

Article

A.I. in Religion, A.I. for Religion, A.I. and Religion: Towards a Theory of Religious Studies and Artificial Intelligence

Randall Reed

Department of Philosophy and Religion, Appalachian State University, Boone, NC 28608, USA;
reedrw@appstate.edu

Abstract: Artificial intelligence is increasingly used in a variety of fields and disciplines. Its promise is often seen in a variety of tasks, from playing games to driving cars. In this article, I will sketch a theory that opens the door to the use of artificial intelligence in the study of religion. Focusing on the work of Jonathan Z. Smith, I will show that if, following Smith, the study of religion is considered primarily an act of classification, it can be aided by narrow artificial intelligence that excels at classification and prediction. Then, using a web A.I. called EMMA to classify the New Testament texts as Pauline or non-Pauline as a toy example, I will explore the issues that occur in the application of A.I. Finally, I will turn to Bruno Latour and actor–network theory as a way to theorize the larger issues brought up by the productive use of artificial intelligence in the study of religion

Keywords: artificial intelligence; classification; Jonathan Z. Smith; actor–network theory; method and theory



Citation: Reed, Randall. 2021. A.I. in Religion, A.I. for Religion, A.I. and Religion: Towards a Theory of Religious Studies and Artificial Intelligence. *Religions* 12: 401. <https://doi.org/10.3390/rel12060401>

Academic Editor: Takeshi Kimura

Received: 19 March 2021

Accepted: 26 May 2021

Published: 31 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In this article, I want to begin to sketch a theory that opens the door to the use of artificial intelligence¹ in the study of religion. At some level, it might be thought that this is getting the cart before the horse. With rare exceptions, artificial intelligence is not a technology used in the study of religion. Yet, artificial intelligence is increasingly used in a variety of fields and disciplines. It is, often without our knowledge, encroaching on our lives and impacting everything from what we buy to how we vote (Tegmark 2017; Harari 2017).

However, artificial intelligence also provides an opportunity in the study of religion. For in the same way that artificial intelligence can deal with the enormous amount of data generated by clicks on websites, or cars driving on the street, artificial intelligence likewise has the potential to help understand, in new ways, religious practices, beliefs, and texts. If the history of Religious Studies is any guide, it will only be a matter of time before scholars of religion are employing artificial intelligence in their work.²

This, of course, should not be seen in a negative light. Religious Studies' prolific application of other fields' methods has, on the one hand, led to significant advances in the understanding of our data, and on the other raised concerns that perhaps *this time* we have gone too far. However, the ever-present need for scholars of religion to publish has produced cadres of scholars who have damned the torpedoes and continued full-speed ahead, employing new and novel methods in their research. It is unlikely that artificial intelligence will be the rocks upon which we finally founder.³

When thinking about artificial intelligence, two mutually congruent avenues present themselves: On the one hand, scholars may examine the way artificial intelligence is used in religion and the religious implications of the use of A.I. in society. For instance, as we find ourselves increasingly surveilled to the point where Bentham's panopticon seems an almost laughable forerunner in its limitations, we might want to think about the role religion plays in public acceptance or resistance. After all, Christianity made the all-seeing deity a doctrine of faith. God's ever-watchful eye is always on the world, seeing the

sparrow fall, the grass grow and every secret transgression being committed. The church has long leveraged this religious panopticon in hopes of inspiring good behavior. However, we might rightly ask when A.I. can do the same, will the faithful be inured to it, having transferred the omnipresence/omniscience of God to Google? Or will they resist, seeing this as an incursion on divine territory? Such questions I believe will long occupy our scholarly work as artificial intelligence becomes more commonplace.

On the other hand, the second avenue is to think about how scholars of religion can conduct experiments using artificial intelligence and machine learning in an attempt to better understand religion and conversely to use religion to better understand artificial intelligence. Wesley Wildman and collaborators (Shults et al. 2017; Saikou et al. 2019; Shults et al. 2018a, 2018b; Gore et al. 2018) have already begun such experiments attempting to use simulation software to examine trends and theories in religious studies. Their work has already generated insights into the way religion works.

Yet, such a move requires a theorization of the task. In what ways can artificial intelligence be effectively used in the study of religion? This paper then seeks to explore this theorization as well as potential problems in its implementation. My task here is to construct a theoretical justification for the use of A.I. in religious studies. To that end, I take seriously the ideas of J.Z. Smith, whose work on the methodology of religious studies focused on classification and redescription and will apply these ideas to an understanding of how artificial intelligence may be understood as part of the methodological pantheon of religious studies.

However, I also want to point out some of the problems that are inherent in the usage of artificial intelligence; specifically, I will discuss the nature of the “black box” problem that artificial intelligence has, and the reality that while we can see what artificial intelligence produces, we often can only understand why A.I. generates what it does indirectly. To that end, I use as an example EMMA A.I. EMMA is a plagiarism detector that claims to use artificial intelligence as part of its system. It is a good example of the problems I want to engage for two reasons: (1) it is proprietary so the user does not have access to its model, (2) its only output is a conclusion and confidence level. For my example, my research group and I made use of EMMA because its simple web interface might be attractive to non-technical humanities researchers, and because it highlights a problem that would likewise be the case in more technical implementations of the same task, the issue of explainability.

With the problem of explainability in mind, I then return to a theoretical discussion. What EMMA exemplifies is the different kinds of classification and evaluation that an A.I. engages in when analyzing problems in religious studies—a process different from humans. To theorize this, I turn to Bruno Latour whose actor–network theory recognizes both the independence and the interdependence of the apparatus of analysis coupled with the human actor. With artificial intelligence, this relationship becomes even more pronounced.

However, to be clear, this paper does not attempt to forward an application of using artificial intelligence in religious studies. The example I use is more a toy example, useful for pointing out the problems inherent in the use of artificial intelligence, but certainly more sophisticated models and analysis would be necessary to contribute to the problem of authorship that I have focused upon. While such a project is commendable, it is not the focus of this paper. Thus the practitioner of artificial intelligence may be dismayed at the elementary results that EMMA produces, the lack of clarity about its model, and the failure to produce accepted evaluation metrics like f1, perplexity or accuracy scores. However, for my project, this is a feature, not a bug. For as I will make clearer below, EMMA offers us a chance to think about A.I. without getting bogged down in the technical details. That EMMA fails at her task is irrelevant; ultimately it is the processes and constraints that are a part of the use of artificial intelligence that I am interested in, not the results.⁴

In the meantime, while the implementation of artificial intelligence in the study of religion seems like a question of when, not if, it is worthwhile to think carefully about what

A.I. might do for the study of religion and to think about the theoretical implications of doing this.

