Introduction

Fake News

- **Fake News:** News written to intentionally deceive or sway public opinion
- 64% of US adults said that “made-up news” has caused a “great deal of confusion” about the facts of current events
- False news stories are 70% more likely to be retweeted than stories from reliable sources
- Fake news is currently mostly written by humans, but computer-generated fake news is a foreseeable problem

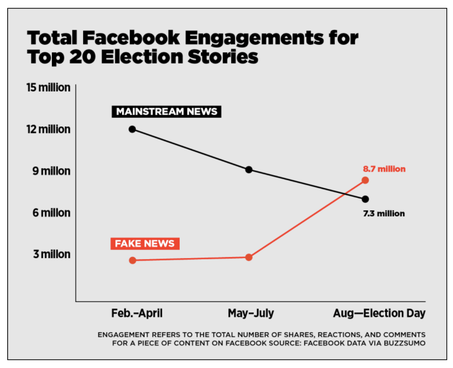


Figure 1: Sharing of news during 2016 election season

Human fact-checking websites exist…
→ cannot keep up with the rapid output of fake news
→ computer models rapidly generate these articles

Natural Language Processing

Natural Language Processing (NLP): Study of interaction between computer and human language

https://sadasystems.com/blog/google-translation-api

https://thetechpanda.com/a-brief-history-of-voice-recognition/27841/

https://usingtechnologybetter.com/check-spelling-in-a-gmail-message/

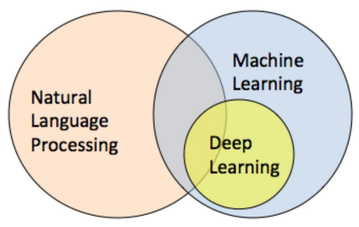
Natural Language Processing Tasks:
→ Sentiment Analysis: Detects emotion in text data
→ Stance Detection: Analyzes the position of the author on a topic
→ Answering of simple questions within a larger text

Machine Learning

- NLP research uses machine learning models to test data
- Uses algorithms to find correlations in large datasets
- Models can identify features, or characteristics, that separate one type of datapoint from another

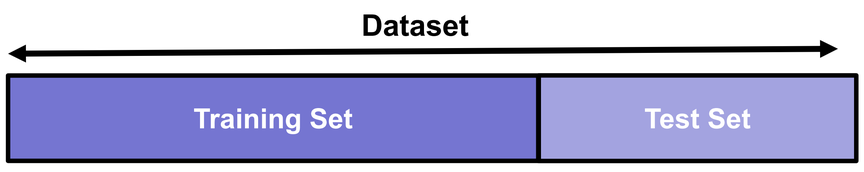
Deep Learning: Layered machine learning models creating a larger neural network

- A greater number of layers makes a more precise and accurate model



https://www.eventshigh.com/detail/bangalore/b8bf5130003a638ad1c32bfe8e3d35e6-bootcamp-on-natural-language-processing

- A neural model must be trained with a set of data, then tested with another
- Neural networks loosely resemble the human brain in the methods with which they learn and process information
- A train/test split separates a set of data to mostly training data, with some test data to evaluate the model’s accuracy



Grover Learning Model

- Grover is a deep learning NLP model capable of identifying and creating neural fake news
- Grover was developed to prepare for malicious of neural fake news generation
- Learn to combat neural fake news before it is an issue
- The Grover model is open access for both discrimination and generation

Review of Literature

Language Models are Unsupervised Multitask Learners

Goal: To use language modeling to train an unsupervised deep learning model

A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever. “Language Models are Unsupervised Multitask Learners”. In: (2019).

