Natural Language Processing Article Reviews

Phoe Aung, Dylan Bortolon, Laurence Burden, and Amanda Burfield

Purdue Global University

IN400: Artificial Intelligence (AI) - Deep Learning and Machine Learning

Robert Kayl

26 September 2023

# 

**Phrase-Based & Neural Unsupervised Machine Translation**

**By: Phoe Aung**

**Article Summary**

Machine translation (MT) is the act of translating one language to another via the usage of models and machine learning. The authors mentioned that older models of MT were slow and at times inaccurate. MT also faces the challenges of having to rely on large parallel corpora to be effective. What this essentially means is that the input dataset needs to contain many parallel sentences between two languages for the model to be accurate. So, languages that are distant or lack low resources tend to struggle when trying to be translated from one language to another. For this reason, the article went over two different methods that allow for a more efficient MT; these methods are called neural machine translation (NMT) and phrase-based statistical MT(PBSMT). As the name implies the NMT uses a neural network whereas PBSMT uses a phase-based model. Overall, the results of these models were extraordinary in that they outperform some of the state-of-the-art machine translations, and they did it by not using a single parallel sentence in their model.

**Usage and contribution relating to NLP**

Natural language processing (NLP) is the idea of trying to create a machine that can interpret and comprehend human language. Machine translation then utilizes this feature to help translate one language to another. Essentially MT cannot be developed without the use of NLP, and as mentioned above the two NLP models the authors utilized were neural network and phase-based model. The article was able to contribute to NLP by creating new methods that help compute and interpret different languages. An example is when the authors use byte-pair encoding (BPE) instead of bilingual dictionaries to reduce the vocabulary size and eliminate unknown words in their NMT model. This allows the model to be much more effective and simpler without the need to translate each word one by one.

**Ethical Issues**

Ethically the usage of NLP in the field of MT tends to lean more on the acceptable side since the main goal is to provide better translation between two languages through the usage of machine learning. The only ethical concern that might arise is that the translations are inaccurate or unprofessional, which can cause more harm than good for the user.

**Advantages and disadvantages**

NLP does a lot for machine translation and is even considered to be the backbone of MT. Without NLP a lot of the translation cannot be done since the machine is not able to compute the words there will be no computed word to translate. NLP however does have its disadvantages, while it is used as a language processor, it is not able to detect tone, attitude, and ambiguity as well as a human can. This then causes a problem when trying to translate in a professional setting, as the word choices for the translation might be proper for the user. A lot of things can go wrong when the word choices are inappropriate or just plain inaccurate, and as a result, can cause misinterpretation.

**Article one: Amazon Web Services MT**

Article one stated that Amazon Systems has integrated a few machine translations into their business. However, once all the machine translation is done, Amazon sends the data to a human translator for editing. What this implies is that MT is not as perfect as it could be, and it still requires a professional to look over the translated work. This is important to consider since the article above wanted to compare their model to the previous MT model, but never really started how well machine translation works. In truth, there isn’t a perfect machine translation yet, and there are flaws that exist that these models from the article cannot fix. For this reason, many organizations such as Amazon would still require a human translator to correct bad translations made by the MT.

**Article two: Shortcomings of MT**

While many people use machine translators like Google Translate with ease, it is to note that it isn’t perfect. As mentioned above it is quite common for MT to inaccurately translate a word, due to a lack of resources or because the languages are too distant. This can cause a lot of misunderstanding and the worst part is you can’t tell if your translation was incorrect. The article then proceeded to state that if the translation was used incorrectly in a professional setting, it can start to become expensive as you would most likely need to hire a professional translator to fix the mistake.

**Article Three: MT Usage and Caution**

Article three dives more into concerns that might reside in public machine translators. The article stated that since most machine translators are free and open source like Google Translate, the catch is that they might not be confidential. Ethically this can be of concern. On one hand, most of these translators have a term and agreement that announces that they are collecting the user data. On the other hand, collecting user data has always been an ethical concern of modern discussion, and to an extent, it has become normalized and to be expected.

