Research Article

Dane Rich*

Something to Do with Paying Attention: A Review of Transformer-based Deep Neural Networks for Text Classification in Digital Humanities and New Testament Studies

https://doi.org/10.1515/opth-2025-0052
received November 10, 2024; accepted June 17, 2025

Abstract: Researchers in the field of New Testament and Religious Studies were engaged in using digital techniques from the origins of what is now known as the field of Digital Humanities (DH). Even so, New Testament researchers have not kept abreast of the tools and techniques coming out of DH research. In particular, the latest abilities of transformer-based deep neural networks (TB-DNNs) such as BERT have yet to be comprehensively applied for text classification purposes of New Testament texts. To remedy this lacuna, we offer an exploration of recent TB-DNN usage in DH for text classification to highlight its potential for NTS. On the way, we review some of the previous text classification work done in DH and NTS. Finally, we discuss some of the barriers to implementing TB-DNN models. It is hoped that this article will stimulate NTS researchers to consider TB-DNN models in their text classification work.

Keywords: digital humanities, New Testament studies, text classification, computational linguistics, transformer, deep neural networks, BERT

1 Introduction

The advent of transformer-based deep neural networks (TB-DNNs) in 2017[1] marked a distinct shift in digital humanities (DH) tooling, with particular relevance for New Testament studies (NTS). TB-DNNs have shown state-of-the-art performance in text classification (TC) tasks, such as authorship attribution, genre classification, and text dating, which are central to NTS research. Despite their relevance and being almost a decade old, NTS scholars have yet to tap into the potential of these TB-DNNs. This lag in implementation is, unfortunately, endemic to much of NTS engagement with DH techniques. We believe this hesitancy to implement DH techniques is due at least in part to a general unfamiliarity – and perhaps even distrust – of DH techniques within NTS. This article seeks to remedy this lack of familiarity and distrust and promote the use of TB-DNNs for New Testament research.

To accomplish this goal, this article is divided into three sections. The first is a brief history of DH that explores the challenges of defining the field and NTS's fundamental but tentative relationship to it. The next section constitutes the bulk of the article. It offers an introduction to the process of text classification and the

1 Vaswani et al., *Attention Is All You Need.*

*** Corresponding author: Dane Rich,** Gordon-Conwell Theological Seminary, South Hamilton, United States,
e-mail: Drich1@gordonconwell.edu

ways in which TB-DNNs have progressed it. Following this introduction, each subcategory of TC will be reviewed with relevance to NTS. It is hoped that these brief reviews will show the high relevance and potential of DH broadly, and TB-DNNs in particular, for NTS research. The last section will discuss some challenges with implementing TB-DNNs, including overcoming data scarcity issues with ancient texts, validating findings, and ethical concerns. While this review article covers much ground, it is hoped that it will function as a kind of DH primer for NTS scholars and promote the use of advanced DH techniques in the field.

2 A Brief History of Digital Humanities and New Testament Studies

Tradition places the advent of DH in the mid-twentieth century with the creation of the massive Index Thomisticus, a digitized concordance of the works of Thomas Aquinas.[2] Scholars now complicate this history by noting various other works done before and around this same time,[3] and over the last two decades, the history of DH has become a focus in its own right, with multiple works devoted to its elaboration.[4] The reader is encouraged to consult these histories for in-depth treatment of the subject.

Along with the increased attention on the history of DH has come additional reflection on how to define DH and whether it constitutes its own field. There is no real agreement about how to define DH or what belongs within its scope of practice.[5] While DH has been dominated by text analysis tasks,[6] many believe that defining DH based on this work neglects to do justice to the full scope of DH research.[7] As of now, the definition of DH is still up for debate. Though there is pushback to defining DH based on its text analysis approaches, it is these approaches that we are most interested in for this article.

What does have broad agreement – though not without critique – is the status of DH as its own legitimate field of inquiry. By analyzing the scope of journal contributions in DH, Luhmann and Manuel show that DH is not only its own legitimate field of inquiry, but is highly interdisciplinary.[8] Yet even as the field is finding its footing as an established field, some hope for the day beyond the current moment:

> we hope that the day will soon come when scholars no longer think in terms of a distinct field called the "Digital Humanities," but rather expect the Humanities to be studied using digital methods and sources[9]

This is a bold statement indeed. What seems assumed in its articulation is the challenge of integrating digital tools and methods into the humanities in a way that is "natural." The divide between the humanities and scientific endeavors like computing has at times felt antagonistic.[10] This perceived antagonism can be limiting to either field, as stereotypes around what it means to be in science versus the humanities can often discourage their integration. The presence of computing in humanities departments may feel cold or even threatening, and *vice versa*.

The field of NTS seems particularly slow in integrating DH techniques. This slow integration exists despite the fact that the history of DH is actually quite intertwined with NTS. After the creation of the Index Thomisticus, the origin of DH has been inseparably connected to Biblical and Theological Studies. In fact, many of the digital tools created in the subsequent decades were made for the study of the Bible.[11] Yet DH work

2 Busa, "The Annals of Humanities Computing."
3 Sula and Hill, "The Early History of Digital Humanities."
4 Nyhan and Flinn, *Computation and the Humanities*; McCarty, *Humanities Computing*; Hockey, "The History of Humanities Computing."
5 Terras et al., "Selected Definitions from the Day of Digital Humanities."
6 Nyhan and Flinn, *Computation and the Humanities*.
7 Sula and Hill, "The Early History of Digital Humanities."
8 Luhmann and Burghardt, "Digital Humanities – A Discipline in Its Own Right?"
9 Clivaz and Savant, "Introduction."
10 Bouterse and Karstens, "A Diversity of Divisions."
11 Clivaz and Allen, "The Digital Humanities in Biblical Studies and Theology."

in NTS has waxed and waned over the decades, with bursts of activity separated by extended periods of relative inactivity.[12] This lack of consistent engagement with DH is lamentable, as NTS scholars might not only benefit from the DHs, but NTS could, in turn, significantly contribute back to the field of DH. Many of the key issues of NTS research are also issues within the DH, and the study of ancient texts in NTS provides a unique challenge to DH techniques. Greater collaboration between the fields is likely to yield much fruit.

It is worth noting that many significant and important steps have been taken to more robustly integrate digital methods into NTS. The formation of the Centre for Digital Theology in Durham, UK was an important progression. Projects like the MARK16[13] and the publishing endeavors of Brill in the Digital Biblical Studies series. The Society of Biblical Literature (SBL) has hosted several program units on Digital Humanities in Biblical, Early Jewish, and Christian Studies. There is clearly interest and worthwhile contributions being made toward incorporating digital computing techniques in NTS. It is hoped that this article will build upon these efforts to stimulate greater integration of DH techniques into NTS. To that end, we now turn to text classification in digital humanities and New Testament studies.

3 Text Classification in Digital Humanities and New Testament Studies

Open to the introduction of any modern biblical commentary and one can discern the persistent issues and questions of the field: who wrote the text? When was it written? What is the genre? These questions all fit within the umbrella task of text classification. Text classification (TC) is concerned with sorting texts into discrete, pre-defined categories.[14] Categories relevant to New Testament Studies (NTS) include authorship, date, and genre. Each of these categories, their relevance for NTS, and their history in DH research will be discussed in the following sections, along with recent advancements in transformer-based deep neural network (TB-DNN) models. Before that, however, a brief introduction to TC is given.

3.1 Introduction to Text Classification in Digital Humanities

3.1.1 Classic Text Classification Approaches

Classical TC can be decomposed into four distinct steps: feature extraction, feature dimensionality reduction, classification method selection, and result evaluation.[15] Text pre-processing may also be necessary if the text is in poor shape, which is common for ancient texts. No matter what category one is dealing with – authorship, date, genre – these steps will be employed, with added customization as necessary. We will deal briefly with each of these steps.

Pre-processing is a step in TC where elements of a text are added or removed to make it better suited for machine processing. Common pre-processing techniques are stopword removal, stemming, and lemmatization.[16] Some ancient texts that have been digitized using document scans and optical character recognition (OCR) contain errors that may need to be corrected.[17] Pre-processing is often time-

12 Schroeder, "The Digital Humanities as Cultural Capital."
13 Clivaz, "The Impact of Digital Research."
14 Gasparetto et al., "A Survey on Text Classification Algorithms."
15 Kowsari et al., "Text Classification Algorithms."
16 Orellana et al., "A Study on the Impact of Pre-Processing Techniques in Spanish and English Text Classification over Short and Large Text Documents."
17 Tzogka et al., "OCR Workflow."

consuming, but usually necessary, although its impact on classification outcomes should be carefully considered.[18]

Feature selection and reduction is the next step in the TC process. Texts can be thought of as high-dimensional, sparse feature sets. The goal of feature selection and reduction is to isolate the most consequential elements of a text for classification purposes, as well as improve the efficiency and speed of classification methods. Classical approaches to feature selection were heavily dependent on domain experts to identify the textual features that are relevant for the classification task.[19] A recent NTS example is Robinson's attempt to discern authentic Pauline authorship using "hand-coded" texts with "twenty literary characteristics" that were "selected by learned scholars of the primary sources."[20] These literary characteristics are mostly rhetorical features of the text that Robinson believes characterizes Paul's "style." The problem with the domain expert approach is that it may be too subjective and idiosyncratic to be repeatable across multiple domains and experiments, making evaluation and benchmarking challenging.