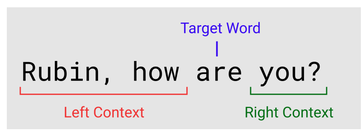
Results:
- A model capable of realistic human writing
- Creates coherent writing based off a minimal prompt
- 48 layer model
- Predicts the next word repeatedly (unidirectional language)

Bert: Pre-training of deep bidirectional transformers for language under-standing

Goal: To create a language model capable of performing common NLP tasks

Results:
- A neural model capable of Question Answering and Sentiment Analysis (analyzing the emotions of an author), among other NLP tasks
- More practically useful than GPT-2, cannot generate equally convincing writing
- Bidirectional, analyzes full context of sentences, does not analyze word-by-word

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language under-standing arXiv preprint arXiv:1810.04805, 2018.



https://medium.com/@plusepsilon/the-bidirectional-language-model-1f3961d1fb27

Figure 2: Bidirectional learning

How to Build a "Fake News" Classification Model

Goal: To create a natural language dataset of reliable and fake news

McIntire, G. (2018, April 18). How to Build a "Fake News" Classification Model. Retrieved from https://opendatascience.com/how-to-build-a-fake-news-classification-model/.

Results:
- ~11,000 article dataset of equal parts fake and real news
- Used data scraping using All Sides’ reliable news from various political sources
- Annotated labels of real and fake
- Fake news classification
- Heavily sources from news during the 2016 election cycle

Gap in Research

Neurally-generated news has never been used to train for identification of fake news.

Testing Computer Generated Disinformation with Human Written Fake News Using the Neural Model Grover

Discussion

Application of Research

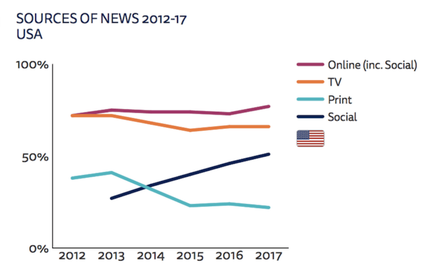
- Grover demonstrates that in its current state, neural fake news is not a suitable replacement for human-written fake news, and they cannot be identified the same way
- A better understanding of the future of fake news will be key to stopping it, or at least preventing it from exercising the same impact in the 2020 presidential election as it did in the previous
- The development of neural fake news should be explored, but it is currently ineffective and imperfect
- Neural fake news will be easier to identify to a computer than human fake news

https://www.fastcompany.com/90309287/this-ai-generates-fake-news-about-anything-you-want-try-it

Figure 3: Grover is capable of detecting, as well as generating fake news articles

Fake News

- Once a fabricated article reaches virality, its impact is over
- The top 1% of successful fake news reaches 1,000 to 100,000 people, the truth rarely diffuses to 1,000
- The only method of preventing fake news is to identify before it becomes viral
- Fake news poses a larger issue than small-scale deception of individual readers: it is the greatest source of negative publicity towards businesses
- Fake news impacts politics of America
  - → alters public opinion
  - → tarnishes reputation of politicians with false information



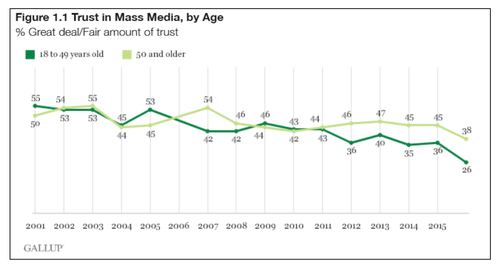
https://www.businessofapps.com/news/social-media-apps-are-the-go-to-news-source-for-51-of-people/

Figure 4: People consume news from social media at an increasing rate

- Social media bots maintain constant popularity of fake news

Effects on Society

- Trust in media is at an all time low at 32%
- Need for improvement to mass media is high
- Media trust continues to decline year by year



https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook#.qqEPoaOJ3

Figure 5: American people’s trust in mass media

Limitations

- Grover’s mega size model (48 layers) could not be used due to financial limitations
- Instead, the large model was used (28 layers) in Google Colab
- Deep Learning, as a directly superior form of machine learning, is more taxing on computer hardware and disadvantages researches with less funds
- Recent and up to date datasets are hard to find in a constantly shifting field

Future Research

Year 1:
- Expand study to include a second human dataset
- Use mega-sized Grover model
- Test model with numerous datasets of both human and neural news

Year 2:
- Expand study to include GPT-2 and BERT models
- Create a deep learning model mostly using GPT-2, BERT, and Grover’s designs

Conclusion

Goal
To determine if Grover’s discrimination model could be used for detection of neural fake news

Hypothesis
Since the model is trained on the test data, it is assumed that the human data testing round will perform better.