2. Artificial Intelligence—What It Is

What are we talking about when we talk about artificial intelligence? This is more confusing than it might appear at the outset, because A.I. has not been solely the purview of computer science departments and tech companies. The discussion around artificial intelligence has already been shaped by a myriad of Hollywood films, from Hal in *2001: A Space Odyssey* to *The Terminator* and the *Matrix* movies, to the BBC series “Humans” and HBO’s “Westworld”. When most people think about artificial intelligence they think about robots, world-dominating or otherwise. With every article written about artificial intelligence in the popular press, the question that usually arises somewhere in the article is, “is this beginning of the robot uprising?” to which the answer is invariably “no” (Geraci 2012).

However, the fact that A.I. has a presence in our cultural imagination means that our discussion about it is already overdetermined. We ask the existential question, and having heard that humanity as a species can rest easy in its survival, we fail to ask more important and perhaps more disturbing questions that may not have such easy reassuring answers. In discussions about A.I., it is common to divide A.I. into two categories: narrow A.I. (ANI) and general A.I. (AGI). The *Matrix*, the *Terminator*, etc. are all examples of AGI. Artificial general intelligence is a machine that can think as we do, that can learn from new experiences, and can form conclusions on its own. Often it is defined as a machine intelligence that is able to function effectively in multiple domains. An AGI, like us, could drive a car, have a conversation, play chess, and read and produce a summary of a paper. Perhaps not all at the same time, but importantly AGI is not limited to a single domain.

We currently do not have an AGI, and A.I. researchers do not have clarity regarding when or if we will. To be sure, there are computer scientists who are working on the problem (Goertzel 2016), yet there is currently no consensus. For instance, is it possible that AGI requires an actual body to develop? Suzanne Gildert, founder of Sanctuary A.I. (a robotics company), believes bodies are a necessary precondition to intelligence (Creative Destruction Lab 2017). Likewise, is the route to AGI found through one master learning algorithm, as Open A.I. and Google’s Deep Mind seem to infer, or is to be found through a series of decentralized but connected algorithms as Ben Goertzel has argued (Goertzel et al. 2014)? We do not know, and at this point, we are far enough from the goal that such talk is mostly speculative, though work along each of those lines continues, including important discussions about safety after the development of AGI (Bostrom 2016; Russell 2019).

What is present now is artificial narrow intelligence (ANI). Narrow artificial intelligence is often talked about as deep learning and sometimes as machine learning (although these are not the same thing, the difference need not concern us at this point). In broad strokes, ANI is very good in terms of three quite different tasks: classification, prediction, and generation (Mitchell 2020). There is a fourth kind of ANI, reinforcement learning, which attempts to learn through trial and error by maximizing a reward signal.⁵ For the purposes of this discussion, I will not engage reinforcement learning; however, it is possible that in simulation studies like those done by Wesley Wildman and LeRon Shultz, reinforcement learning will play an important role in future religious studies research.

I will be focusing on the first two kinds of tasks that ANI can do (classification and prediction). The processes of creating an A.I. task of classification or prediction begins with a dataset divided into training and testing data. The model first learns the commonalities of the training dataset. The model is then applied to the testing dataset. After these steps, the model either classifies or predicts based on the type of model. For instance, in a dataset of cats, given any picture, the model should be able to classify whether the picture is a cat or not. In terms of prediction, given a certain input, the model should be able to predict the next element or event. For instance, a predictive type model will, given a certain set of words, be able to predict what the next word should be. Google’s predictive type in email

is an example of this. I type “Let us try to get” and it will suggest “together this week.” In response to a student email that says, “I was in a car accident this weekend, can I take the midterm next Thursday,” Google will predict (and suggest) a response of “You can do that”.

The third task, which I include here only for completeness, is that of generation. Generation likewise starts with a dataset to learn the characteristics of data. The A.I. then produces another example of what it sees in the dataset. Often, this is coupled with what is termed “a discriminative network”, which then classifies the output as either similar to the data in the dataset or not. Thus far, generative networks have not shown application to Religious Studies.

In all these cases, data are fed back to the model indicating success or failure, and the model accordingly adjusts its parameters. Thus the A.I. learns from each attempt and ultimately improves over the number of times it is run. It is this component that is the “learning” in “deep learning”.

It is perhaps important to talk about the kind of data that are used in ANI. Generally, the kind of data can be divided into two distinct categories: image/video and text. In Religious Studies, we are generally text-based; whether we are talking about primary and secondary source material or interviews and case studies, ultimately the type of data we are engaging is textual data. This, however, may be a limitation in terms of the kinds of data we privilege, rather than the demand of our methods. Insofar as we are dealing with the distant past, our access is generally limited to texts; certainly, this is true for Biblical Studies as well as textual studies in other religions and traditions. However, this may betray the protestant bias in our field, where the notion of “sola scriptura” has focused us on textual representations.

For those scholars who are studying more contemporary expressions of religion, it may very well be that textual representations fail to capture much of what is actually happening in religion. Perhaps for these scholars, religious behaviors would be better recorded with video and images. Take for instance the commonplace happening of Catholic parishioners coming forward to take communion. The variety of expressions, from prayerful, to joyful, to rote, to bored, might be captured in their fullness through video in more precise detail than it would be through a set of field notes taken by the scholar. While clearly the camera may not capture everything, it may give the scholar access to something very different and expansive beyond textual representations.

The future of using this sort of visual data then presents a different kind of problem than that of text: how do we analyze it? Does the quantity of data that one might acquire from simply recording a normal church service exceed the ability of the single scholar to detect patterns and deviations? It is here where scholars may seek to turn to artificial intelligence to aid in data analysis. Given the enormous quantity of data, A.I. does well in tasks of classification.

However, of course, this is not to exclude textual data in lieu of visual data when it comes to the use of A.I. in the study of religion. Certainly, artificial intelligence has been deployed in the analysis and classification of text, which may be a boon to the textual study of religion. The application of A.I. to text is called “Natural Language Processing” (NLP). NLP is used in a variety of applications from topic modeling to text generation. Additionally, yet again, given the large quantities of data that we find in the textual tradition of religions, both contemporary and historical, artificial intelligence may be able to produce classificatory results that are intriguing.

Additionally, several traditional questions in textual studies might be given new purchase by the application of NLP techniques. Questions like the authenticity of sayings in the quest for the historical Jesus, the authorship of Pauline letters, or analysis of Christian versus Greco-roman literature might well find new answers with NLP approaches.⁶ Similar kinds of insights might be found when applied to literature in a variety of historical moments or when tracing cultural and religious trends. Some early work has been undertaken in this regard using Google Books (Finke and McClure 2017), and certainly, with more

intelligent machine agents, we might well expect continuing progress. The result may be new data and new analyses in the study of religion.

Just as intriguing is the possibility of insights into artificial intelligence that might be generated by its application to Religious Studies. What we must recognize is that artificial intelligence does not think in the same way as humans do. While there is a rough analogy between neural networks and the human brain, such comparisons are more evocative than accurate. Thus, there is particular use in giving an A.I. a Religious Studies problem and then looking at how it might solve that problem. We have seen examples in the way that A.I. has often defied conventions and demonstrated ingenuity in playing games like Go ([AlphaGo: The Story So Far n.d.](#)), might we not see this in other domains as well like religious studies?