Moving away from this domain expert approach, several features have become standard for TC purposes over the decades. These features include n-grams, bag-of-words, and frequency-based term vectors.[21] A helpful review of more computationally based feature selection methods is provided by Shah and Patel.[22] Recently, word embeddings have been used to represent document features.[23] Word embeddings have the advantage of capturing both semantic and lexical-syntactic features of a text. Word embeddings are the bedrock of modern text classification work using TB-DNN methods, as will be discussed shortly.

After features have been selected and isolated, further dimensionality reduction may be desired or necessary. Mirończuk and Protasiewicz provide a helpful list of some classic and more modern dimension reduction techniques.[24] The advantage of more modern techniques is the ability to handle high amounts of text features, removing the need for dimensionality reduction.

Choosing a classification method is the central decision of TC. Many classification methods have been designed over the decades. Gasparetto et al. have provided a very helpful overview of some of the more popular classification methods, including probabilistic classification, k-nearest neighbors (KNN), support vector machine (SVM), decision trees, and shallow neural networks.[25] Kowsari et al. add the Naïve-Bayes classifier among others and provide helpful mathematical formulations of all these techniques. When NTS has used computational methods, it has traditionally heavily relied on the more probabilistic classification techniques. Unfortunately, it has done this to the near exclusion of more machine learning-based techniques, such as KNN, SVM, shallow neural networks, and Naïve-Bayes. Neglecting to implement these more modern techniques has impoverished NTS.

The last step in TC is evaluation. Evaluation is necessary to provide confidence that classification methods are, in fact, valid and performative. It also helps to determine which methods perform better than others. Evaluation metrics are based on various combinations of the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Popular metrics based on these numbers are accuracy, recall, precision, and f-measure.[26] In reality, the task of evaluation is challenging. Datasets, benchmarking, and evaluation metrics have not been fully standardized, limiting the field's ability to perform comparative analysis. A key task for NTS is to develop robust and standardized datasets and benchmarks to further the field's ability to evaluate TC results.

---

18 Chaerul Haviana et al., "The Effects of Stopwords, Stemming, and Lemmatization on Pre-Trained Language Models for Text Classification;" Uysal and Gunal, "The Impact of Preprocessing on Text Classification."

19 Forsyth and Holmes, "Feature-Finding for Text Classification."

20 Roy and Robertson, "Applying Cosine Similarity."

21 Kowsari et al., "Text Classification Algorithms."

22 Shah and Patel, "A Review on Feature Selection and Feature Extraction for Text Classification."

23 Ge and Moh, "Improving Text Classification with Word Embedding."

24 Mirończuk and Protasiewicz, "A Recent Overview of the State-of-the-Art Elements of Text Classification."

25 Gasparetto et al., "A Survey on Text Classification Algorithms."

26 Hossin and Sulaiman, "A Review On Evaluation Metrics For Data Classification Evaluations."

3.1.2 TB-DNN Classification Approaches

Deep neural network (DNN) approaches have been around for the last two decades and have contributed significantly to TC in DH. "Deep" neural networks distinguish themselves from "shallow" neural networks with their many internal layers of neurons. These added layers contribute greatly to the pattern recognition and generalization abilities of neural networks. Modern large language models (LMMs) boast dozens to hundreds of internal layers. Kowsari et al. provide a helpful summary of the early deep learning approaches, including recurrent neural networks (RNN)[27] and long short-term memory (LSTM) networks, among others. These techniques are excellent and worth implementing in NTS. This article, however, is focused on attention-based architectures and so will not explore these approaches further.

The first attributed use of attention mechanisms in DNNs is in a machine translation article by Bahdanau.[28] Broadly conceived, attention in DNNs is the ability to focus and attend to the most meaningful aspects of a sequence in order to retain contextually significant information.[29] The significant contribution of Vaswami in 2017 was the idea that a translation model could be designed primarily using this attention mechanism.[30] The resulting model is known as the transformer. This new model is highly scalable and is the fundamental innovation enabling modern LLMs, or what we're calling TB-DNNs. The reader is referred to Gesparetto's incredibly helpful and accessible introduction to TB-DNNs in their review of modern text classification techniques.[31]

An important thing to highlight about these new TB-DNNs is their ability to perform feature-finding based on syntax *and* semantics. The enabling technology for this ability is known as word embeddings.[32] Embeddings are word representations that store hundreds or even thousands of word syntactic and semantic features in a feature space.[33] These features are learned and stored in embeddings automatically during TB-DNN model training. Thus, feature selection is no longer limited to reduced dimensionality as with classical TC methods. Rather, these embeddings may hold hundreds of features that models can train on and then represent in their model weights.

A popular and important TB-DNN model is the Bidirectional Encoder Representations from Transformers, or BERT, model.[34] As the name implies, BERT uses an encoder-only transformer architecture, which means it is intended to learn and not necessarily generate language. BERT's ability to encode language rules has made it a popular and powerful method for text analysis purposes.[35] Many versions of BERT have emerged for specialized use, including a more lightweight distilBERT,[36] an improved roBERTa model,[37] and a cross-lingual xlmBERT.[38] The following exploration of TC sub-tasks will highlight recent uses of BERT models in TC.

3.2 Sub-Tasks in Text Classification

3.2.1 Authorship Attribution

Discussion regarding the authorship of various biblical and extra-biblical books extends back into the early church.[39] The authorship of most books of the Bible is now debated, including the Synoptics, Pauline epistles,

---

27 Lipton et al., "A Critical Review of Recurrent Neural Networks for Sequence Learning."
28 Bahdanau et al., "Neural Machine Translation by Jointly Learning to Align and Translate."
29 Soydaner, "Attention Mechanism in Neural Networks."
30 Vaswani et al., *Attention Is All You Need*.
31 Gasparetto et al., "A Survey on Text Classification Algorithms."
32 Mikolov et al., "Efficient Estimation of Word Representations in Vector Space."
33 Mikolov et al., *Linguistic Regularities in Continuous Space Word Representations*.
34 Devlin et al., "BERT."
35 Rogers et al., "A Primer in BERTology."
36 Sanh et al., "DistilBERT, a Distilled Version of BERT."
37 Liu et al., "RoBERTa."
38 Conneau et al., "Unsupervised Cross-Lingual Representation Learning at Scale."
39 Reed, "Pseudepigraphy, Authorship, and the Reception of 'The Bible' In Late Antiquity."

Gospel of John, letters of John, Revelation, Hebrews, etc. While all of these authorship debates are interesting and worthwhile in their own right, the issue of Pauline authorship is particularly apt for computational approaches. This aptitude is due to the fact that multiple documents in the New Testament are attributed to Paul, even by modern scholars. Although the determination of what is "authentic" Pauline versus "inauthentic" is continuously in flux,[40] even a few confidently "authentic" documents are helpful in identifying an author's style. But just how large of a sample is required for stylometric methods to capture an author's style? According to Eder, 3,000 consecutive tokens in Ancient Greek is a helpful cutoff.[41] Since Romans – which is considered an authentic Pauline letter by both ancient and contemporary researchers – consists of over 8,000 words,[42] it is at least theoretically possible to capture Paul's authentic style.

While AA can be traced back several millennia,[43] the birth of modern DH approaches to AA is usually traced back to the nineteenth-century musings of Augustus de Morgan[44] or the mid-twentieth-century work of Mosteller and Wallace on the Federalist Papers.[45] Since these original formulations, the field of AA has inspired significant contributions. Many excellent surveys of the field have been written that trace the history of AA and its techniques.[46]

Notably for this article, over the last two decades, there has been an influx of machine-learning-based contributions to the field.[47] These contributions are significant and point to an important evolution within the field of stylometry, namely, the increasing adoption of machine-learning techniques. Some of the more popular machine-learning classification methods include those already mentioned for TC, including k-nearest neighbors, SVM, nearest shrunken centroids, and regularized discriminant analysis.[48]

As mentioned above, the Pauline question is an excellent use case for these techniques. Yet despite its appropriateness, few advanced AA techniques have been applied to it. For instance, we could find no studies applying SVM[49] or compression techniques to the Pauline question.[50] Many computational studies on Pauline authorship have, instead, relied on distance-based methods.[51] Considering how heated the debate around Pauline authorship can get, greater efforts should be made to use mitigating methods, including TB-DNN models.