**Future Research and Concern**

The models from this article are indeed breakthroughs for MT technology. These models are much more effective and simpler than older MT models but still have a long way to go before they are perfect. Further research such as using other NLP models or creating a new algorithm needs to be investigated. Until the concern of MT not being able to 100% accurately translate between any language is developed, it is best to not use MT in a professional setting. Being able to understand the limitations of the machine transition as of now will help the audience know that current models are not perfect, and the best practice is to have a professional edit the translation or do it themselves.

# Peer Review by Amanda Burfield

**Contributions to NLP**

Phoe Aung reviewed “Phrase-Based & Neural Unsupervised Machine Translation.” The contributions identified are that Machine translation uses interpretation and comprehension of human language. The contributions of neural network and phase-based models within NLP allow use towards Machine Translation.

**Ethical Issues regarding NLP**

An ethical issue discussed within the paper included misinterpretation, inaccuracy, or unprofessional translation. Although there are advantages to Phrase-Based & Neural Unsupervised Machine Translation, the ethical concern of a possible inaccuracy is something to be reviewed and investigated.

**Topic Overview**

My topic on ULMFiT in relation to Phrase-Based & Neural Unsupervised Machine Translation is that it raises the ethical concern of inaccuracies. Due to many meanings of individual words, there is a great possibility of possible inaccurate translations.

**Peer Review by Laurence Burden**

**Contribution to NLP**

The article summary provided reviews the use of neural machine translation and phrase-based statistical machine translation in regard to overall machine translation algorithms. These efforts were made to improve machine translation without the need for one-to-one sentences between two different languages. The researchers were able to help improve our use of computers to translate words and phrases between human languages more easily.

**Ethical Issues Discussed**

Only one ethical issue was discussed in regard to these new machine translation algorithms. The developer of the model must ensure that their translations are accurate and professional. Relationships of varying degrees may rely on an accurate translation. The developer should do everything in their power to not inadvertently hurt those relationships.

**Relation to Artificial Human Attention in NLP**

# There is not a direct relationship between the MT methods discussed and artificial human attention models. The two models may be able to be used together to better translate creative works, though. Artificial human attention could be introduced to the training of the model, before the translation of languages, to determine what a human would most likely process from the story. The results could then be fed through the MT processes discussed to come to a more faithful translation to what the original author intended.

**Peer Review by Dylan Bortolon**

**Contributions to NLP**

Phoe Aunt summarizes the article “Phrase-Based & Neural Unsupervised Machine Translation”. The summary discussed many contributions to NLP, some of the contributions brought up were NMT (Neural Machine Translation) and PBSMT (Phrase-Based Statistical Machine Translation). These two new two models were a great success and actually were able to outperform most of the state-of-the-art machine translations models. Overall after research there was a great improvement to computers in terms of translation.

**Ethical Issues regarding NLP**

From my understanding there are a few ethical concerns regarding the article reviewed, which were unprofessionalism and inaccuracy. These are two important ethical concerns to have when handling and executing machine translation. The dependency of accurate translations is vital for decision making and information retention. There will need to be more research done to make sure inaccuracy and unprofessionalism is handled correctly.

**Topic Overview**

I wrote my article review on BERT, BERT and Phrase-Based & Neural Unsupervised Machine Translation don’t have a direct link to one another or really share any similarity. You can though include BERT and Phrase-Based & Neural Unsupervised Machine Translation together as BERT is pre-trained to understand phrases. The collaboration of both are still in the early stages of research but the results seem promising.

**Review of BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.**

**By: Dylan Bortolon**

**Paper Summary**

The article that I choose to review is about BERT or also known as Bidirectional Encoder Representations from Transformers. “BERT is designed to pretrain deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers”(Devlin et al, 2019). What makes BERT so special is that most language models before BERT came into existence could only read left to right and right to left. BERT makes its special appearance with the ability to read both directions, this was the first language model that can achieve non sequential text input reading.

There are two steps to the BERT framework that make it work so flawlessly to an extent. These two steps are pre-training and fine-tuning. These steps are essential for the success and progression of BERT. Pre-training is used very commonly in NLP to train a language model such as BERT on a vast amount of unlabeled data. Pre-training BERT models is important as it helps learn what the contents of the documents/data means. After the pre-training stage is finished we would start fine-tuning, this is equally just as important as pre-training. Fine-tuning is used to get desired tasks out of the pre-trained models for example you get questions answered and in depth analysis reports. BERT can function without fine-tuning but it is highly not recommended as the performance will be weak and information could be inaccurate.