Artificial intelligence, then, may have a positive impact on the study of Religion. However, apart from whatever practical results we might see now or in the future, I want to take a moment to reflect on the theoretical concerns that might lie behind the use of artificial intelligence in Religious Studies.

3. J.Z. Smith and the Study of Religion

To begin, I want to first address the question of whether narrow artificial intelligence really has any application in Religious Studies. After all, scholars often wax poetic about the complexity of religion. If ANI is limited to classification and prediction, does it really have use in religious studies? I want to argue that classification is, in fact, essential to the work of religious studies. To support this position I want to turn to the work of Jonathan Z. Smith. There is no other scholar who is as respected when it comes to the study of religion as J.Z. Smith. For the more than four decades of his career, Smith was continually interested in the question of what the study of religion is. His concern about method and theory in the field launched a thousand articles, as scholars took up his call to think seriously about the field.

One of Smith's most explicit expositions of his thinking about the method in religious studies is found in his work *Drudgery Divine* ([Smith 1994](#)). Here, he takes on the issue of comparison. For Smith, comparison is the essence of the study of religion, particularly in his area of the history of religions. Categories such as "unique" and the more theological "wholly other" ultimately act as a way of insulating religion from critical inquiry. He notes that, instead of those categories,

What is required is the development of a discourse of 'difference', a complex term which invites negotiation, classification, and comparison and at the same time avoids too easy a discourse of the same. ([Smith 1994](#), p. 42)

What we might note here is that "discourse of 'difference'" has three components, "negotiation", "classification" and "comparison". The notion of negotiation in Smith remains unexplored, though he seems to be focused on the question of the elements of comparison. In a quote from another of his works, Smith argues "comparison requires the acceptance of difference as the grounds of its being interesting and a methodical manipulation of the difference to achieve some stated cognitive end" ([Smith 1992](#), p. 14). Negotiation thus has to do with "manipulation", which, for Smith, involves maintaining some differences and then also "defensibly relax[ing] and relativiz[ing]" other differences ([Smith 1992](#), p. 14). Inevitably, this requires an answer to the question of the selection of what is being compared; why are those things necessarily the object of comparison rather than other things?

In *To Take Place*, Smith further develops the category of classification. Without doubt, comparison as a project is founded upon classification. Thus Smith takes the process of classification seriously. While it would not be appropriate to say that Smith elucidates a theory of classification per se in this work, the issue of classification does concern a great deal of his discussion in the second chapter. There he is particularly interested in the notion of classification as employed by Durkheim and Mauss (E. [Durkheim and Mauss 1967](#)). The ultimate question seems to be, for Durkheim and Mauss, one of origins with regard to the classificatory systems of aboriginal tribesmen. It is here that Durkheim and Mauss

ultimately ask the question: where does the process of classification originate? The answer, for them, is that there is a base duality that lies at the root of classification, and it has a basis in the social order. Ultimately Durkheim continues this line of thinking in *Elementary Forms of Religion* (É. Durkheim 2008).

Yet, Smith is not really interested in this question of origins, as he uses his discussion of classification to highlight the notion of “gap”. Smith argues that there is always a distance between the reality of a thing and the ritual/belief about it. With this notion of “gap”, he unmasks the ideological nature of classification, that classification is never value-free. Rather, classifications “are hierarchically ranked in relations of superordination and subordination with radically different valences.” (Smith 1992, p. 41). Classification then is a human project that sometimes subtly, other times overtly, inscribes the social process into the very parameters of a human understanding of reality. Thus, correspondingly, classification is, for Smith, a key component of Religion and Religious Studies as well. Classification in the hands of the scholar of religion becomes the essential tool in both the creation of our data and our understanding of that data.

Still, at this point, I might supplement this with a further discussion of Smith’s development of modes of classification in *Drudgery Divine*. In a footnote on pages 47–48, Smith discusses the difference in biology between homology and analogy. For our purposes, the full distinction is of less importance; what is more pressing is that this relates to a scientific notion of classification. The question at stake is how one classifies a given biological structure. Smith concludes his argument,

There is one paramount issue which appears to distinguish the biological enterprise of phylogenetic comparison from those within the human sciences. The phylogeneticist strives, at least in theory, for a ‘natural’ classification; the human scientist must always propose an ‘artificial’ classification. This is because there is, arguably, nothing comparable to genetic processes within the sphere of human culture. (Smith 1994, pp. 47–48)

This ties classification into his conclusion about comparison in general. The result is a principle, now widespread in Religious Studies, that Smith would become known as the originator of: comparison is an act of the scholar, and that act constructs the ultimate result.

⁷ Thus,

There is nothing ‘natural’ about the enterprise of comparison. Similarity and difference are not ‘given’. They are the result of mental operations. In this sense, all comparisons are properly analogical . . . (Smith 1994, p. 51)

I believe it is not taking Smith too far to suggest that what is true about comparison is likewise true about classification in general; that both are ultimately about a combination of interests and operations whereby we make distinctions between objects or collections of objects of study.

Smith engages in a further sustained reflection on the notion of classification in a 2000 article in Braun and McCutcheon’s *Guide to the Study of Religion* (Smith 2000). Here, he reiterates some of the points explicitly that were implicit in his earlier work. He highlights that from the very start, anthropology (Lévi-Strauss) and Enlightenment philosophy (Hobbes) see an “interest in taxonomy and the necessity of classificatory projects as characteristic of either science and cognition” (Smith 2000, p. 36). However, note the subtle shift that occurs. Now classification is an inherent part of the intellectual process, one that is carried through from biology to the philosophy of Wittgenstein and Austin in the notions of “family resemblances” that have been picked up by religion scholars. At the same time, Smith cites an interest in classification coming from cognitive science in Religion Studies, noting “cognitive studies of classification have begun to be influential in some recent theoretical works on religion” (Smith 2000, p. 37). Classification has thus shifted from a tool employed by and in the study of religion, to something at the root of intellect and brain function itself.

Smith also here expands on his work on classification. Smith argues that classification is polythetic rather than monothetic, meaning that the classification of a given item into a class need not be based on a singular component or some element(s) that all components of the class have in common, but rather may be based on a collection of components that may be only partially found in another element of the class. This brings him to the idea of prototype theories whereby “the match, in a polythetic fashion, does not have to be exact; it requires a judgment that the object (or concept, or word) is sufficiently similar, a judgment not made primarily on the basis of visual appearance.” (Smith 2000, p. 38).