3.2.1.1 TB-DNN in Authorship Attribution

Recent years have seen the advent of new transformer-based approaches to AA. Fabien et al. fined-turned a BERT model for use in AA and achieved state-of-the-art results.[52] Their work was tested against more traditional n-gram-based methods in Tyo et al.'s article.[53] Tyo et al. found that traditional methods still excel on some tasks, while BertAA proved best for AA datasets with large amounts of attributed works, and also in AV applications. They made their benchmarking and model implementations available for public use.[54]

40 White, "The Pauline Tradition," 39–53.

41 Eder, "Does Size Matter?"

42 Savoy, "Authorship of Pauline Epistles Revisited."

43 Delcourt, "Stylometry."

44 Holmes and Kardos, "Who Was the Author?"

45 Mosteller and Wallace, "Inference in an Authorship Problem."

46 Koppel et al., "Computational Methods in Authorship Attribution;" Grieve, "Quantitative Authorship Attribution;" Holmes, "The Evolution of Stylometry in Humanities Scholarship;" Neal et al., "Surveying Stylometry Techniques and Applications;" Lagutina et al., "A Survey on Stylometric Text Features;" Juola et al., "A Prototype for Authorship Attribution Studies;" Stamatatos, "A Survey of Modern Authorship Attribution Methods."

47 Muldoon et al., "Modern Stylometry;" Savoy, *Machine Learning Methods for Stylometry*.

48 Jockers and Witten, "A Comparative Study of Machine Learning Methods for Authorship Attribution."

49 Diederich et al., "Authorship Attribution with Support Vector Machines."

50 Pavelec et al., "Compression and Stylometry for Author Identification."

51 Savoy, "Authorship of Pauline Epistles Revisited;" Roy and Robertson, "Applying Cosine Similarity to Paul's Letters."

52 Fabien et al., *BertAA*.

53 Tyo et al., "On the State of the Art in Authorship Attribution and Authorship Verification."

54 https://github.com/JacobTyo/Valla.

Only a few studies have applied TB-DNN methods to works of antiquity, and none were found that applied them to New Testament texts. Yamshchikov performed AA analysis on Plutarch's work using a custom Ancient Greek BERT model.[55] The custom model was made using transfer learning to compensate for the low availability of Ancient Greek training data.[56] They then used few-shot learning to train the model on different authors' styles. Their model achieves an 80% accuracy in AA.

3.2.2 Text Dating

Text dating can be divided into two separate foci. The first focus is concerned with tracking the shift in an author's style over time and sorting their texts accordingly. This approach is known as Stylochronometry. Very little work has been done in this field, and it has yet to amass a suite of proven techniques.[57] Yet previous work shows promise,[58] and Stylochronometry as a concept appears to have theoretical backing.[59]

Paul's letters would be an interesting case to explore with Stylochronometry. For example, many debates regarding Pauline authorship of the pastoral epistles point to features such as sentence length and hapax legomenon. Changes in average sentence length and vocabulary are two concerns of Stylochronometry. Understanding how style and vocabulary shift through aging may shed much light on the authorship of these works and their inclusion in the authentic Pauline corpus. Additionally, the timeline of Paul's travels and their relationship to Acts has been long debated. Knowing the sequence in which Paul wrote his letters through Stylochronometry would be a tremendous contribution to these debates.

The second focus of text dating is categorizing texts according to larger historical time periods. Since text dating according to time periods is a classification problem, the typical DH methods for TC have been applied. Typical period classification resolution is centuries or decades.[60] With this resolution, text dating methods are relevant for NTS since NTS is often concerned with whether a New Testament text was written in the first or second century. The authorship date of Luke-Acts, for instance, is debated along these first-second century lines,[61] with Pervo being a prominent voice initiating and defending second-century dating.[62] A first-century date is perhaps still the preferred scholarly opinion.[63] Unfortunately, we found no DH text dating methods applied to questions of New Testament texts. The next section will explore TB-DNN methods for text dating that may be relevant for future NTS applications.

3.2.2.1 TB-DNN in Text Dating

This section will explore the DH text dating competition hosted by Evalita. Evalita is a conference focused on natural language processing in Italian. The competition guidelines are described in Menini et al.'s article.[64] The competition had four tiers with combinations of the same-genre vs cross-genre corpus of a single author. The targeted range of dating these corpora was coarse (a decades-sized window) to fine-grained (1- and 5-year windows).

---

55 Yamshchikov et al., *BERT in Plutarch's Shadows*.

56 See Section 3.1 of this article for a discussion on challenge of data scarcity in Ancient Greek.

57 Neal et al., "Surveying Stylometry Techniques and Applications."

58 Can and Patton, "Change of Writing Style with Time."

59 Stamou, "Stylochronometry."

60 Baledent et al., "Dating Ancient Texts;" Boldsen and Wahlberg, "Survey and Reproduction of Computational Approaches to Dating of Historical Texts;" Toner and Han, "Dating Medieval Texts by Classification with Flexible Time Intervals;" Niculae et al., *Temporal Text Ranking and Automatic Dating of Texts*.

61 Dicken, "The Author and Date of Luke-Acts."

62 Pervo and Tyson, "Dating Acts;" see also, Nordgaard Svendsen, "Luke's Readers and Josephus. Paul and Agrippa II as a Test Case."

63 Adams and Pahl, *Issues in Luke-Acts*.

64 Menini et al., *DaDoEval @ EVALITA 2020*.

Two approaches were submitted for this competition. The first was a more classical SVM model,[65] and the second was a transformer-based model.[66] The SVM model used word and character n-grams as the basis of its classification scheme. The transformer model used a Sentence-BERT (SBER)[67] model to create sentence embeddings of the text. Sentence-BERT takes a regular BERT model and uses it to embed entire sentences for classification. This article also used a bag-of-entities approach to analyze named entities. Named entities were used to help identify periods in which the corpus could be classified.

As expected, both submissions struggled with cross-genre dating and fine-grained dating. The performance on cross-genre dating dropped by nearly 100% compared to same-genre. This dramatic drop in accuracy indicates that dating documents across different genres significantly increases the difficulty of the problem space. Interestingly, the challenges of cross-genre dating seemed to come from syntactic rather than semantic factors. Across the genre space, vocabulary remained fairly constant, while sentence complexity varied significantly.

Between the two models, the SVM model performed better than the SBERT model. In another article by Westin, SBERT was deployed for the time-period classification of fiction literature[68] and also performed poorly on this task compared to other methods, such as Time Frequency-Inverse Document Frequency (TD-IDF) and Latent Dirichlet Allocation (LDA). From these results, it appears that SBERT is a poor model for classifying works by time period. Additional research is needed to identify how BERT might be best modified for text dating purposes.

Some intriguing work has been done by Ren et al. on the semantic shift of words over time.[69] Their article proposes a new AI architecture that includes an analysis of the semantic evolution of words over time. With this method they are able to achieve state of the art results as compared to other models, such as SBERT.

3.2.3 Genre Classification

Genre is a fascinating and frequently debated aspect of NTS. While much of New Testament literature exemplifies the various genres of the time, it also adapts and expands these genres for its own purposes, making neat classification challenging. Texts such as Revelation[70] and Luke-Acts[71] pose particularly challenging problems, while the genre of "gospel,"[72] questions fundamental genre categories. In general, genre classification can be viewed as another offshoot of the broader text classification task, making it apt for DH classification techniques. But genre is more challenging than other TC categories because genre is such a difficult category to define.[73] Therefore, applying TC techniques to the genre of the New Testament is precarious, but has great potential.

Genre classification has seen DH contributions over the last few decades. Kuzman and Ljubešić provide a helpful discussion about the various approaches taken in the literature and associated datasets.[74] They also provide a helpful review of the various machine learning techniques utilized for genre classification, including SVMs, Naïve-Bayes, and fastText – fastText being superior to the previous two. They also note that transformer-based machine learning models have shown step-up performance over the previous generation of automated approaches. They speculate that this improvement is due to the fact that these models utilize both semantic information and feature frequency, while previous methods relied heavily on feature frequency.

---

65 Brivio, "Matteo-Brv @ DaDoEval."

66 Massidda, "Rmassidda @ DaDoEval."

67 Reimers and Gurevych, "Sentence-BERT."

68 Westin, "Time Period Categorization in Fiction."

69 Ren et al., *Time-Aware Language Modeling for Historical Text Dating*.

70 Reddish, "The Genre of the Book of Revelation."

71 Smith and Kostopoulos, "Biography, History and the Genre of Luke-Acts."

72 Calhoun et al., *Modern and Ancient Literary Criticism of the Gospels*.

73 Chandler, "An Introduction to Genre Theory."

74 Kuzman and Ljubešić, "Automatic Genre Identification."

3.2.3.1 TB-DNN in Genre Classification

Much DNN work on genre is focused on "register" identification of web-content. For our purposes, register and genre may be considered interchangeable, as both seek to classify texts based on stylistic differences that are context dependent.[75] TB-DNN models have been applied to this task of register/genre classification with recent work exploring performance improvements through multi- and cross-lingual transfer models.[76] Kuzman et al. show that BERT-based models not only perform better in genre classification than traditional methods, but they also generalize better to out-of-domain texts.[77] We found no examples of TB-DNN work applied to questions of New Testament genres.