One of the trained models within BERT is Masked LM or also known as MLM. This is a trained predictive language model that is actually very commonly used. How masked LM works is by masking data, it will take sentences and mask a small percentage of the words. The trained masked language model will then predict and provide recommendations for that masked word. For example I could use the sentence “The dog [MASK/BLANK] in the woods at night”. The pre-trained model will then predict that the word is “ran”. With some added fine-tuning, BERT and masked LM have changed the technology world in terms of language processing and can now deliver state-of-the-art performance on many different tasks such as text classification.

GLUE or General Language Understanding Evaluation benchmark is a set collection of datasets that are used for evaluation and training purposes for NLP models. GLUE is highly important to the overall analysis of performance for models such as BERT. To summarize the research done on this article by Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova out of 5 total systems including BERT base and BERT large both BERT systems scored higher in GLUE testing. The other systems included for comparison were Pre-OpenAI SOTA, BiLSTM+ELMo+Attn, and OpenAI GPT. BERT base scored an average of 79.6, BERT large scored an average of 82.1 and the other systems all scored an average below 76. GLUE consists of 9 different datasets with different questions to test systems like BERT. These datasets have questions that the system will try to solve such as similarity, accuracy and grammatically. As stated previously the reason why BERT outperforms other models in tests such as GLUE in terms of accuracy and performance is due to its bidirectional nature while other models only read texts sequentially.

The article doesn’t go far into the ethical concerns of BERT with NLP but that doesn’t mean that there are none. BERT with NLP has brought up many concerns as it is a very powerful model that can understand human-like functions such as generating text that are very similar to what an actual human would write. With any great powerful technological product, system, device or model there will always be concerns of it being misused. Biased results are the biggest ethical concern as BERT will only learn and understand the data that is fed or inputted to it.

**BERT usage and contributions with NLP**

Throughout this paper and the research done within by the authors/researchers you can tell that BERT has made many incredible breakthroughs within NLP and Language Modelling tasks. BERT has improved and advanced accuracy within many different NLP systems such as SQuAD v1. BERT has also actually been able to surpass and outperform humans by 2.0% which is a technological milestone.

**BERT Advantages and Disadvantages**

Anadvantage that was stated in the chosen article was that fine-tuning is very beneficial within BERT and actually relatively inexpensive. It is not just simply inexpensive but it is also resource saving. Within the article we are presented with many results graphics. The article states “All of the results in the paper can be replicated in at most 1 hour on a single Cloud TPU, or a few hours on a GPU, starting from the exact same pre-trained model” (Devlin et al, 2019). A disadvantage would be that BERT is relatively slow to train simply because of its framework and its overall large size.

**BERT Additional Research** **and Support**

From the three different articles that I have chosen regarding the same topic I would say it’s a good variety of reviews on BERT and NLP. All of the articles together do have a common ground which is how technologically advanced BERT is and how it almost outperforms anything similar. In my research they all support the aspect of NLP and also briefly speak upon the performance BERT within NLP brings by outperforming humans in questions answering tests.

**Future Research** **Regarding BERT**

Considering the facts and research that we have already done and documented, it seems like BERT has a long way to go due to its large potential. BERT as already stated can outperform humans in multiple tests including question answering, meaning that BERT is already very advanced. There will need to be a lot of research done while also exhausting a lot of resources to make BERT more advanced than it already is. Google is already actively doing research on BERT to improve it and try to implement the model more in terms of speech-text technology within its services.

# Peer Review by Amanda Burfield

**Contributions to NLP**

Dylan Bortolon reviewed the article “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” The contributions of BERT are that it has the capability to read in multiple directions whereas other technologies only read right to left or left to right. The usage of pre-training and fine-tuning in NLP allows for training on unlabeled data. BERT also allows for Masked Language Models which can predict words within a sentence as suggested by “The dog [MASK/BLANK] in the woods at night.” BERT would give the predicted word as “ran” to be inserted into the sentence allowing for text classification. Other contributions of BERT include General Language Understanding Evaluation (GLUE) which is used for evaluation and training in NLP models. BERT improved and advanced NLP systems.