Results
The human data testing round performs significantly better than the neural testing round

References

[1] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever. “Language Models are Unsupervised Multitask Learners”. In: (2019).
[2] Adams, C. (2019, July 17). Expensive, Labour-Intensive, Time-Consuming: How Researchers Overcome Barriers in Machine Learning. Retrieved from https://journal.binarydistrict.com/expensive-labour-intensive-time-consuming-how-researcher-overcome-barriers-in-machine-learning/.
[3] Baly, R., Karadzhov, G., Alexandrov, D., Glass, J., & Nakov, P. (2018). Predicting Factuality of Reporting and Bias of News Media Sources. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing.
[4] Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 U.S. Presidential election online discussion. *First Monday*, *21*(11). doi: 10.5210/fm.v21i11.7090
[5] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language under-standing.arXiv preprint arXiv:1810.04805, 2018.
[6] Jiawei Zhang, Limeng Cui, Yanjie Fu, and Fisher B Gouza. 2018. Fake News Detection with Deep Diffusive NetworkModel.arXiv preprint arXiv:1805.08751 (2018)
[7] Kosoff, M. (2018, January 29). The Fake-News Epidemic Is Worse Than We Imagined. Retrieved from https://www.vanityfair.com/news/2018/01/the-fake-news-epidemic-is-worse-than-we-imagined.
[8] Lazer DMJ, Baum MA, Benkler Y et al.: The science of fake news.Science 359:1094–6, 2018x
[9] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521,** 436–444 (2015) doi:10.1038/nature14539
[10] McIntire, G. (2018, April 18). How to Build a "Fake News" Classification Model. Retrieved from https://opendatascience.com/how-to-build-a-fake-news-classification-model/.
[11] Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. 2019. Defending Against Neural Fake News arXiv:1905.12616.
[12] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online. Science359, 1146–1151 (2018).
[13] Welch, R. (2017, April 20). PolitiFacts guide to fake news websites and what they ... Retrieved from https://www.politifact.com/punditfact/article/2017/apr/20/politifacts-guide-fake-news-websites-and-what-they/.

Problem Statement / Goal / Hypothesis

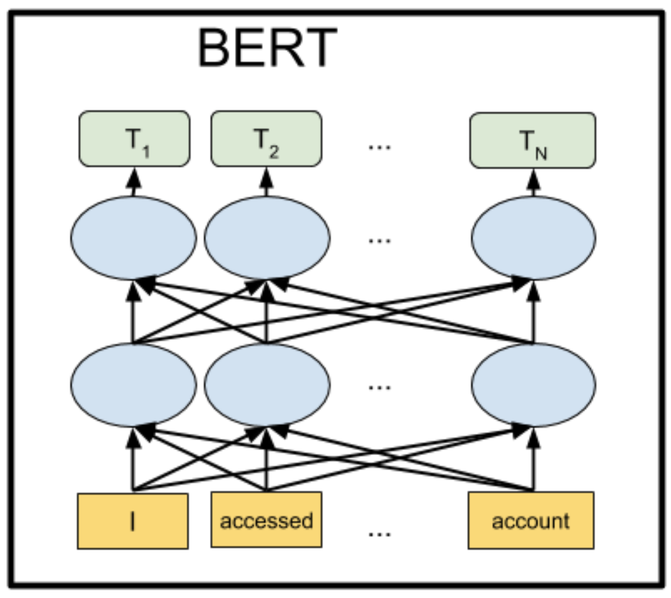
Problem:
Grover has never been tested as a discriminatory language model for real and fake human news

Goal:
Train and test Grover's discrimination model with a dataset of real and fake human news to compare neural fake news with human fake news

Hypothesis:
Grover will perform significantly better when trained with human news than with neural news