4 Challenges with Implementing TB-DNNs in New Testament Studies

So far in this article, we have reviewed the importance of text classification in digital humanities and New Testament Studies. By showing the importance of text classification for both fields, we have highlighted the ways in which DH techniques and tools have been and may be used in New Testament Studies. We have also highlighted the growing work done using TB-DNN models on questions of TC. While exploring these DH techniques, it has become clear that NTS have not consistently stayed abreast of the tools and techniques that DH affords. In failing to implement these tools, NTS has not enjoyed their potential benefits. To remedy this lack of implementation, we have introduced the latest suite of TB-DNN tools and how they might be utilized for NTS. We have shown that these new models are powerful and produce state-of-the-art results in TC. We believe their use in NTS will greatly enhance the field and inform its future directions.

Despite they're great potential, it is not simple to use these TB-DNN models. They are challenging to understand and are actively being developed and changed at a rapid rate. Before implementing them, there is a need for data pre-processing, model tuning, database curation, and, importantly, ethical considerations. The goal of the next section is to address some of these challenges. In particular, we will discuss data scarcity issues, validation considerations, and ethical concerns.

4.1 Data Scarcity Considerations

One challenge in working with ancient texts is the scarcity of labeled data. For example, the Disorios Ancient Greek corpus, a set of Ancient Greek texts dating from Homer to the 5th century AD, contains roughly 10 million tokens.[78] As of 2022, the Perseus library boasts 32 million Greek words in its collection.[79] The Thesaurus Linguae Graecae nearly quadruples the Perseus library with 125 million words in Ancient Greek, but it includes texts up through the Byzantine period and even some into the nineteenth century.[80] These corpora are significant, but the size disparity is apparent when compared to the typical size of training data used for TB-DNNs. BERT, for instance, was trained on 3.3 *billion* tokens of unlabeled data.

Solutions to this data problem will have to be creative. The corpus of Ancient Greek will not change dramatically, so trying to compete with the accuracy of modern models when training a BERT model on Koine Greek is a non-starter. Thankfully, much work has gone into exploring how to perform analysis and various

75 Ibid.

76 Repo et al., "Beyond the English Web;" Rönnqvist et al., *Multilingual and Zero-Shot Is Closing in on Monolingual Web Register Classification*.

77 Kuzman et al., "Automatic Genre Identification for Robust Enrichment of Massive Text Collections."

78 Vatri and McGillivray, "The Diorisis Ancient Greek Corpus."

79 Perseus News and Updates. Perseus Digital Library. (2024, January 15). http://www.perseus.tufts.edu/hopper/#:~:text=The%20most%20up%2Dto%2Ddate,and%20sources%20in%20other%20languages.

80 The thesaurus Linguae Graecae®: Our mission and our projects. TLG. (n.d.). https://stephanus.tlg.uci.edu/tlg.php.

tasks on domains with limited data. This field is broadly known as transfer or low-resource learning. Transfer learning takes a model originally trained on one domain and adapts it to work with another domain.

There are various techniques for accomplishing transfer learning. For text analysis on Ancient Greek, Yamshchikov used the technique of fine-tuning to modify Greek BERT into an Ancient Greek BERT model.[81] A similar approach was undertaken by Krahn et al.[82] Both models are freely available online.[83] Another technique that this article will cover is known as meta-learning. Both transfer learning and meta-learning have been used extensively in text classification tasks.

4.1.1 Transfer Learning

As stated earlier, transfer learning involves taking a model trained on one set of domain data and using it in a new domain. There are several ways to perform transfer learning on a model. A method known as sequential transfer learning involves pre-training a model on a large amount of unlabeled data and then fine-tuning the model for the new domain.[84] Pre-training a BERT model is usually done using two techniques: language modeling and masked language modeling.[85] These techniques help BERT to learn general linguistic and syntactic patterns in language. Once the model has acquired general linguistic skills, it can be further fine-tuned for task-specific applications,[86] including text classification.[87] A challenge with fine-tuning is that it can be computationally expensive.

4.1.2 Meta-Learning

Meta-learning is all about models "learning to learn."[88] In other words, creating models that don't simply hard-code parameter weights based on training datasets but also have structures that can generalize to many kinds of data, even with limited exposure. Many meta-learning models have been proposed, including metric-based,[89] optimization-based (i.e., model-agnostic meta-learning (MAML)),[90] transfer-based,[91] self-supervised,[92] and memory-based.[93] Meta-learning has emerged as a powerful way to augment models to work in low-resource languages for TC tasks.[94] A particularly significant application of meta-learning is the ability to use high-resourced languages to train models on broad linguistic rules and then utilize that linguistic knowledge to learn a low-resourced language – such as Koine Greek. This technique is known as cross-lingual transfer.[95] Some BERT models – such as mBERT[96] and XLM-R[97] – are multilingual and are excellent candidates

---

81 Yamshchikov et al., *BERT in Plutarch's Shadows*.
82 Krahn et al., "Sentence Embedding Models for Ancient Greek Using Multilingual Knowledge Distillation."
83 https://huggingface.co/altsoph/bert-base-ancientgreek-uncased; https://github.com/kevinkrahn/ancient-greek-datasets.
84 Ruder et al., *Transfer Learning in Natural Language Processing*.
85 Devlin et al., "BERT."
86 Raffel et al., "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer."
87 Sun et al., *How to Fine-Tune BERT for Text Classification?*
88 Lee et al., "Meta Learning for Natural Language Processing;" Yin, "Meta-Learning for Few-Shot Natural Language Processing;" Li et al., "A Concise Review of Recent Few-Shot Meta-Learning Methods."
89 Gharoun et al., "Meta-Learning Approaches for Few-Shot Learning;" Yin, "Meta-Learning for Few-Shot Natural Language Processing."
90 Devos and Dandi, "Model-Agnostic Learning to Meta-Learn."
91 Sun et al., "Meta-Transfer Learning Through Hard Tasks."
92 Bansal et al., "Self-Supervised Meta-Learning for Few-Shot Natural Language Classification Tasks."
93 Santoro et al., "Meta-Learning with Memory-Augmented Neural Networks."
94 Amity School of Engineering & Technology and Kumari, "Exploring the Potential of Meta-Learning in Natural Language Processing."
95 Liu et al., "A Study of Cross-Lingual Ability and Language-Specific Information in Multilingual BERT."
96 Pires et al., "How Multilingual Is Multilingual BERT?"
97 Conneau et al., "Unsupervised Cross-Lingual Representation Learning at Scale."

for this technique. Meta-learning is an active area of research and holds much promise for NTS research, in which small datasets and low-resourced languages are a chronic issue.

4.1.3 GAN-BERT

A powerful method to improve model text classification results on limited labeled datasets is GAN-BERT. GAN stands for General Adversarial Network and was introduced in 2014 by Goodfellow et al.[98] GANs can produce new data that simulates real labeled data by using two components – a generator and a discriminator. The generator creates synthetic data from a randomized vector input. The synthetic data are then fed into a discriminator, which tries to tell whether the data are synthetic or real. The generator aims to create something so similar to real data that it fools the discriminator into believing the "fake" data are real. On the other hand, the discriminator attempts to accurately label the generator's output as fake while labeling the real data as real. Both models are working against each other in a game-like min-max scenario. Ideally, the generator will become good enough that the discriminator cannot distinguish its output from real data.

Using this basic approach, Croce and Basili introduced GAN-BERT.[99] Croce and Basili use a BERT model as the discriminator/classifier in the two-part GAN architecture. Since the BERT model not only discriminates but also classifies the input data, it enables it to improve its ability to identify not only real data from fake but also its ability to classify texts. GAN-BERT is particularly useful in situations in which labeled data are scarce. Some applications of this model have been implemented with high success rates, including in hate speech detection,[100] intent classification,[101] and authorship attribution.[102]

4.2 Validity Considerations

A glaring challenge to using TB-DNN models in research is establishing their trustworthiness. A model may assert that a particular letter in Scripture is Pauline, but how is that verified? How does one trust that result? Trust comes from continued rigorous assessment and peer review, but more involvement from the NTS and DH community is needed to foster this kind of trust. Another important component of model validation is universal benchmarking datasets. A popular benchmark for linguistic competence is the General Language Understanding Evaluation (GLUE) benchmark.[103] GLUE evaluates models based on nine diagnostic tests, which assess the model's natural language understanding. This benchmark is one of the most widely used and is a good indicator of a model's linguistic trustworthiness.

An interesting experiment was performed by Nangia et al. where humans were tested on the GLUE benchmark and the results were compared to BERT's performance. The researchers concluded, "we find that state-of-the-art models like BERT are not far behind human performance on most GLUE tasks."[104] This article was written in 2019, shortly after BERT was first introduced, and recent articles have asserted BERT's superiority in human comparisons.[105] This finding is a strong indicator of BERT's trustworthiness as a research tool.