**Ethical Issues regarding NLP**

The paper touched on advantages and disadvantages of NLP with the use of BERT. With additional research I found that the other ethical concerns are the misuse of the technology since it can be used to predict language and could be used to spread misinformation as well as propaganda. The main issue around it is that it could create deep cake text. Because BERT relies on the data that is inputted, it could contain other bias results based on the data that is inputted. (Frackiewicz, 2023)

**Topic overviews**

The topic I chose was article #5: “Universal Language Model Fine-tuning for Text Classification.” Dylan’s topic relates to mine because BERT dethroned ULMFiT for its capabilities. Although ULMFiT was the steppingstone in Fine-tuning for Text Classification, eventually BERT was developed and was able to do predictive text and text classification better and more efficiently than ULMFiT. Thus, ULMFiT was dethroned by BERT in Text Classification with its ability to perform in predictive text with the use of AI.

**Peer Review by Laurence Burden**

**Contribution to NLP**

BERT was designed to be the first bidirectional language model. In short, it reads text from both left to right and from right to left while training the model. This training method was shown to be better than four other standard training methods when using the General Language Understanding Evaluation (GLUE) benchmark.

**Ethical Issues Discussed**

Usage of BERT needs to have the same ethical considerations as any other natural language processing algorithm. Designers must pay special attention to removing any possible bias from their datasets and the final model. Privacy and transparency must also be considered as top priorities.

**Relation to Artificial Human Attention in NLP**

# The BERT training method is not directly related to using artificial attention in NLP model design. Both methods set out to improve NLP model creation in terms of accuracy. The BERT was shown to beat an algorithm that included artificial attention by more than 3 points when using the GLUE dataset.

# 

# Review of Sequence Classification with Human Attention

**By Laurence Burden**

## Selected Paper Summary

The provided paper was written by five researchers who had the goal of introducing natural human attention data into a natural language processing algorithm. The human attention process is described as the natural reading method by humans to omit certain function words and “focus on less predictable content words” (Barrett et al., 2018). This data is then used across NLP models to detect sentiment analysis, grammatical error detection, and detection of abusive language.

The researchers created a recurrent neural architecture using both the standard recurrent parameters and the attention function created from open-source eye-tracking data. They used a grouping of labeled sentences and sequences that associate a scalar value that represents human attention as their input set. The researchers point out that their protocol does not use, and does not need in-task eye-tracking recordings, and instead relies on already existing models of human reading attention (Barrett et al., 2018). Using these methods allows the researchers to compare human versus machine attention and to use the human attention data to influence the way a machine reads information. These methods were then used to create the various models for the three use cases they set out to compare against the usual model creation methods.

The first use case the team tackled was sentiment analysis. Sentence-level analysis was selected, rather than word or phrase analysis. They also reduced the sentiment outcome to a binary-level outcome. Either the sentiment was negative or non-negative. Next, the team looked at grammatical error detection.

Grammatical error detection software and plugins have become widely popular for both students and professionals. The team used a dataset that contains essays penned by students who were in the process of learning English. The dataset had professional annotations for where the grammar was not written correctly. The model for this task also provided a binary result. Either a sentence was considered grammatically correct or it was not. The researchers then moved on to their final use case.

Hate speech detection was the final model created by this team. They used two datasets containing 20,940 manually annotated tweets (Barrett et al., 2018). The team reduced any tweets marked as either racist or sexist into one category for their final model. The team then set about conducting their comparative experiments, after creating their three models using the human attention parameters.

The results of the test are measured using the *F*1 score. This is calculated using both precision and recall. The team’s finding is that their “human attention model, based on regularizations from mean fixation durations in publicly available eye-tracking corpora, consistently outperforms the recurrent architecture with learned attention functions” (Barrett et al., 2018). The mean error reduction over the baseline claimed by the paper is 4.5%.