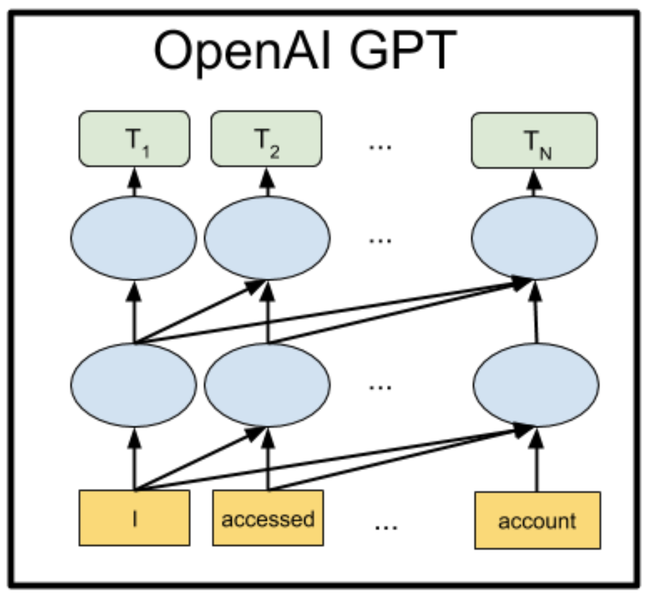
Methodology

Grover's Learning Model



https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html

Bidirectional Learning



Unidirectional (Left to Right) Learning

Grover

- Mostly modeled after GPT-2
- Grover is a unidirectional learning model (similar to GPT-2)
- Uses the same architecture as the original GPT model (which processes are in which layers)
- Grover's three sizes are on par with the BERT and GPT-2, with Grover Mega (it's largest size) at 48 layers and 1.5 billion parameters

Training and Testing

Datasets and Preprocessing

McIntire Fake News Dataset n = 10,558 Samples

- **Fake** → **Kaggle dataset:** fake news scraped from 244 websites
- **Real** → **All Sides:** Real articles from various political sources

Original

3608, Kerry to go to Paris in gesture of sympathy, U.S. Secretary of State John F. Kerry said Monday that he will stop in Paris later this week, amid criticism that no top American officials attended Sunday's unity march against terrorism., REAL

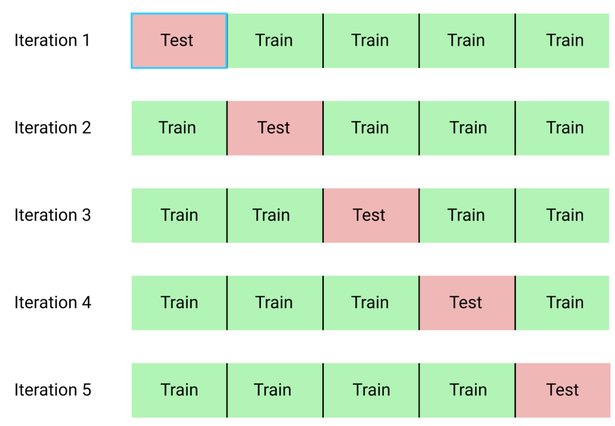
Transformed

{"title": "Kerry to go to Paris in gesture of sympathy": " U.S. Secretary of State John F. Kerry said Monday …authors": null, "date": null, "domain": null, "ind30k": "1571", "url": null, "label": ”real", "orig_split": null, "split": "test", "random_score": null},

Training and Testing

- **Train Grover with original input data**
- **Train Grover with McIntire's Fake News Dataset**

→ **Test Grover with McIntire's Fake News Dataset**



Cross Validation is a procedure used if a quantity of data is insufficient for a test/train split

Results

Figure 3: Results of Grover model tested with the MicIntire Fake News Dataset, trained with different datasets

	NFN	HFN
# of Training Instances	10,000	6,336
# of Training Instances	6,336	6,336
Accuracy	44.2%	95.4%

Table 1: Experimental Results with dataset instances