Finally, model validity is often assessed on task-specific applications. For example, Yamshchikov et al. validated their Greek BERT classifier on ancient works of known authors before using their model to

98 Goodfellow et al., "Generative Adversarial Networks."

99 Croce et al., "GAN-BERT."

100 Mnassri et al., *Multilingual Hate Speech Detection Using Semi-Supervised Generative Adversarial Network*.

101 Tanvir et al., *A GAN-BERT Based Approach for Bengali Text Classification with a Few Labeled Examples*.

102 Silva et al., *Forged-GAN-BERT*.

103 Wang et al., "GLUE;" Hossin and Sulaiman, "A Review on Evaluation Metrics for Data Classification Evaluations."

104 Nangia and Bowman, "Human vs Muppet."

105 Yao and Yuan, "Research on the Application and Optimization Strategies of Deep Learning in Large Language Models."

discriminate authorship attribution of Plutarch's contested works. They're assessment found high validity in their model.[106] Fabien et al. validated their BERTAA model using a dataset of IMDb blog posts with known authors.[107] While a researcher cannot claim with certainty that the results of a valid model are accurate, a rigorously assessed model improves its outcome's trustworthiness and academic esteem. As models improve and become more accurate/valid, their use in making assertions will become increasingly important.

4.3 Ethical Considerations

With the increased usage of DNNs in society, there ought to be equally maturing reflections on their ethical implications. Efforts to offer ethical groundwork are being made in the fields of medicine,[108] education,[109] academics,[110] and society broadly.[111] Much of this reflection, however, is geared toward the generative aspect of DNNs, i.e., the use of ChatGPT for academic research. The ethical use of DNNs for analysis purposes – such as text classification – has had less scrutiny. In this section, we will offer some preliminary considerations regarding the use of DNNs in NTS research, although a fuller treatment from ethicists is greatly needed. Below, we consider three ethical concerns of using DNNs in NTS: big data, cultural humility, and model bias.

4.3.1 The Ethics of Big Data

Central to the ethical heart of DNNs is its reliance on "big data."[112] Gonzalez and Rodrigues wrote a helpful paper considering the use of big data in DH.[113] They describe the way in which information may go on a "data journey" that de-contextualizes and re-contextualizes it. This data journey becomes even more complex with the development of DNNs, which absorb massive amounts of data into various matrix weights. Much legal and ethical consideration is happening around the rights of companies to mine data across the internet without permission from users, particularly from those writing personal or intellectual property. These data are often used to train the most powerful DNNs – DNNs that researchers are most likely to use. While these issues are being sorted out in society, DH and NTS researchers who utilize these tools ought to recognize how they function within the evolving ethical conversation. If certain tools or models are deemed to be unethically sourced, DH and NTS researchers ought not to use them and perpetuate continued exploitation. In the future, DH and NTS researchers should consider training and establishing open-source DNNs that are powerful, ethical, and transparent.

4.3.2 Cultural Humility

The use of computational methods to explore and study human creations should be done with appropriate sensitivity. Humans and their creations are not reducible to numbers, even if numbers may be used to describe their movements and behaviors. As such, tasks such as attributing texts to one author or another, or accusing a supposed author of plagiarism, ought to be done carefully and considerately. Especially when dealing with ancient texts, where authors have no voice and data can be scarce, conclusions should be humbly

106 Yamshchikov et al., *BERT in Plutarch's Shadows*.
107 Fabien et al., *BertAA*.
108 He et al., "A Survey of Large Language Models for Healthcare."
109 Yan et al., "Practical and Ethical Challenges of Large Language Models in Education."
110 Lund et al., "ChatGPT and a New Academic Reality."
111 Weidinger et al., "Ethical and Social Risks of Harm from Language Models."
112 Pence, "What Is Big Data and Why Is It Important?"
113 Gonzalez and Rodrigues, "Digital Humanities."

regarded and carefully considered. Additionally, the field of NTS is not only concerned with historical and anthropological questions, but also questions involving ethnic and religious group identity. While the use of computational tools to discover true realities related to religious and ethnic concerns is a worthwhile endeavor, these computational tools offer no consideration to those communities practicing authentic faith or culture. In light of these tools' inabilities, researchers ought to show appropriate deference to the religions and communities they research.

4.3.3 Bias in DNNs

One last consideration regards the biases endemic to DNNs. While it may be tempting to see these models as unbiased mediators of persistent debates, the reality is that these models have multiple layers of biases.[114] These biases originate not only from the selected training data used to train them, but also from the philosophy of their design. The fundamental mathematics and training of DNNs make claims regarding the structure and nature of language. While these presuppositions may lead to the successful generation and analysis of language, this success does not necessarily imply that the model's philosophy of language is correct. It may simply be excellent at faking or approximating language. Regardless, since these models are fundamentally based on trained data from our world, as long as inequities and biases persist in society, they will find their way into the training and implementation of DNNs. Researchers ought to be aware of these biases and seek to create fair and equitable training sets for training and benchmarking purposes.

5 Conclusion

This article has explored the use of transformer-based deep neural networks for New Testament studies. It has situated the use of digital and computational tools in New Testament studies within their broader use in digital humanities. In doing so, we have shown that New Testament researchers have failed to stay abreast of the advances in the field of digital humanities. In particular, very little work in New Testament studies has utilized the latest suite of advanced tools, namely, transformer-based deep neural networks such as BERT. It is recommended that New Testament researchers consider using these tools to augment and complement their work. Doing so, however, is not simple. We discuss barriers to using models such as BERT, including data scarcity, validation challenges, and ethical considerations. If deep neural network models are adopted for New Testament studies, there will inevitably be many pain points. Over time, however, these will give way to worthwhile contributions to the field and greater heights of discovery.

Acknowledgments: I am deeply grateful and indebted to Kevin Krahn and his patient proof-reading of this paper. I am also thankful to Erich Pracht and his invitation to submit for this volume.

Funding information: Author states no funding involved.

Author contribution: The author confirms the sole responsibility for the conception of the study, the presented results, and the manuscript preparation.

Conflict of interest: Author states no conflict of interest.

Data availability statement: Data sharing is not applicable to this article.

114 Navigli et al., "Biases in Large Language Models."

References

Adams, Sean A. and Michael Pahl. *Issues in Luke-Acts: Selected Essays*. Gorgias Handbooks 26. Piscataway, NJ: Gorgias Press, 2012.
Amity School of Engineering & Technology and Lalita Kumari. "Exploring the Potential of Meta-Learning in Natural Language Processing." *International Journal of Scientific Research in Engineering and Management* 7, no. 5 (2023). doi: 10.55041/IJSREM21190.
Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural Machine Translation by Jointly Learning to Align and Translate." 2016. doi: 10.48550/arXiv.1409.0473.
Baledent, Anaëlle, Nicolas Hiebel, and Gaël Lejeune. "Dating Ancient Texts: An Approach for Noisy French Documents." Marseille, France, 2020. https://hal.science/hal-02571633.
Bansal, Trapit, Rishikesh Jha, Tsendsuren Munkhdalai, and Andrew McCallum. "Self-Supervised Meta-Learning for Few-Shot Natural Language Classification Tasks." 2020. doi: 10.48550/arXiv.2009.08445.
Boldsen, Sidsel and Fredrik Wahlberg. "Survey and Reproduction of Computational Approaches to Dating of Historical Texts." In *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, edited by Simon Dobnik and Lilja Øvrelid, 145–56. Reykjavik, Iceland (Online): Linköping University Electronic Press, 2021. https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-471190.
Bouterse, Jeroen and Bart Karstens. "A Diversity of Divisions: Tracing the History of the Demarcation between the Sciences and the Humanities." *Isis* 106, no. 2 (2015), 341–52. doi: 10.1086/681995.
Brivio, Matteo. "Matteo-Brv @ DaDoEval: An SVM-Based Approach for Automatic Document Dating." In *EVALITA Evaluation of NLP and Speech Tools for Italian - December 17th, 2020*, edited by Valerio Basile, Danilo Croce, Maria Maro, and Lucia C. Passaro, 398–402. Torino: Accademia University Press, 2020. doi: 10.4000/books.aaccademia.7593.
Busa, R. "The Annals of Humanities Computing: The Index Thomisticus." *Computers and the Humanities* 14, no. 2 (1980), 83–90. https://www.jstor.org/stable/30207304.
Calhoun, Robert Matthew, David P. Moessner, and Tobias Nicklas. *Modern and Ancient Literary Criticism of the Gospels: Continuing the Debate on Gospel Genre(s)*. Wissenschaftliche Untersuchungen Zum Neuen Testament 451. Tübingen: Mohr Siebeck, 2020.
Can, Fazli and Jon M. Patton. "Change of Writing Style with Time." *Computers and the Humanities* 38, no. 1 (2004), 61–82. doi: 10.1023/B:CHUM.0000009225.28847.77.
Chaerul Haviana, Sam Farisa, Sri Mulyono, and Badie'Ah. "The Effects of Stopwords, Stemming, and Lemmatization on Pre-Trained Language Models for Text Classification: A Technical Study." In *2023 10th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 521–27, 2023. doi: 10.1109/EECSI59885.2023.10295797.
Chandler, Daniel. "An Introduction to Genre Theory." 2004. https://www.semanticscholar.org/paper/An-Introduction-to-Genre-Theory-Chandler/a98beb1ba5c1ca683993c10c071601fe4ed832b5.
Clivaz, Claire. "The Impact of Digital Research: Thinking about the MARK16 Project." *Open Theology* 5, no. 1 (2019), 1–12. doi: 10.1515/opth-2019-0001.
Clivaz, Claire and Garrick V. Allen. "The Digital Humanities in Biblical Studies and Theology." *Open Theology* 5, no. 1 (2019), 461–65. doi: 10.1515/opth-2019-0035.
Clivaz, Claire and Sarah Bowen Savant. "Introduction: The Dissemination of the Digital Humanities within Research on Biblical, Early Jewish and Christian Studies." In *Ancient Manuscripts in Digital Culture*, edited by David Hamidović, Claire Clivaz, and Sarah Bowen Savant, 1–13. Boston, MA: Brill, 2019. doi: 10.1163/9789004399297_002.
Conneau, Alexis, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. "Unsupervised Cross-Lingual Representation Learning at Scale." 2020. doi: 10.48550/arXiv.1911.02116.
Croce, Danilo, Giuseppe Castellucci, and Roberto Basili. "GAN-BERT: Generative Adversarial Learning for Robust Text Classification with a Bunch of Labeled Examples." In *Online: Association for Computational Linguistics*, edited by Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, 2114–9, 2020. doi: 10.18653/v1/2020.acl-main.191.
Delcourt, Christian. "Stylometry." *Revue belge de philologie et d'histoire* 80, no. 3 (2002), 979–1002. doi: 10.3406/rbph.2002.4651.
Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, volume 1 (long and short papers), 4171–86, 2019. doi: 10.18653/v1/N19-1423.
Devos, Arnout and Yatin Dandi. "Model-Agnostic Learning to Meta-Learn." In *NeurIPS 2020 Workshop on Pre-Registration in Machine Learning*, 155–75. PMLR, 2021. https://proceedings.mlr.press/v148/devos21a.html.
Dicken, Frank. "The Author and Date of Luke-Acts: Exploring the Options." In *Issues in Luke-Acts: Selected Essays*, edited by Sean A. Adams, Michael Pahl, F. Scott Spencer, Karl Shuve, Brandon Crowe, Kenneth Litwak, Frank Dicken, et al., 7–26. Piscataway, NJ: Gorgias Press, 2012. https://www.degruyter.com/document/doi/10.31826/9781463223984-005/html.
Diederich, Joachim, Jörg Kindermann, Edda Leopold, and Gerhard Paass. "Authorship Attribution with Support Vector Machines." *Applied Intelligence* 19, no. 1 (2003), 109–23. doi: 10.1023/A:1023824908771.
Eder, Maciej. "Does Size Matter? Authorship Attribution, Small Samples, Big Problem." *Digital Scholarship in the Humanities* 30, no. 2 (2015), 167–82. doi: 10.1093/llc/fqt066.
Fabien, Maël, Esau Villatoro-Tello, Petr Motlicek, and Shantipriya Parida. *BertAA: BERT Fine-Tuning for Authorship Attribution*, edited by Pushpak Bhattacharyya, Dipti Misra Sharma, and Rajeev Sangal, 127–37. Indian Institute of Technology Patna, Patna, India: NLP Association of India (NLPAI), 2020. https://aclanthology.org/2020.icon-main.16/.