Smith then goes on to argue, in the detailed fashion for which he is famous, that classification both by and about religions has ultimately played into theological concerns and apologetic agendas. In contrast to the scholarly failure to classify religions without engaging in essentialist and colonialist agendas, Smith commends the more limited taxonomies of church, sect, and cult studies or ritual studies and folklore classifications. Yet, these limited successes and grand failures are not an argument to abandon classification or to resort to merely repeating indigenous categories⁸, but instead calls for “rectification”. It is incumbent upon the scholar to work through the subtle issues of colonialism and theology that have permeated the use of classification and find a way to employ classification free of those traps. However, the tool of classification must, for Smith, be preserved. For as Smith forcefully concludes, “For the rejection of classificatory interest is, at the same time, a rejection of thought” (Smith 2000, p. 43).

Thus for Smith the study of religion is intricately tied up with the process of classification. In fact, it would not be hyperbole to suggest that for Smith the study of religion requires classification, and at some level never exceeds it. Smith is often well known for his call for “re-description” as the ultimate task of religious studies. This process is one of changing classification schemas. Thus for Smith, the very act of studying religion involves the classificatory endeavor. Or as Smith phrases it, “scholarly labor is a disciplined exaggeration in the direction of knowledge; taxonomy is a valuable tool in achieving that necessary distortion” (Smith 2004, p. 174). It is classifications all the way down.

If it is correct that classification is, in fact, an essential foundation of the study of religion, then the use of artificial intelligence becomes clear. As I noted above, one of the primary uses of artificial intelligence is the act of classification. A.I., once properly trained, can classify an enormous amount of material. Additionally, such classification is almost always polythetic in Smith’s sense of the word. However, of course, the rub is contained in the phrase “once properly trained.” For particularly with supervised data (data in which the classificatory schema is given by the scholar), it is certainly the case that the same sort of biases that Smith worries about, either implicit or explicit, may very well color the classification of future data. The incident of Google photos identifying photos of people of color as “gorillas” is just one example of this problem.⁹

Thus, the principle of classification, the formation of taxonomic categories, ultimately remains the purview of the scholar, not the machine. What the machine can do is apply the same categories to unreviewed data. This may seem like a time-saving effect, but one may fear that it is not really adding to the creation of scholarly knowledge. However, A.I. also might act as a kind of check on the integrity of the scholarly process. Take, for instance, the authentic Pauline texts vs. the Pastoral epistles. While most scholars are convinced for a variety of reasons that the Pastoral epistles are pseudonymous, it would be interesting to see if a neural network trained on Pauline literature and free of theological agenda might come to the same conclusion, and its reasoning in the process. Yet, in the very designation of this problem, there are some issues that beg to be explored, both in terms of the possibilities and limits of A.I. as well as in its application.

4. Some Experiments with EMMA

As an example of both the problems and potential of A.I., I and my undergraduate research team sought to do some experiments using an author identification artificial intelligence called EMMA at emmaidentity.com. EMMA is an A.I.¹⁰ designed for plagiarism

detection. The program has a web interface and texts are uploaded, the user names the author (this becomes the training set), and then another unidentified text is provided and EMMA attempts to determine if the author is the same.

Our use of EMMA was not an attempt to find new insights regarding the authorship of the Pauline corpus (both accepted and disputed texts); rather, our task here was to attempt to understand how an A.I. might evaluate the problem. Given a clearly identified corpus (the accepted Pauline epistles), how would EMMA make a determination regarding the disputed epistles and other biblical texts? Our goal here was to try to trace the reasoning behind the program's decisions and to understand what might influence and bias its decisions. Thus, if the reader feels that Biblical Studies has not really been forwarded by this experiment, they should understand that such was not the intention. Instead, insofar as was possible, we sought to understand the reasoning the A.I. used and expose problems and benefits therein.

To begin, we gave EMMA the NIV translation of Romans (identified as Pauline) and 1 Timothy (unidentified), and it concluded that it was "100 percent sure that this text was written by Paul".¹¹ I also gave it some of my own writings, including a couple of articles, and a blog post I had written, and then some other samples not written by me. EMMA had a 3/5 accuracy rate for my writings.¹²

Regardless of EMMA's failure to re-establish the conclusions of 250 years of Biblical Studies, what would be more helpful is if EMMA could explain why it thought 1 Timothy was written by Paul. However, while EMMA was 100% certain that Paul was the author, it could not give any more information. A presentation on Emma's development company website stated that EMMA uses "more than 50 math parameters to define writing identity" and is 85% accurate, resulting in the "highest result in authorship identification ever achieved by engineers worldwide" ([EMMA. Defining Writing Identity. Disrupting Plagiarism n.d.](#)). Yet, the product of Emma's work is merely a "yes" or "no", and often a "certainty" percentage. Still, even when it produced the number, the reasoning behind the conclusion necessarily remains opaque.

Thus, we may want more explanation as to how EMMA determines authorial authenticity. One of the explanations we might think we could safely dispense with is that EMMA does not classify 1 Timothy as Pauline because it is in the canon. Additionally, yet because EMMA works with English texts, it is dependent on translation and translators. These translators do know that many believe that Paul wrote 1 Timothy as well as Romans. Could it be that these translators tended to downplay linguistic differences in their translations?

My undergraduate research group decided to do a series of experiments to answer this question. First, while it seems unlikely, we wondered if the fact that both texts contain a claim that they were each written by Paul was influencing the algorithm (since this is certainly the case for many conservatives who argue for Pauline authorship of both texts). EMMA's algorithm is not public, so we do not know whether a claim of authorship is a factor or not. Thus we deleted the first verse in both books and re-ran the comparison (here we used the New International Version). The results were identical—EMMA claimed 100% certainty that Paul had written both documents.

Second, we wanted to try a series of other translations and see what results EMMA produced. Significant deviation might point to a translator eliding stylistic differences in their translations between Paul and the pseudonymous letters. We fed EMMA a variety of translations (New American Standard Bible, Tree of Life Version, New Revised Standard, Young's Literal Translation, and the New King James Version); we started with either Romans or 1 Corinthians, and then fed it a series of Pauline Letters and then disputed Pauline letters.¹³

The results were quite consistent across versions: When given the other authentic Pauline letters Romans, Galatians 1 and 2 and Corinthians, EMMA reported each as having been written by the same author. On the other hand, EMMA uniformly rejected Philemon, as well as Philippians (both of which are generally held to be authentically Pauline). 1 Thessalonians was either scored low (around 40% or lower probability for the NASB, YLT,

NKJV, and NRSV) or rejected altogether (NIV). Thus EMMA portrayed a skepticism that Biblical Scholars do not about some of the Pauline corpus.