## Paper Analysis

### Research Use and Contributions

The chosen paper lends its research to improving the accuracy and usefulness of natural language processing. Sentiment analysis is widely used on large amounts of text from hundreds and thousands of posts across social media to determine the public’s general feelings on a topic. Increasing the accuracy of this analysis will lead companies to have more actionable data on how to navigate a market, politicians will be able to better craft their messages, and researchers will have a better understanding of public consensus.

### Advantages and Disadvantages of the Discussed Research

The idea of having a computer read text the same way that a human does comes with multiple advantages. This allows the computer to better approximate the same understanding of a piece of text as a human would. This can lead to a greater level of agreement between what the computer says the text means and a plurality of humans.

Some disadvantages may occur when using the ideas outlined in this paper. Namely, overreliance on the accuracy of this model may be used by some governments to generate a sentiment analysis model that is used to flag those considered dissidents. The paper also states that the new model was not tested against many models that came before that also artificial human attention in their analysis. In fact, there are studies out there that suggest attention is not a reliable means to assist in neural network development

### Additional Research on the Topic

This paper is not the only one on this topic. Others have attempted similar uses of artificial human attention both before and since the creation of the reviewed paper. Sood et al. found a similar increase in *F*1 score, 5%, with their own model using artificial human attention parameters (2020). The field of NLP can clearly benefit from the use of these new parameters to help computers analyze text in a way that more closely approximates how humans do. We must ensure that we are acting ethically, regardless of how useful this research looks on the surface.

### Reflections on Ethical Considerations

The fictional character Dr. Ian Malcolm once stated, “Your scientists were so preoccupied with whether they could that they didn’t stop to think if they should” (Spielberg, 1993). No matter how useful a new technology looks on the surface, we must ensure its use is ethical in nature. The continued use of NLP models has the capacity to raise concerns about bias, privacy, and transparency (Bhattacharyya, 2023).

Bias concerns arise when a trained model provides results that are beneficial or detrimental to one group of individuals over another. Privacy concerns arise when normally private information is shared with people who should not have access to it through the use of NLP models. Finally, transparency needs to be an over-arching goal of all researchers so that anyone affected by an algorithm can know how the decision was reached.

### Future Research

More research will need to be conducted on this topic to fully know how useful artificial attention is to creating NLP models. Continuing to develop models that are harder to distinguish from real human beings may not be the best outcome for society. This process will surely help in general research of publicly available text, but there is a high likelihood of it being used for harmful outcomes, too.

***Conclusion***

Researchers will continue to make machine learning algorithms more closely model that of real human behavior. Adding parameters that mimic that of a human’s attention during reading has been shown by multiple papers to generate more accurate models. This has not been shown to be definitively true, though. The ethics of this process is an area that still requires more research, as well.

# Peer Review By Amanda Burfield

# Contributions to NLP

# Laurence Burden reviewed the article “Sequence Classification with Human Attention.” In his review it discusses the contribution of NLP to the improvement of accuracy and usefulness within NLP. Other contributions could include the ability to perform analysis to gain the public’s feelings on the topic. This provides companies with the ability to have data that will allow them to navigate marketing, politics, and additional research.

# Ethical Issues regarding NLP

# The ethical issue discussed in Laurence’s write-up discusses the idea of even though we could do something, to always ask a question of if we should do something. It’s important to constantly evaluate new technology with ideas towards bias, privacy, and transparency to ensure that the technology is being used ethically within NLP.

# Topic overviews

# Laurence’s article relates to my article on the ideas of how data is processed with the use of NLPs. It rings true that there are a lot of ethical concerns when it comes to the development of these new technologies and that with each advancement it’s imperative to monitor to ensure that it is reacting in ethical standards needed to make proper business decisions without bias, with transparency, and privacy protection.

# 

# 

# Peer Review By Dylan Bortolon

# Contributions to NLP

This article review on “Sequence Classification with Human Attention” was done by Laurence Burden. Within the summary many topics were talked about including how the use of human attention can benefit humanity in general. Human attention is used “across NLP models to detect sentiment analysis, grammatical error detection, and detection of abusive language” as the article and Laurence state. Human attention contributes to NLP by improving overall accuracy and usefulness

# Ethical Issues regarding NLP

After reviewing Laurences summary of the article I can detail that the ethical concerns with human attention fall on bias, privacy, and transparency concerns. All trained models will have ethical concerns regarding bias results simply due to the understanding that the model is only as good as the data you give it. If you provide biased data you will most likely get bais results. Privacy and transparency issues were briefly discussed in the matter of personal information not being kept private through other models and transparency concerns regarding being honest about algorithm effects.