Forsyth, Richard S. and David I. Holmes. “Feature-Finding for Text Classification.” *Literary and Linguistic Computing* 11, no. 4 (1996), 163–74. doi: 10.1093/llc/11.4.163.

Gasparetto, Andrea, Matteo Marcuzzo, Alessandro Zangari, and Andrea Albarelli. “A Survey on Text Classification Algorithms: From Text to Predictions.” *Information* 13, no. 2 (2022), 83. doi: 10.3390/info13020083.

Ge, Lihao and Teng-Sheng Moh. “Improving Text Classification with Word Embedding.” In *2017 IEEE International Conference on Big Data (Big Data)*, 1796–805, 2017. doi: 10.1109/BigData.2017.8258123.

Gharoun, Hassan, Fereshteh Momenifar, Fang Chen, and Amir H. Gandomi. “Meta-Learning Approaches for Few-Shot Learning: A Survey of Recent Advances.” *ACM Computing Surveys* 56, no. 12 (2024), 1–41. doi: 10.1145/3659943.

Gonzalez, Maria Eunice and Mariana Vitti Rodrigues. “Digital Humanities: Ethical Implications and Interdisciplinary Challenges.” *Humanities Bulletin* 5, no. 1 (2022), 111–25. http://journals.lapub.co.uk/index.php/HB/article/view/2357.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. “Generative Adversarial Networks.” *Communications of the ACM* 63, no. 11 (2020), 139–44. doi: 10.1145/3422622.

Gorovaia, Svetlana, Gleb Schmidt, and Ivan P. Yamshchikov. *Sui Generis: Large Language Models for Authorship Attribution and Verification in Latin*, edited by Mika Hämäläinen, Emily Öhman, So Miyagawa, Khalid Alnajjar, and Yuri Bizzoni, 398–412. Miami, USA: Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.nlp4dh-1.39.

Grieve, Jack. “Quantitative Authorship Attribution: An Evaluation of Techniques.” *Literary and Linguistic Computing* 22, no. 3 (2007), 251–70. doi: 10.1093/llc/fqm020.

He, Kai, Rui Mao, Qika Lin, Yucheng Ruan, Xiang Lan, Mengling Feng, and Erik Cambria. “A Survey of Large Language Models for Healthcare: From Data, Technology, and Applications to Accountability and Ethics.” *Information Fusion* 118 (2025), 102963. doi: 10.1016/j.inffus.2025.102963.

Hockey, Susan. “The History of Humanities Computing.” In *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth, 1–19. Malden, MA, USA: Blackwell Publishing Ltd, 2004. doi: 10.1002/9780470999875.ch1.

Holmes, David I. “The Evolution of Stylometry in Humanities Scholarship.” *Literary and Linguistic Computing* 13, no. 3 (1998), 111–7. doi: 10.1093/llc/13.3.111.

Holmes, David I. and Judit Kardos. “Who Was the Author? An Introduction to Stylometry.” *Chance* 16, no. 2 (2003), 5–8. doi: 10.1080/09332480.2003.10554842.

Hossin, Mohammad and Sulaiman MN. “A Review on Evaluation Metrics for Data Classification Evaluations.” *International Journal of Data Mining & Knowledge Management Process* 5, no. 2 (2015), 1–11. doi: 10.5121/ijdkp.2015.5201.

Jockers, Matthew L. and Daniela M. Witten. “A Comparative Study of Machine Learning Methods for Authorship Attribution.” *Literary and Linguistic Computing* 25, no. 2 (2010), 215–23. doi: 10.1093/llc/fqq001.

Juola, Patrick, John Sofko, and Patrick Brennan. “A Prototype for Authorship Attribution Studies.” *Literary and Linguistic Computing* 21, no. 2 (2006), 169–78. doi: 10.1093/llc/fql019.

Koppel, Moshe, Jonathan Schler, and Shlomo Argamon. “Computational Methods in Authorship Attribution.” *Journal of the American Society for Information Science and Technology* 60, no. 1 (2009), 9–26. doi: 10.1002/asi.20961.

Kowsari, Kamran, Kiana Jafari Meimandi, Mojtaba Heidarysafa, Sanjana Mendu, Laura Barnes, and Donald Brown. “Text Classification Algorithms: A Survey.” *Information* 10, no. 4 (2019), 150. doi: 10.3390/info10040150.

Krahn, Kevin, Derrick Tate, and Andrew C. Lamicela. “Sentence Embedding Models for Ancient Greek Using Multilingual Knowledge Distillation.” 2023. doi: 10.48550/arXiv.2308.13116.

Kuzman, Taja and Nikola Ljubešić. “Automatic Genre Identification: A Survey.” *Language Resources and Evaluation* 59, no. 1 (2025), 537–70. doi: 10.1007/s10579-023-09695-8.

Kuzman, Taja, Igor Mozetič, and Nikola Ljubešić. “Automatic Genre Identification for Robust Enrichment of Massive Text Collections: Investigation of Classification Methods in the Era of Large Language Models.” *Machine Learning and Knowledge Extraction* 5, no. 3 (2023), 1149–75. doi: 10.3390/make5030059.

Lagutina, Ksenia, Nadezhda Lagutina, Elena Boychuk, Inna Vorontsova, Elena Shliakhtina, Olga Belyaeva, Ilya Paramonov, and P. G. Demidov. “A Survey on Stylometric Text Features.” In *2019 25th Conference of Open Innovations Association (FRUCT)*, 184–95, 2019. doi: 10.23919/FRUCT48121.2019.8981504.

Lee, Hung-yi, Shang-Wen Li, and Ngoc Thang Vu. “Meta Learning for Natural Language Processing: A Survey.” 2022. doi: 10.48550/arXiv.2205.01500.

Li, Xiaoxu, Zhuo Sun, Jing-Hao Xue, and Zhanyu Ma. “A Concise Review of Recent Few-Shot Meta-Learning Methods.” *Neurocomputing* 456 (2021), 463–8. doi: 10.1016/j.neucom.2020.05.114.