As EMMA is ultimately a black box and her programming is proprietary rather than open-source, the various models and programs that EMMA used were inaccessible for inspection by my team.¹⁴ Thus, my research team tried to determine if it could deduce why EMMA made the decisions it did. The team tried varying the initial Pauline book that was fed to EMMA, which did seem to make a difference. If 1 Corinthians (NRSV) was the first document seen by EMMA, she was far more skeptical of Titus (NRSV), rating it 35% likely to have been written by Paul. However, that probability nearly doubles when Romans is the first book EMMA sees, to 69%. We hypothesized that perhaps style similarities between Romans and Titus explained the difference. In the same way, the rejection of Philemon across the board would likewise point to style being predominant, as Philemon is the only letter to an individual in the authentic Pauline corpus.

On the other hand, we concluded that while style seemed to be a factor, EMMA was not just captured by “Bible” language. Genesis and Revelation were rejected as non-Pauline across the board, though (what would probably be to Paul’s ultimate consternation) EMMA was less sanguine about Matthew, occasionally suggesting a probability of common authorship of just over 50% in several cases.

However, my research team did find that length may have been overly valued. When the input documents were made uniform lengths (though cutting or combining), EMMA’s certainty increased to deem even Philemon (ever reviled by EMMA as inauthentic) as probably authentic. Likewise, when the training case was longer (Romans has 9494 words) and the test cases were all shortened to 200 words (the minimum EMMA would take), EMMA uniformly rejected all Pauline letters with the exception of Galatians (54% probability). Thus, it appears that EMMA may have put undue emphasis on particular proxies for authorship, such as length and formal vs. informal style.¹⁵

The results of this work reveal that while algorithms may avoid some of the classificatory problems that have beset Religious Studies scholars in the past (for instance, the algorithm itself holds no theological standpoint that it is eager to justify), an A.I. may lack the ability to deal with composite reasoning, and often substitutes unreliable proxies that we can discern only indirectly. Algorithmic approaches in other fields have shown a similar result. Particularly in the areas of criminal justice (O’Neil 2016) and gender-biased hiring (Thompson 2019), algorithms that are trained on biased data have been seen to replicate that bias. As artificial intelligence enters into the toolbox of the Religious Studies scholars, we must be alert to not only our particular field’s biases, but also the concerns that have been seen in other domains.

5. Actor–Network Theory

If classification is an essential part of both artificial intelligence and Religious Studies, then it follows that artificial intelligence can function as a tool in this regard. However, the experiment above suggests that artificial intelligence has issues that scholars of religion need to be aware of. What our experiment has shown is that as a tool, A.I. functions independently and opaquely. Many of the criteria by which EMMA made her classification decision were not clear, and experiments designed to determine the parameters of these decisions often showed a line of reasoning very different to what a human might employ. The benefits of this is that, ideally, subjective ideals such as theological convictions and a history of colonialism are not present, or present to a lesser degree. However, the downside is that often not clear what has replaced these ideas. Thus, A.I. as a tool requires further consideration, particularly as we contemplate the integration of the two fields of religious studies and computer science. Along these lines, it may be fruitful to look at the theoretical insights of actor–network theory (ANT). Rising out of science and technology studies (a subfield of sociology), ANT is a form of analysis concerning the study of society in general, and more specifically technology’s role in it. ANT’s theoretical contribution to this project is that it is an analysis that appraises the role of technology as something more than just

a passive extension of human action, as an actor (or “actant”) itself. When we consider artificial intelligence, we reach a new level of tool, a tool that unlike many other tools works both independently and in concert with the human user. To this end, artificial intelligence is thus the final expression of what is anticipated in actor–network theory. I want to return to our previous discussion of classification in general and suggest that an integration with ANT may yield fruitful insights.

The primary goal of actor–network theory is to examine the network of connections of actants that makes the “structure” of any system possible. ANT focuses on “associations”, meaning the connections between the various actants in a network. ANT has some immediate application to our concern here, because an actant does not have to be human; indeed, Latour sees this as an inherent benefit of ANT, in that it takes seriously the non-human actant. Actants simply do things and are most often *made* to do things. Actants may likewise make others do things. Each actant thus is located in the chain of an action network. As Latour says, “action is dislocated”; it may involve, but does not require, a human actor (Latour 2007, p. 46).

Instead of focusing on actors, one might spend more time profitably talking about mediators and intermediaries. Intermediaries are largely transparent, the “inputs predict output fairly well” (Latour 2007, p. 58). Mediators, on the other hand, make changes to the input before producing the output. There is a certain level of unpredictability, particularly if mediators invoke/create other mediators. A simple illustration of this is the keyboard and word processor I am using to write this article. The keyboard is an intermediary; each letter is tapped on the keyboard and reproduced on the screen. The word processor, however, is a mediator. It constrains what I can do (hand drawings are out), it silently formats my text, it corrects my spelling and grammar. It occasionally substitutes one word for another, often without my permission. The input and output are not necessarily matched; ideally, the mediator improves the output, though this is not guaranteed, but it certainly does change it.

Latour maintains that most objects start as mediators but soon become “invisible, asocial intermediaries” (Latour 2007, p. 80). However, Latour suggests that when objects create innovations, they “live a clearly multiple and complex life.” My keyboard was not always an intermediary. When I first took typing class in 10th grade, it was a strange illogical creature that most often failed to repeat my intentions. However, this changed with practice, and it is when objects fade into the background (like my keyboard) that they become intermediaries. Latour maintains that in certain circumstances, objects may return to the forefront, particularly when they cease to work as expected (anyone with a sticking key on their keyboard can attest to this). I would add that when objects add something new through upgrade or revision they may also return to mediator status.

As mediators, an object can constrain and/or impel action. Latour discusses the NRA’s slogan, “guns do not kill people, people kill people” (Latour 1999). For Latour, this fails to understand the mediating nature of the object. The person with the gun may in fact have wished their intended victim harm before picking up the gun, and yet the gun itself, through its form (its portability and potential for concealment) and its mechanical nature (its ability to deliver lethal force) is more than just a passive intermediary for the perpetrator’s desire. The gun mediates that desire, the gun forms the kind of result that can be accomplished. A different weapon, a knife or baseball bat, might also deliver injury and death, and yet the type of injury, the type of defense, the type of result is different. My baseball bat-wielding assailant needs a lot more room to swing the bat around, while my knife-wielding assassin needs much greater proximity but less room to inflict maximum harm. Move into a narrow hallway and the bat wielder struggles. Separate me from the knife-wielder with a ditch and their attack is stymied or reduced to the accuracy of their throw. Each of these weapons changes the assailant; it facilitates and limits their choices. We use different language regarding the combination of assailant and weapon that indicates this difference: shooter vs. knife-wielder. Returning to the shooter, however, none of the defensive measures I have outlined above for the other weapons are effective

against the gun. The gun, then, is not innocent. It too is an actant that causes the assailant to act in particular ways when combined with the assailant actant. As Latour says, “When the propositions are articulated, they join into a new proposition. They become ‘someone, something’ else” (Latour 1999, p. 180).