**Topic overviews**

While BERT and artificial attention do not have a direct link to one another they do share some valuable similarities. BERT and artificial attention both improve or aim to improve the overall quality and accuracy of many NLP models. BERT also shares similarity in ethical concerns, bias and privacy are both a common concern with BERT.

**Topic # 5 by Amanda Burfield**

**Universal Language Model Fine-tuning for Text Classification**

**By Jeremy Howard and Sebastian Ruder**

**Article Summary**

The article I chose to review is the Universal Language Model File-tuning for Text Classification (ULMFiT). The purpose of ULMFiT is to apply it to a task within Natural Language Processing (NLP) and allow it to fine-tune the model. The article states it can reduce error between 18 and 24 percent on datasets. The model and coding were proposed to be open source. (Howard et al, 2018)

Computer vision (CV) models use fine-tuning on different datasets. It is rare or unlikely that the data within CV models are trained to perform a certain way. Text classification usages in real-world applications can be the spam filter in your email, fraud, and ect. The fine tuning within text classification and inductive transfers targets the first layer within a model and has been shown to be unsuccessful for NLP. (Howard et al, 2018)

The article continues to explain that the two authors don’t appear to be very confident in their ability to train a language model (LM). The two lack the knowledge on training the model themselves resulting in overfitting and catastrophic forgetting with classifiers. In turn they recommended a new method. ULMFiT will address the issues and allow for the fine-tuning for certain models and allow for six text classification tasks. (Howard et al, 2018)

They proposed the following contributions of ULMFiT which included CV-like transfer learning for NLP tasks, fine-tuning with triangular learning and gradual unfreezing to retain information and avoid catastrophic forgetting. They also state that ULMFiT will outperform six representative text classification datasets with the previously stated error reduction percentage, ablation analysis, and enable wider adoption. (Howard et al, 2018)

ULMFiT has three stages that consists of the LM being trained to capture the different layers of the language, fine-tuning the data and slanted triangular learning rates (STLR) to learn specifics, and fine-tuning the classifier with gradual unfreezing and ect. The fine tuning to each layer is emphasized in importance as well. It was also suggested that fine tuning the last layer’s learning rate would assist in fine-tuning of the other layers within the model. (Howard et al, 2018)

They later used ULMFiT in experiments to test out the text-classification in relation to real-world applications. They evaluated datasets with different numbers and documents in different lengths. Doing the evaluation allowed for them to use text classification and learning approaches to allow for sentiment analysis and question classification. The sentiment analysis used IMDB movie reviews and Yelp reviewed datasets that were compiled by Zhang et al (2015). The use of an open domain with fact questions was also applied. With elevation of the datasets, it allowed them to be placed in special text-classifications for uppercase words, repetitions, and ect. They then established baselines and comparison models between other datasets that were similar. In conclusion they reported the results with three different charts depicting validation error rate by percentage as the y-axis and # of training examples as the x-axis. They used from scratch, supervised learning, and unsupervised learning to depict the experiment. The from scratch showed a very distinct line near the top of the y-axis and decreasing whereas the supervised and unsupervised learning showed gradual decreases in reference to the validation error rate. (Howard et al, 2018)

They later concluded in the article they proposed ULMFiT as a sample-efficient transfer learning method to apply to any NLP task. They also raised the suggestions to prevent catastrophic forgetting for ULMFiT and stating once again it outperforms other transfer learning techniques with six representative text classification tasks. (Howard et al, 2018)

NLP presents the possibility of ethical concerns. Within the article, the authors discussed that the use of ULMFiT with any NLP outperforms other applications in real-world applications. The best way to ensure that it remains ethical is to ensure there is no bias within the datasets themselves. The use of an IMDB dataset is great but if you are comparing the ratings of different movies, it may require the use of other datasets that have moving ratings besides IMDB. It’s also good to ensure that there is transparency within the datasets and allows for privacy. When ensuring these things, it’s best to locate various laws as well that pertain to privacy and security and adhere to local and federal laws around them.