Lipton, Zachary C., John Berkowitz, and Charles Elkan. “A Critical Review of Recurrent Neural Networks for Sequence Learning.” 2015. doi: 10.48550/arXiv.1506.00019.

Liu, Chi-Liang, Tsung-Yuan Hsu, Yung-Sung Chuang, and Hung-Yi Lee. “A Study of Cross-Lingual Ability and Language-Specific Information in Multilingual BERT.” 2020. doi: 10.48550/arXiv.2004.09205.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. “RoBERTa: A Robustly Optimized BERT Pretraining Approach.” 2019. doi: 10.48550/arXiv.1907.11692.

Luhmann, Jan and Manuel Burghardt. “Digital Humanities – A Discipline in Its Own Right? An Analysis of the Role and Position of Digital Humanities in the Academic Landscape.” *Journal of the Association for Information Science and Technology* 73, no. 2 (2022), 148–71. doi: 10.1002/asi.24533.

Lund, Brady D., Ting Wang, Nishith Reddy Mannuru, Bing Nie, Somipam Shimray, and Ziang Wang. "ChatGPT and a New Academic Reality: Artificial Intelligence-Written Research Papers and the Ethics of the Large Language Models in Scholarly Publishing." *Journal of the Association for Information Science and Technology* 74, no. 5 (2023), 570–81. doi: 10.1002/asi.24750.

Massidda, Riccardo. "Rmassidda @ DaDoEval: Document Dating Using Sentence Embeddings at EVALITA 2020." In *EVALITA Evaluation of NLP and Speech Tools for Italian - December 17th, 2020*, edited by Valerio Basile, Danilo Croce, Maria Maro, and Lucia C. Passaro, 403–8. Torino: Accademia University Press, 2020. doi: 10.4000/books.aaccademia.7603.

McCarty, Willard. *Humanities Computing*. Basingstoke [England]: Palgrave Macmillan, 2005. http://catdir.loc.gov/catdir/enhancements/fy0624/2005043356-t.html.

Mealand, David. "Measuring Genre Differences in Mark with Correspondence Analysis." *Literary and Linguistic Computing* 12, no. 4 (1997), 227–45. doi: 10.1093/llc/12.4.227.

Menini, Stefano, Giovanni Moretti, Rachele Sprugnoli, and Sara Tonelli. "DaDoEval @ EVALITA 2020: Same-Genre and Cross-Genre Dating of Historical Documents." In *Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020)*, 391–7. Torino, Italy: Accademia University Press, 2020.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. "Efficient Estimation of Word Representations in Vector Space." 2013. doi: 10.48550/arXiv.1301.3781.

Mikolov, Tomas, Wen-Tau Yih, and Geoffrey Zweig. *Linguistic Regularities in Continuous Space Word Representations*, edited by Lucy Vanderwende, Hal Daumé III, and Katrin Kirchhoff, 746–51. Atlanta, Georgia: Association for Computational Linguistics, 2013. https://aclanthology.org/N13-1090/.

Mirończuk, Marcin Michał and Jarosław Protasiewicz. "A Recent Overview of the State-of-the-Art Elements of Text Classification." *Expert Systems with Applications* 106 (2018), 36–54. doi: 10.1016/j.eswa.2018.03.058.

Mnassri, Khouloud, Reza Farahbakhsh, and Noel Crespi. *Multilingual Hate Speech Detection Using Semi-Supervised Generative Adversarial Network*, edited by Hocine Cherifi, Luis M. Rocha, Chantal Cherifi, and Murat Donduran, 192–204. Cham: Springer Nature Switzerland, 2024. doi: 10.1007/978-3-031-53503-1_16.

Mosteller, Frederick and David L. Wallace. "Inference in an Authorship Problem: A Comparative Study of Discrimination Methods Applied to the Authorship of the Disputed Federalist Papers." *Journal of the American Statistical Association* 58, no. 302 (1963), 275–309. doi: 10.1080/01621459.1963.10500849.

Muldoon, Connagh, Ahsan Ikram, and Qublai Ali Khan Mirza. "Modern Stylometry: A Review & Experimentation with Machine Learning." In *2021 8th International Conference on Future Internet of Things and Cloud (FiCloud)*, 293–8, 2021. doi: 10.1109/FiCloud49777.2021.00049.

Nangia, Nikita and Samuel R. Bowman. "Human vs Muppet: A Conservative Estimate of Human Performance on the GLUE Benchmark." 2019. doi: 10.48550/arXiv.1905.10425.

Navigli, Roberto, Simone Conia, and Björn Ross. "Biases in Large Language Models: Origins, Inventory, and Discussion." *ACM Journal of Data and Information Quality* 15, no. 2 (2023), 1–21. doi: 10.1145/3597307.

Neal, Tempestt, Kalaivani Sundararajan, Aneez Fatima, Yiming Yan, Yingfei Xiang, and Damon Woodard. "Surveying Stylometry Techniques and Applications." *ACM Computing Surveys* 50, no. 6 (2017), 1–36. doi: 10.1145/3132039.

Niculae, Vlad, Marcos Zampieri, Liviu Dinu, and Alina Maria Ciobanu. *Temporal Text Ranking and Automatic Dating of Texts*, edited by Shuly Wintner, Stefan Riezler, and Sharon Goldwater, 17–21. Gothenburg, Sweden: Association for Computational Linguistics, 2014. doi: 10.3115/v1/E14-4004.

Nordgaard Svendsen, Stefan. "Luke's Readers and Josephus. Paul and Agrippa II as a Test Case." In *Luke's Literary Creativity*, edited by Jesper Tang Nielsen, 266–79. London: loomsbury T&T Clark, 2016.

Nyhan, Julianne and Andrew Flinn. *Computation and the Humanities: Towards an Oral History of Digital Humanities*. Springer Series on Cultural Computing. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-20170-2.

Orellana, Gerardo, Belen Arias, Marcos Orellana, Victor Saquicela, Fernando Baculima, and Nelson Piedra. "A Study on the Impact of Pre-Processing Techniques in Spanish and English Text Classification over Short and Large Text Documents." In *2018 International Conference on Information Systems and Computer Science (INCISCOS)*, 277–83, 2018. doi: 10.1109/INCISCOS.2018.00047.

Pavelec, Daniel, Luiz S. Oliveira, E. Justino, F. D. Nobre Neto, and Leonardo Vidal Batista. "Compression and Stylometry for Author Identification." In *2009 International Joint Conference on Neural Networks*, 2445–50, 2009. doi: 10.1109/IJCNN.2009.5178675.

Pence, Harry E. "What Is Big Data and Why Is It Important?" *Journal of Educational Technology Systems* 43, no. 2 (2014), 159–71. doi: 10.2190/ET.43.2.d.

"Perseus Digital Library." Accessed March 24, 2025. http://www.perseus.tufts.edu/hopper/#:~:text=The%20most%20up%2Dto%2Ddate,and%20sources%20in%20other%20languages.

Pervo, Richard I. and Joseph B. Tyson. "Dating Acts: Between the Evangelists and the Apologists." *Religious Studies Review* 34, no. 4 (2008), 299. doi: 10.1111/j.1748-0922.2008.00324_30.x

Pires, Telmo, Eva Schlinger, and Dan Garrette. "How Multilingual Is Multilingual BERT?." 2019. doi: 10.48550/arXiv.1906.01502.

Raffel, Colin, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer." *Journal of Machine Learning Research* 21, no. 140 (2020), 1–67. http://jmlr.org/papers/v21/20-074.html.

Reddish, Mitchell G. "The Genre of the Book of Revelation." In *The Oxford Handbook of the Book of Revelation*, edited by Craig R. Koester. 19–35. Oxford, UK: Oxford University Press, 2020. doi: 10.1093/oxfordhb/9780190655433.013.1.

Reed, Annette Yoshiko. “Pseudepigraphy, Authorship, and the Reception of ‘The Bible’ In Late Antiquity.” In *The Reception and Interpretation of the Bible in Late Antiquity*, edited by Lorenzo DiTommaso and Lucian Turcescu, 467–90. Boston, MA: Brill, 2008. doi: 10.1163/ej.9789004167155.i-608.116.

Reimers, Nils and Iryna Gurevych. “Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks.” 2019. doi: 10.48550/arXiv.1908.10084.

Ren, Han, Hai Wang, Yajie Zhao, and Yafeng Ren. *Time-Aware Language Modeling for Historical Text Dating*, edited by Houda Bouamor, Juan Pino, and Kalika Bali, 13646–56. Singapore: Association for Computational Linguistics, 2023. doi: 10.18653/v1/2023.findings-emnlp.911.

Repo, Liina, Valtteri Skantsi, Samuel Rönnqvist, Saara Hellström, Miika Oinonen, Anna Salmela, Douglas Biber, Jesse Egbert, Sampo Pyysalo, and Veronika Laippala. “Beyond the English Web: Zero-Shot Cross-Lingual and Lightweight Monolingual Classification of Registers.” 2021. doi: 10.48550/arXiv.2102.07396.