The importance of ANT for the question of Religious Studies and A.I. is that it allows us to focus on the non-human. As we look at artificial intelligence in its narrow form, ANT provides the categories to treat this as an actant without it having to rise to the level of general A.I. Narrow A.I. as an actant acts on both its users and its programmers. Its work of classification, prediction, and generation forbids certain kinds of work and compels others.

To return then to our experiments with EMMA. EMMA is unable to replicate the results of Biblical criticism because, as I mentioned before, the designation of authentic Pauline literature is dependent on a set of composite reasons. Style and vocabulary play an important role in the argument, but so do inferred communal structures (which differ between the Pastorals and the authentic letters) as well as anachronistic statements (such as the reference to the demise of the Temple in 2 Thessalonians). How can what is essentially a language classification model include such reasoning or discover it? Additionally, if this is not possible (or at least not possible for EMMA), then might it be the tool itself that seeks to mediate knowledge that occludes, rather than facilitates, discovery?

To return to Smith’s work on classification, he expressed concern that in the process of classification, what was inscribed as the classificatory parameters was the distinction between Christianity and all other religions. This distinction then served as a proxy for true and false religion (Smith 2000). Smith never asked the question of whether the tool was the problem, whether classification itself might not precipitate this issue, choosing instead to see the mistake as dualistic classification. I do not want to second guess him on this point. Rather, I would suggest that this brings together precisely the kind of project suited to ANT. What Smith uncovers is a series of actants, some of which are human (like individual scholars he names), but others that are non-human, like the Christian faith shared by these scholars, the imperialist ideals and structures of the countries most were from and, not to be forgotten, the method of classification itself. These actants came together to form a network that mediated the process of classification into a system that ultimately reinscribed a particular theological agenda.

To see artificial intelligence as an actant thus provides us with an opportunity to examine, at a more granular level, the kinds of challenges that the object may present. It is not that artificial intelligence is necessarily exceptional in this; Latour and his collaborators offer a myriad of examples of this sort of thing in a variety of fields. What ANT shows us is that this is the case of all networks.

Yet, artificial intelligence may differ in two ways: first, in the rapidity with which it has invaded so much of our daily lives (often unbeknownst to the average person), and second, in the way it masks its own decision-making. Safiya Noble and Cathy O’Neil (Noble 2018; O’Neil 2016) have both pointed out the way that the black box of artificial intelligence functions to insulate institutions of power (from Google to U.S. News’ College Rankings) from being called to accountability. The proprietary nature of the algorithms used by these companies along with the inherent obscurity of the algorithms themselves can perpetuate already existent networks of discrimination. We saw the same results with EMMA. The decisions EMMA made were shrouded in its proprietary nature, and we could only indirectly determine why EMMA made the classificatory choices she made. Additionally, while in this case we were not able to determine the decisions were motivated by an implicit theological bias, there is no reason to suspect that, given a dataset that was formed by such bias, it would not be replicated by the A.I.

There are moves away from this opacity towards a greater openness. In 2019, Microsoft open-sourced a library hoping to improve machine learning intelligibility (Glass Is the New Black—Intelligible A.I. Is within Your Grasp 2019). Likewise, in August of the same year, IBM introduced the A.I. Explainability 360 Open Source Toolkit (A.I. Explainability 360 n.d.), and shortly thereafter in October, Facebook introduced Captum, its model inter-

pretability library for the popular PyTorch framework ([Open-Sourcing Captum: A Model Interpretability Library for PyTorch n.d.](#)). What the development of all the frameworks that attempt to create a way of explaining the actions of A.I.s indicates is a recognition of the danger inexplicable A.I. models present. Moreover, non-neural network-based A.I., such as symbolic reasoning (sometimes called “good old fashion A.I.” ([Marcus and Davis 2019](#))) and multi-agent artificial intelligence (J. E. [Lane 2013](#)), do not have the explainability problems of deep learning approaches, and yet these approaches have narrower application than deep learning and are not as widely used.

Additionally, what ANT urges us to do is to consider the whole network, not just pieces of the model. The network of corporations, programmers, models, computer equipment, results, publication (or lack thereof), and implementation all work together as a combination of actants that produce the oblique models that threaten to classify human recipients, often to their detriment as credit risks, recidivist risks, poor students, etc., without the possibility of appeal. The answer may lie in the addition of further actants, as the frameworks listed above seek to do, but it may also require a different network that includes different actants that are founded upon openness rather than profit. It is by engaging in a careful analysis of those networks that we can see the problems and seek solutions.

6. Conclusions

The task for Religious Studies scholars who want to use artificial intelligence as part of their work is two-fold. First, we cannot lose sight of the limits of A.I. as I briefly mentioned above. A.I.s cannot deal with facts and logical inferences ([Marcus and Davis 2019](#)). Perhaps in the future, this will not be a limitation of A.I.; however, as of now, reasoning is beyond the abilities of even the most advanced models. The composite nature of the reasoning when dealing with the authorship of 2 Thessalonians offers an example. Furthermore, it is not impossible to substitute classification for reasoning in some instances. For example, with 2 Thessalonians, one might start with a classification model of a large segment of texts classified as written before 70 and after 70. After that, this kind of chronological classification algorithm might then be combined with one that examines the possible Pauline texts stylistically as EMMA did. This two-step process could act as a way of classifying the data using multiple models, thus putting 2 Thessalonians outside the probable time of Paul even if the text was not initially excluded based on stylistic issues. The combination of multiple models towards an end (sometimes called a “pipeline”) like this is quite common in artificial intelligence, with results that increasingly reach or surpass human levels ([Gerrish and Scott 2018](#)). Additionally, such a project is not reasoning, but a proxy for it. It is to return to narrow A.I.’s strength, which is classification. Additionally, as I have argued above for the scholar of religion, this is, in fact, a key part of their task. A.I. may offer us new applications of this pivotal method.

Second, it is necessary to understand what contribution artificial intelligence can make to the questions at hand. In reality, an artificial intelligence experiment may have two different results. On the one hand, A.I. may help the scholar create new insights into the object of religious beliefs, texts, or practices. In so far as this is the case, though, the limitations of artificial intelligence quickly become clear. ANT provides the tools to map the networks that are a part of such work, and can reveal precisely some of those limitations. However, with this caveat, A.I. can be an important aid in classification and prediction, and any task that includes these might well benefit from the implementation of A.I. as A.I. has the ability to deal with enormous amounts of material and consider incredibly large numbers (millions and billions) of variables.