**ULMFiT Usage with NLP**

ULMFiT involves a transfer learning technique applied within NLP tasks. Therefore, to perform, it needs to be paired with an NLP to complete its transfer and perform its purposes. It requires various datasets and can be used in supervised or unsupervised learning processes. (Kriplani, 2019) Advantages of using ULMFiT in real world applications can revolve around social media for example. The collection of data between various posts on Facebook, Instagram, and ect and training it to perform with ULMFiT. Once trained and identifying text-classifiers, the set can then be trained to identify information for specified tasks. (Faltl, et al, 2019)

**ULMFiT Contributions to NLP**

ULMFiT’s contributions included CV-like transfer learning for NLP tasks, fine-tuning with triangular learning and gradual unfreezing to retain information and avoid catastrophic forgetting. (Howard et al, 2018) It was found later that ULMFiT was “dethroned by BERT and BERT was then dethroned by XLNet for text classification. But overall, the use of ULMFiT was a stepping stone moving forward with Universal Language Model Fine-Tuning for transfer learning. (Kriplani, 2019)

**Advantages and Disadvantages of ULMFiT to NLP**

Advantages described in the article allowed for lower percentage rates for errors as well as being a top tier use of Universal Language Model Fine-Tuning. Another advantage is that it requires a small amount of data to train for a deep learning model. Disadvantages were that it needs to be trained before use and starting from the bottom layer of data and ultimately working in reverse. Lastly, it can be timely to train before being able to utilize ULMFiT. (Karbowy, 2021)

**Other ULMFiT Articles**

The additional articles I found relating to ULMFiT gave various mixed reviews as two of them state that ULMFiT was “dethroned” by other NLP software. However, all three articles do come to the conclusion of the important impact ULMFiT had on the development of text-classification and its development in data analytics. Overall, they all seem to support the argument of the original article and allow for a better understanding of how technology is always evolving.

**Future research with ULMFiT**

# Technology and various software is constantly evolving to suit the needs of day-to-day or real life applications. The future of ULMFiT was stated in the article by Howard Kriplani stating that it was dethroned by another application and later that another application had dethroned the previous. Overall, because of the ever-evolving world of technology and various environments, the future of ULMFiT may produce more fine-tuning software down the line that will allow for more specific task identification.

**Peer Review by Laurence Burden**

**Contribution to NLP**

ULMFit was created to bring certain computer vision learning techniques to natural language processing. Specifically, the researchers used transfer learning algorithms to reduce errors in NLP by 18-24% (Howard et al., 2018). The algorithm design showed progress but was ultimately dethroned by other training algorithms. However, lessons learned from the ULMFit research can continue to be used in NLP and other machine-learning endeavors.

**Ethical Issues Discussed**

The ethical concerns discussed include privacy and bias. Bias can occur when a dataset contains data that leads the algorithm to be unjustly weighted against a group of individuals. Any data scientist must be sure to account for the possibility of bias and to follow ethical privacy and transparency guidelines.

**Relation to Artificial Human Attention in NLP**

# Both ULMFit and the introduction of artificial human attention were used to help natural language processors result in more human-like behavior. Their methods relied on introducing another step into the model training steps to increase accuracy and reduce errors. Both groups of researchers accomplished their goals but through different means.

# 

**Peer Review by Dylan Bortolon**

# Contributions to NLP

ULMFit has been a great contributor to NLP and the research/progress of it over the last couple of years. Yes, it is technically not the best anymore but it was basically the start of it all. Everything is always the best until something new and better comes out in technology. Even with ULMFit not being the best learning algorithm anymore it is still widely used in research to continue to make better algorithms and models. Gradual unfreezing was briefly spoken about in the article but I believe that it is one of the biggest contributors ULMFit brought to NLP.

# Ethical Issues regarding NLP

Ethical concerns regarding ULMFit are the same as almost all NLP’s which are bias, transparency, and privacy. If a system is biased it can end up becoming the downfall of that model or algorithm. Ensuring that ULMFit is not biased will continue to lead to its success. There are always best practices and standards to follow to ensure these concerns do not actually happen when executing actions. Being alert and attentive will ensure no ethical concerns become an ethical problem.

