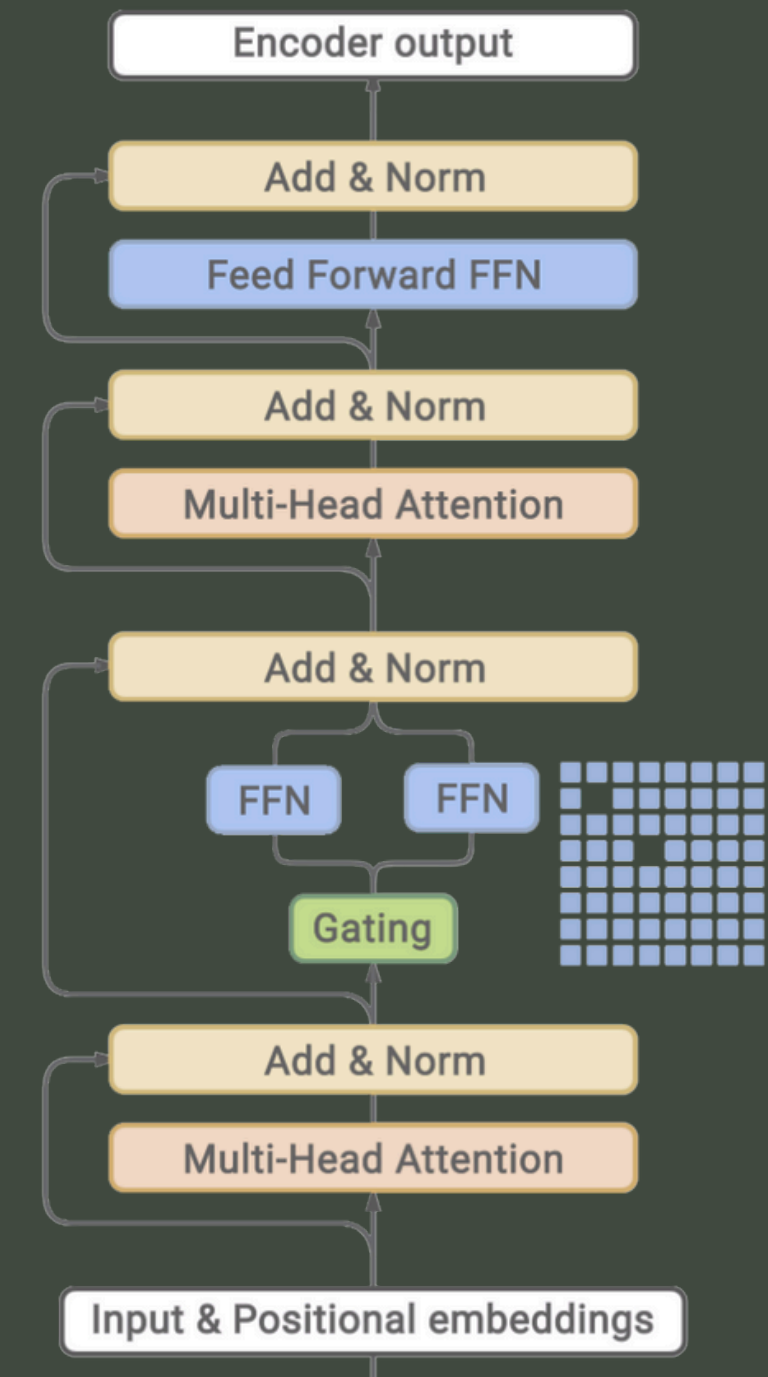
[1] Loconte et al. Verbal lie detection using large language models. 2023.

Ethnic similarities and differences in linguistic indicators of veracity and lying in a moderately high stakes scenario. Journal of Police and Criminal Psychology