On the other hand, another possible contribution of A.I. experiments to Religious Studies may well have nothing to do with the object of study, and instead may explore the thinking of the A.I. instead. A.I. is literally an “artificial intelligence”, and thus an “alien intelligence” in the sense of being something different from a human intelligence. In the experiments that my research group performed with the Pauline/deutro-Pauline texts, we were not interested in finding new insights about the authorship of the Pauline

corpus. Instead, we were interested in understanding the limits and processes of EMMA, by applying EMMA to this well-trodden area of Biblical Studies. It was possible that insights regarding authorship attribution could have been forthcoming along the way; indeed, as we sought to illuminate EMMA's reasoning process, we gained a new appreciation for the work of Biblical Scholars in this area, but the goal was more to understand the alien intelligence that is A.I. Our explorations with EMMA taught us about the priorities and function of the A.I., but offered no real new perspectives on Paul. Again, the tools of ANT can help us see the actants in a network that includes A.I. and recognize that such alien thinking found in A.I. is not just a matter of algorithmic models that are disconnected from human processes of thinking, but actually a part of an entire network that calls for analysis. To my mind, such a project is worthwhile in itself, but its goal needs to be clear.

Thus, following J.Z. Smith, I have argued that classification is the hallmark of the work of the religious scholar. Artificial intelligence today can aid in classification, particularly in situations where there is an immense quantity of data (for example the project digitizing the Danish theologian N.F.S. Grundtvig's writings (Lane 2019)), and yet as Smith noted, and as the practitioners of ANT have reiterated, classification itself, whether a part of A.I. or separate from A.I., is not a neutral intermediary—the tool constructs the worker and the task. Artificial intelligence likewise represents the same danger and the same promise.

Funding: This research received no external funding.

Acknowledgments: My sincerest thanks to the student members of my research team: Trey Adams, Carson Burleson, Elizabeth Henry, Ryan Schlecht, and Cameron Small. Particular thanks go to Courtney Dixon who scraped the data used in the team's experiments.

Conflicts of Interest: The author declares no conflict of interest.

Notes

- ¹ For the purposes of this paper, I will be discussing only narrow A.I. and specifically that part of narrow A.I. that is most often described as "Deep Learning." Other kinds of artificial intelligence both in the present and for the future are outside the domain of my inquiry for this paper.
- ² Since the 19th century when the critical study of religion came into its own, it has always seen farther because it stood on the shoulders of giants from other fields. From the phenomenologists (Studstill 2000; McCutcheon 2001) to the post-modernists (Moore 1994; Caputo 2007; McCutcheon 1997), Religious Studies has always been a "bricolage" field, often cobbling together various approaches from a variety of fields like Philosophy, Literary Studies, and Cognitive Science. We beg, borrow and steal because we are not, and, since the abandonment of theology as our sole methodology, cannot be, a discipline. We are a field of study, we import methods, apply those methods to our data and export explanations.
- ³ In Biblical Studies, the Semeia Journal published between 1975 and 2002 was the best example of this, with each issue exploring a new methodological approach to understanding the Biblical texts.
- ⁴ Scholars presenting at the newly founded Artificial Intelligence and Religion Research Seminar at AAR (2019–2024) are engaging in these sorts of sophisticated experiments with intriguing results.
- ⁵ Reinforcement learning (RL) gives an A.I. a task (usually in a simulated environment) and then rewards the A.I. for accomplishing the task. This is most often used in games, where the A.I. has a score that functions as a reward and then develops a series of strategies to maximize the score. While I think RL has real potential for some applications, its drawback is that it needs a simulated environment where it can continuously try less than optimal strategies to accomplish the goal. This makes it ideal for learning to play games like Chess, Go, Dota, or StarCraft (to name a few that have been in the news), but it is less conducive to more real-world applications that cannot afford the implementation of less effective approaches. To date I have yet to see an application of this technology for more pedestrian tasks such as moving a column of names from one application to another. Whether there is application for this in simulation studies remains to be seen. The goal of most reinforcement learning is to master a domain given a set of rules where the goal is clear (most often winning). Whether this can be expanded to open-ended Religious Studies problems in simulation is unclear. Even if so, the question arises of what machine decision making will tell us about human decision making, as often the goal of RL is to allow machine creativity, which often innovates in different ways than humans.
- ⁶ The quest for the historical Jesus has largely been a classificatory endeavor, as scholars have sought to separate authentic Jesus sayings and actions from those invented by the tradition. Likewise, the debate of the authorship of some of the Pauline letters hinges on whether certain texts were written by Paul, or were the product of a later church writing in his name. Finally, there is the on-going work about whether and to what degree Greco-Roman sources influenced early Christian work.
- ⁷ See a similar argument about religion in (Smith 1998).

- ⁸ A perspective he condemns earlier (Smith 2000, p. 36).
- ⁹ The problem of Google photo's A.I. making racist identifications shows the real problem of bias. An interesting discussion of this problem by the project manager at Google is worth reading (Zunger 2017). Other discussions of the problem can be found in (Vincent 2018; Thompson 2019; Noble 2018), while Cathy O'Neil has explored the larger problem of algorithmic bias in a variety of domains (O'Neil 2016). There is a growing literature about the problem of bias in A.I. by scholars who are particularly concerned about (though not exclusively) racial bias. (Buolamwini and Gebru 2018; Raji et al. 2020).
- ¹⁰ It is unclear what model EMMA is using. Given that it was released in 2016–2017, it seems likely that it is using word2vec for its language model, though blog posts pseudonymously written by the A.I. maintain that multiple stylometric algorithms were used (Identity 2017). Attempts to contact the authors to get more information about the inner workings of EMMA have gone unanswered. The information I do have was gleaned from promotional material on the web site, and a few short articles on technology blogs. There was a running blog in EMMA's name on Medium in 2017 (<https://medium.com/search?q=emma%20identity> accessed on 20 June 2019). Since this paper was written the website for EMMA has become defunct.
- ¹¹ The Pauline corpus is divided by scholars into the authentic works and the disputed works. The authentic works of Paul are Romans, 1 Corinthians, 2 Corinthians, 1 Thessalonians, Galatians, Philipians and Philemon. These texts share a vocabulary, style, theology and generally similar context (Mack 1996; Tabor 2012). Any analysis of the writings of Paul begins with these texts as scholars generally acknowledge these were written by Paul. Sometimes these are referred to as the uncontested letters of Paul. The other letters are more and less disputed. 2 Thessalonians, Ephesians, Colossians are thought by most critical scholars to be written pseudonymously, but this is not a universal perspective (Brown 1997). On the other hand, opinions about the Pastoral letters, Titus, 1 Timothy, and 2 Timothy are more uniform, implying that Paul is not the author. Hebrews, which is not claimed to be written by Paul, but which was sometimes ascribed to him by tradition, is generally rejected by scholars as Pauline (Ehrman 2011).
- ¹² EMMA is able to tell the difference between J.Z. Smith's writing and my own.
- ¹³ One thing we observed was that EMMA does not update her model after being given the initial input. In a sense, she only learns with the first example text given. EMMA does not seem to amend her model. Texts given after that first one, even though the user will identify whether EMMA is correct or incorrect, are not used to update EMMA's conclusions. Thus if one starts with 1 Corinthians and EMMA concludes that Romans is not Pauline, then even if the user identifies Romans as Pauline, if Romans is given again, it will not return an updated assessment, but will still return its initial conclusion.
- ¹⁴ The proprietary nature of EMMA is only partially connected to the problem of explainability that I have highlighted here. Even if EMMA had been open-source, it is the nature of neural networks that their decision making remains opaque. These models consist of a series of weights that are then used by an algorithm to make classifications. The exact meaning of the weights is not comprehensible regardless of whether the model is open- or closed-source. However, as I will discuss below, there are explainability frameworks that are increasingly being used and more model pipelines are including this as an option in both open- and closed-source projects. However, in this case, since EMMA is closed-source, we are unable to determine which models and other programs were being used in her decision making process.
- ¹⁵ A side note. It is unclear exactly the role neural networks play in EMMA's analysis. A project that I am just beginning (presently titled PaulAI) attempted to use a more recent neural network model to determine if it could do better than EMMA. I trained an LSTM model (Howard and Ruder 2018) on wikipedia (which, with over 6 million articles, functions to give the model an understanding of the English language), and then fine tuned the model on the NIV text. Then, I fed it the authentic Pauline texts (excluding disputed texts) and trained a classifier. Ultimately, the model reached a 97% accuracy rate. I then gave it the authentic and pseudonymous Pauline texts. It correctly identified all texts as either Paul or non-Pauline. The work is preliminary and the implications of this have yet to be clearly determined; nevertheless it does seem to show that a properly fine-tuned language model might have more power than we can see with EMMA.