Rogers, Anna, Olga Kovaleva, and Anna Rumshisky. “A Primer in BERTology: What We Know About How BERT Works.” *Transactions of the Association for Computational Linguistics* 8 (2021), 842–66. doi: 10.1162/tacl_a_00349.

Rönnqvist, Samuel, Valtteri Skantsi, Miika Oinonen, and Veronika Laippala. *Multilingual and Zero-Shot Is Closing in on Monolingual Web Register Classification*, edited by Simon Dobnik and Lilja Øvrelid, 157–65. Reykjavik, Iceland (Online): Linköping University Electronic Press, Sweden, 2021. https://aclanthology.org/2021.nodalida-main.16/.

Roy, Ashley and Paul Robertson. “Applying Cosine Similarity to Paul’s Letters: Mathematically Modeling Formal and Stylistic Similarities.” In *New Approaches to Textual and Image Analysis in Early Jewish and Christian Studies*. Boston, MA: Brill, 2022. doi: 10.1163/9789004515116_007.

Ruder, Sebastian, Matthew E. Peters, Swabha Swayamdipta, and Thomas Wolf. *Transfer Learning in Natural Language Processing*, edited by Anoop Sarkar and Michael Strube, 15–8. Minneapolis, Minnesota: Association for Computational Linguistics, 2019. doi: 10.18653/v1/N19-5004.

Sanh, Victor, Lysandre Debut, Julien Chaumond, and Thomas Wolf. “DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter,” 2020. doi: 10.48550/arXiv.1910.01108.

Santoro, Adam, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. “Meta-Learning with Memory-Augmented Neural Networks.” In *Proceedings of The 33rd International Conference on Machine Learning*, 1842–50. PMLR, 2016. https://proceedings.mlr.press/v48/santoro16.html.

Savoy, Jacques. “Authorship of Pauline Epistles Revisited.” *Journal of the Association for Information Science and Technology* 70, no. 10 (2019), 1089–97. doi: 10.1002/asi.24176.

Savoy, Jacques. *Machine Learning Methods for Stylometry: Authorship Attribution and Author Profiling*. 1st ed. 2020 edition. Cham: Springer, 2020.

Schroeder, Caroline T. “The Digital Humanities as Cultural Capital: Implications for Biblical and Religious Studies.” *Journal of Religion, Media and Digital Culture* 5, no. 1 (2016), 21–49. doi: 10.1163/21659214-90000069.

Shah, Foram P. and Vibha Patel. “A Review on Feature Selection and Feature Extraction for Text Classification.” In *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2264–68, 2016. doi: 10.1109/WiSPNET.2016.7566545.

Silva, Kanishka, Ingo Frommholz, Burcu Can, Fred Blain, Raheem Sarwar, and Laura Ugolini. *Forged-GAN-BERT: Authorship Attribution for LLM-Generated Forged Novels*, edited by Neele Falk, Sara Papi, and Mike Zhang, 325–37. St. Julian’s, Malta: Association for Computational Linguistics, 2024. https://aclanthology.org/2024.eacl-srw.26/.

Smith, Daniel Lynwood and Zachary Lundin Kostopoulos. “Biography, History and the Genre of Luke-Acts.” *New Testament Studies* 63, no. 3 (2017), 390–410. doi: 10.1017/S0028688517000091.

Soydaner, Derya. “Attention Mechanism in Neural Networks: Where It Comes and Where It Goes.” *Neural Computing and Applications* 34, no. 16 (2022), 13371–85. doi: 10.1007/s00521-022-07366-3.

Stamatatos, Efstathios. “A Survey of Modern Authorship Attribution Methods.” *Journal of the American Society for Information Science and Technology* 60, no. 3 (2009), 538–56. doi: 10.1002/asi.21001.

Stamou, Constantina. “Stylochronometry: Stylistic Development, Sequence of Composition, and Relative Dating.” *Literary and Linguistic Computing* 23, no. 2 (2008), 181–99. doi: 10.1093/llc/fqm029.

Sula, Chris Alen, and Heather V. Hill. “The Early History of Digital Humanities: An Analysis of Computers and the Humanities (1966–2004) and Literary and Linguistic Computing (1986–2004).” *Digital Scholarship in the Humanities* 34, no. Supplement_1 (2019), i190–206. doi: 10.1093/llc/fqz072.

Sun, Chi, Xipeng Qiu, Yige Xu, and Xuanjing Huang. *How to Fine-Tune BERT for Text Classification?* edited by Maosong Sun, Xuanjing Huang, Heng Ji, Zhiyuan Liu, and Yang Liu, 194–206. Cham: Springer International Publishing, 2019. doi: 10.1007/978-3-030-32381-3_16.

Sun, Qianru, Yaoyao Liu, Zhaozheng Chen, Tat-Seng Chua, and Bernt Schiele. “Meta-Transfer Learning Through Hard Tasks.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, no. 3 (2022), 1443–56. doi: 10.1109/TPAMI.2020.3018506.

Tanvir, Raihan, Md Tanvir Rouf Shawon, Md Humaion Kabir Mehedi, Md Motahar Mahtab, and Annajiat Alim Rasel. *A GAN-BERT Based Approach for Bengali Text Classification with a Few Labeled Examples*, edited by Sigeru Omatu, Rashid Mehmood, Pawel Sitek, Serafino Cicerone, and Sara Rodríguez, 20–30. Cham: Springer International Publishing, 2023. doi: 10.1007/978-3-031-20859-1_3.

Terras, Melissa, Julianne Nyhan, and Edward Vanhoutte. “Selected Definitions from the Day of Digital Humanities: 2009–2012.” In *Defining Digital Humanities*, 279–87. London, UK: Routledge, 2013.

Toner, Gregory and Xiwu Han. “Dating Medieval Texts by Classification with Flexible Time Intervals.” *Digital Scholarship in the Humanities* 35, no. 2 (2020), 459–70. doi: 10.1093/llc/fqz031.

Tyo, Jacob, Bhuwan Dhingra, and Zachary C. Lipton. “On the State of the Art in Authorship Attribution and Authorship Verification,” 2022. doi: 10.48550/ARXIV.2209.06869.

Tzogka, Christina, Fotini Koidaki, Stavros Doropoulos, Ioannis Papastergiou, Efthymios Agrafiotis, Katerina Tiktopoulou, and Stavros Vologiannidis. “OCR Workflow: Facing Printed Texts of Ancient, Medieval and Modern Greek Literature.” 2021. https://www.semanticscholar.org/paper/OCR-Workflow%3A-Facing-Printed-Texts-of-Ancient%2C-and-Tzogka-Koidaki/2fff51e095fde8bcfd618b6a77e5d16de416cd25.

Uysal, Alper Kursat and Serkan Gunal. “The Impact of Preprocessing on Text Classification.” *Information Processing & Management* 50, no. 1 (2014), 104–12. doi: 10.1016/j.ipm.2013.08.006.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. *Attention Is All You Need*, Vol. 30. Curran Associates, Inc., 2017. https://doi.org/10.48550/ARXIV.1706.03762.

Vatri, Alessandro and Barbara McGillivray. “The Diorisis Ancient Greek Corpus.” *Research Data Journal for the Humanities and Social Sciences* 3, no. 1 (2018), 55–65. doi: 10.1163/24523666-01000013.

Wang, Alex, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. “GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding,” 2019. doi: 10.48550/arXiv.1804.07461.

Weidinger, Laura, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng, et al. “Ethical and Social Risks of Harm from Language Models.” 2021. doi: 10.48550/arXiv.2112.04359.

Westin, Fereshta. “Time Period Categorization in Fiction: A Comparative Analysis of Machine Learning Techniques.” *Cataloging & Classification Quarterly* 62, no. 2 (2024), 124–53. doi: 10.1080/01639374.2024.2315548.

White, Benjamin. “The Pauline Tradition.” In *T&T Clark Handbook to the Historical Paul*, edited by Ryan Schellenberg and Heidi Wendt, 1st ed., 39–53. T&T Clark Handbooks, Bloomsbury Publishing, 2022.

Yamshchikov, Ivan P., Alexey Tikhonov, Yorgos Pantis, Charlotte Schubert, and Jürgen Jost. *BERT in Plutarch’s Shadows*, 6071–80, Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022. doi: 10.18653/v1/2022.emnlp-main.407.

Yan, Lixiang, Lele Sha, Linxuan Zhao, Yuheng Li, Roberto Martinez-Maldonado, Guanliang Chen, Xinyu Li, Yueqiao Jin, and Dragan Gašević. “Practical and Ethical Challenges of Large Language Models in Education: A Systematic Scoping Review.” *British Journal of Educational Technology* 55, no. 1 (2024), 90–112. doi: 10.1111/bjet.13370.

Yao, Jerry and Bin Yuan. “Research on the Application and Optimization Strategies of Deep Learning in Large Language Models.” *Journal of Theory and Practice of Engineering Science* 4, no. 5 (2024), 88–94. doi: 10.53469/jtpes.2024.04(05).12.

Yin, Wenpeng. “Meta-Learning for Few-Shot Natural Language Processing: A Survey,” 2020. doi: 10.48550/arXiv.2007.09604.