# Topic overviews

My article summary was regarding BERT or Bidirectional Encoder Representations from Transformers. ULMFit was eventually outperformed by BERT and a few other algorithms, in terms of text classification. ULMFit was the start of it and if it wasn't for it BERT would not be what it is today. Once it was discovered that BERT can perform bidirectional understanding it outperformed competitors such as ULMFit that uses unidirectional understanding. At the end of the day both BERT and ULMFit were a great success but BERT overall just outperformed ULMFit.

# 

# References

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019, May 24). *Bert: Pre-training of deep bidirectional Transformers for language understanding*. arXiv.org. <https://arxiv.org/abs/1810.04805>

Lutkevich, B. (2020, January 27). *What is Bert (language model) and how does it work?*. Enterprise AI. <https://www.techtarget.com/searchenterpriseai/definition/BERT-language-model#:~:text=BERT%20is%20designed%20to%20help,with%20question%20and%20answer%20datasets>.

Wei, J. (2020, September 2). *Bert: Why it’s been revolutionizing NLP*. Medium. <https://towardsdatascience.com/bert-why-its-been-revolutionizing-nlp-5d1bcae76a13>

Synced. (2019, June 26). *Best NLP model ever? Google Bert sets new standards in 11 language tasks*. Medium. <https://medium.com/syncedreview/best-nlp-model-ever-google-bert-sets-new-standards-in-11-language-tasks-4a2a189bc155>

Barrett, M., Bingel, J., Hollenstein, N., Rei, M., & Søgaard, A. (2018). *Sequence classification with human attention*. ACL Anthology. <https://aclanthology.org/K18-1030/>

Galassi, A., Lippi, M., & Torroni, P. (2021, October 10). *Attention in Natural Language Processing*. IEEE Explore. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7169508>

Sood, E., Tannert, S., Muller, P., & Bulling, A. (2020). *Improving natural language processing tasks with human gaze ... - neurips*. Proceedings Neural Information Processing Systems. <https://proceedings.neurips.cc/paper/2020/file/460191c72f67e90150a093b4585e7eb4-Paper.pdf>

Spielberg, S. (1993). *Jurassic Park.* Universal Pictures.

Bhattacharyya, S. (2023, June 13). *The ethical considerations of Natural Language Processing (NLP)*. Analytics Steps. <https://www.analyticssteps.com/blogs/ethical-considerations-natural-language-processing-nlp>

Faltl, S., Schimpke, M., & Hackober, C. (7 February 2019). *Universal Language Model Fine-Tuning. Humboldt Univesrity in Berlin*. [ULMFiT: State-of-the-Art in Text Analysis (humboldt-wi.github.io)](https://humboldt-wi.github.io/blog/research/information_systems_1819/group4_ulmfit/)

Howard, J., & Ruder, S. (23 May 2018). *Universal Language Model Fine-tuning for Text Classification.* ARXIV. [1801.06146.pdf (arxiv.org)](https://arxiv.org/pdf/1801.06146.pdf)

Karbowy, H. (12 July 2021). *Why ULMFiT can still matter: training and inference times*. Edrone. [Why ULMFiT can still matter: training and inference times | edrone | CRM for e-commerce & marketing automation](https://edrone.me/blog/why-ulmfit-can-still-matter-training-and-inference-times)

Kriplani, H. (4 September 2019). *Understanding language modelling (NLP) and Using ULMFIT.* Towards Data Science. [Understanding language modelling(NLP) and Using ULMFIT | by Himanshu Kriplani | Towards Data Science](https://towardsdatascience.com/understanding-language-modelling-nlp-part-1-ulmfit-b557a63a672b#:~:text=Universal%20Language%20Model%20FIne%2DTuning,by%20XLNet%20in%20text%20classification%5D.)

Frackiewicz, M. (6 August 2023). *BERT and the Ethics of AI Language Understanding: A Critical Discussion*. TS2. [BERT and the Ethics of AI Language Understanding: A Critical Discussion (ts2.space)](https://ts2.space/en/bert-and-the-ethics-of-ai-language-understanding-a-critical-discussion/)