FLAN-T5

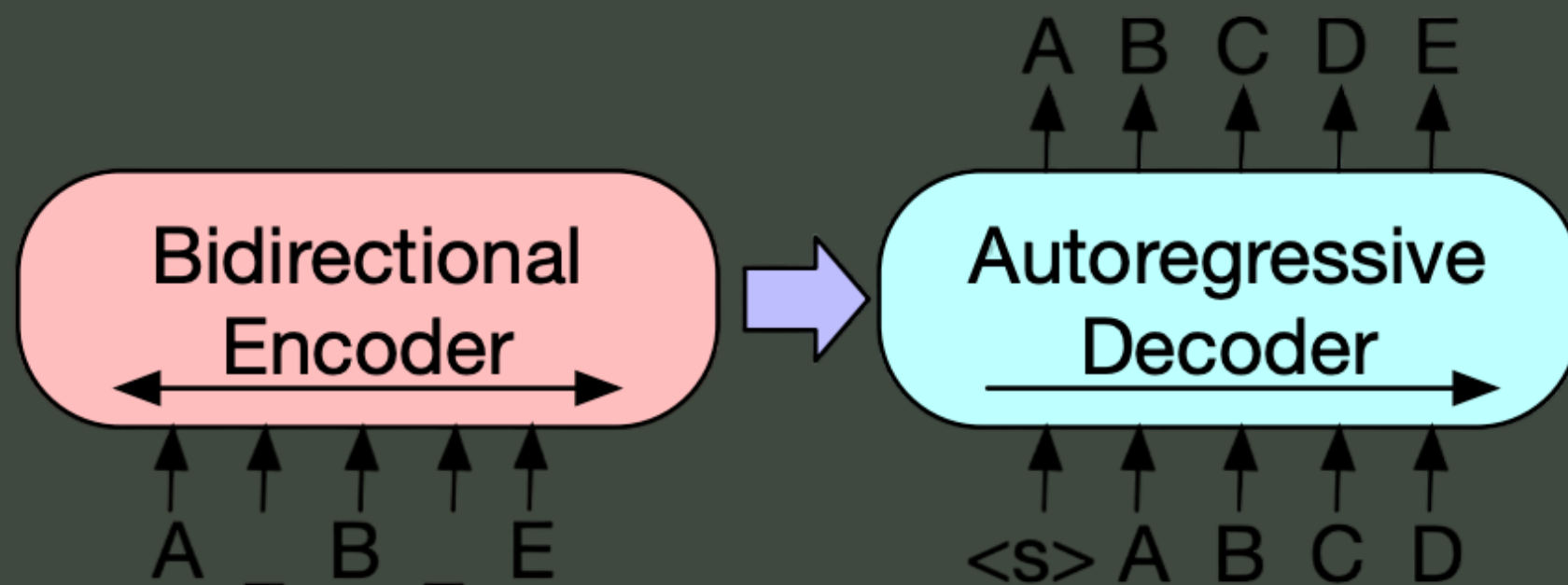
Google researchers developed an improved version of the MT-5 model, a text-to-text general model that excels in many NLP tasks.

This Large Language Model (LLM) balances the trade-off between computational load and the quality of learned representations, enhancing performance without excessive computational demands.



BART

The BART model employs a seq-to-seq architecture, featuring a bidirectional encoder (like BERT) for context-rich input understanding, and a left-to-right decoder (similar to GPT) for generating coherent outputs.



This combination, effective for text generation, also excels in comprehension tasks, leveraging both bidirectional and unidirectional processing.

DATASET: OPINIONS

The dataset contains the opinions from 500 people on five distinct topics: abortion, cannabis legalization, euthanasia, gay marriage and policy on migrants, resulting in a total of 2500 sentences.

In particular, it consists in a total of 1250 truthful and 1250 deceptive opinions balanced across topics.

TRUTHFUL	DECEPTIVE
DOMAIN: ABORTION	
IT	
Penso che ogni donna dovrebbe avere diritto di scegliere se portare avanti o meno una gravidanza. In ogni caso non mi schiero completamente a favore dell'aborto, perchè ci sono circostanze in cui viene comunque interrotta una gravidanza nonostante potrebbero esserci soluzioni alternative, che non comportino né un cambiamento di vita drastico per la donna, né la perdita di una vita.	L'aborto è una cosa inumana e non capisco come possa essere legale. Fortunatamente esistono gli obiettori di coscienza che decidono al posto di quelle sciagurate che hanno pensato bene di rimanere incinta e poi se ne pentono e vogliono uccidere il bambino. Che poi, cosa costa portare a termine la gravidanza e dare in adozione il bambino?
EN	
While I am morally torn on the issue, I believe that ultimately it is a woman's body and she should be able to do with it as she pleases. I believe people should not dehumanize the fetus though, to make themselves feel better. The decision about laws regarding this issue should be left up to the states to decide. To combat this problem, birth control should be easily accessible.	Abortion is the termination of a life and should not be allowed. If a fetus has made it to the point of being able to survive "on its own" outside its mother's body, what right do we have to cut its life short. If the mother's life is in danger, she already chose that she was willing to sacrifice her life to have a child when she consented to procreating.

Table 3: The table shows some examples of the gathered opinions for both the Italian (IT) and standard American English (EN) languages included in the DecOp corpus. The domain considered for the example is Abortion.

METHOD: SCENARIO 1

The model was divided using 10-fold cross validation, fine-tuned on a portion of the dataset and tested on the remaining part.

Fine-tuning of LLMs consists in adapting a pre-trained language model to a specific task by further training the model on task-specific data.

TRAINING SET

450 writers

TEST SET

50 writers

RESULTS

FLAN-T5	BART
82.6 ± 0.03	82.2 ± 0.011

RESULTS

EXP-1	EXP-2	EXP-3
0.7636 ± 0.0182	0.82 ± 0.0140	0.7812 ± 0.0159

RESULTS

INTENSIONS	MEMORIES
0.72561 ± 0.0327	0.794095 ± 0.0165

BART ON ITALIAN DATASET

Eager to see what BART was capable of achieving with the Italian language, we decided to fine tune the model on this new dataset, using the same setting described above.

In order to do so, we made use of a tailored version of Bart pre-trained on Italian text corpora

BART
0.857 ± 0.0128

CONCLUSION AND DISCUSSION

- The results are satisfactory, suggesting promising prospects for the use of LLMs such as BART in the field of Lie Detection
- The additional experiments involving the Italian language give hope for the potential application of these techniques in languages beyond English.