References

- A.I. Explainability 360. n.d. Available online: <http://aix360.mybluemix.net/> (accessed on 19 October 2019).
- AlphaGo: The Story So Far. n.d. Deepmind. Available online: <https://deepmind.com/research/case-studies/alphago-the-story-so-far> (accessed on 21 May 2020).
- Bostrom, Nick. 2016. *Superintelligence: Paths, Dangers, Strategies*, reprint ed. Oxford and New York: Oxford University Press.
- Brown, Raymond E. 1997. *An Introduction to the New Testament (The Anchor Yale Bible Reference Library)*, 1st ed. New Haven: Yale University Press.
- Buolamwini, Joy, and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. Edited by Sorelle A. Friedler and Christo Wilson. Proceedings of Machine Learning Research. New York: PMLR, vol. 81, pp. 77–91.
- Caputo, John D. 2007. *What Would Jesus Deconstruct? The Good News of Postmodernism for the Church*. Grand Rapids: Baker Academic.
- Creative Destruction Lab. 2017. Suzanne Gildert: AGI and Embodiment. Youtube. Available online: <https://www.youtube.com/watch?v=3METDqWpVyY> (accessed on 10 November 2017).
- Durkheim, Émile. 2008. *The Elementary Forms of the Religious Life*. Translated by Joseph Ward Swain. New York: Dover Publications.
- Durkheim, Emile, and Marcel Mauss. 1967. *Primitive Classification*. Translated by Rodney Needham. Chicago: University of Chicago Press.

- Ehrman, Bart D. 2011. *Forged: Writing in the Name of God—Why the Bible’s Authors Are Not Who We Think They Are*, 1st ed. New York: HarperOne.
- EMMA. Defining Writing Identity. Disrupting Plagiarism. n.d. Available online: <https://emmaidentity.com/press-kit> (accessed on 2 August 2019).
- Finke, Roger, and Jennifer M. McClure. 2017. Reviewing Millions of Books: Charting Cultural and Religious Trends with Google’s Ngram Viewer. In *Faithful Measures: New Methods in the Measurement of Religion*. Edited by Roger Finke and Christopher D. Bader. New York: New York University Press, pp. 287–316.
- Geraci, Robert M. 2012. *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality*, reprint ed. Oxford: Oxford University Press.
- Gerrish, Sean, and Kevin Scott. 2018. *How Smart Machines Think (The MIT Press)*, 1st ed. Cambridge: The MIT Press.
- Glass Is the New Black—Intelligible A.I. Is within Your Grasp. 2019. Microsoft Research. Available online: <https://www.microsoft.com/en-us/research/blog/creating-ai-glass-boxes-open-sourcing-a-library-to-enable-intelligibility-in-machine-learning/> (accessed on 10 May 2019).
- Goertzel, Ben. 2016. *AGI Revolution: An Inside View of the Rise of Artificial General Intelligence*. New York: Humanity + Press.
- Goertzel, Ben, Cassio Pennachin, and Nil Geisweiller. 2014. *Engineering General Intelligence, Part 1: A Path to Advanced AGI via Embodied Learning and Cognitive Synergy (Atlantis Thinking Machines)*, 2014th ed. Dordrecht: Atlantis Press.
- Gore, Ross, Carlos Lemos, F. Leron Shults, and Wesley J. Wildman. 2018. Forecasting Changes in Religiosity and Existential Security with an Agent-Based Model. *Journal of Artificial Societies and Social Simulation* 21: 4. Available online: <http://jasss.soc.surrey.ac.uk/21/1/4.html> (accessed on 25 October 2019). [CrossRef]
- Harari, Yuval Noah. 2017. *Homo Deus: A Brief History of Tomorrow*, 1st ed. San Francisco: Harper.
- Howard, Jeremy, and Sebastian Ruder. 2018. Universal Language Model Fine-Tuning for Text Classification. *arXiv* arXiv:1801.06146.
- Identity, Emma. 2017. Writing Styles and Their Juicy Ingredients: Opinion by One Artificial Intelligence. Medium. The Creative Cafe. Available online: <https://thecreative.cafe/writing-styles-and-their-juicy-ingredients-opinion-by-one-artificial-intelligence-a62cd6a84ed4> (accessed on 31 May 2017).
- Lane, Justin. 2013. Method, Theory, and Multi-Agent Artificial Intelligence: Creating Computer Models of Complex Social Interaction. *Journal for the Cognitive Science of Religion* 1: 161. [CrossRef]
- Lane, Justin. 2019. The Promise of Reincarnation in the Grundtvig AI. The Religious Studies Project. Available online: <https://www.religiousstudiesproject.com/2019/10/18/the-promise-of-reincarnation-in-the-grundtvig-ai/> (accessed on 18 October 2019).
- Latour, Bruno. 1999. *Pandora’s Hope: Essays on the Reality of Science Studies*, 1st ed. Cambridge: Harvard University Press.
- Latour, Bruno. 2007. *Reassembling the Social: An Introduction to Actor–network–Theory (Clarendon Lectures in Management Studies)*, 1st ed. Oxford: Oxford University Press.
- Mack, Burton L. 1996. *Who Wrote the New Testament? The Making of the Christian Myth*, 1st ed. San Francisco: HarperOne.
- Marcus, Gary, and Ernest Davis. 2019. *Rebooting AI: Building Artificial Intelligence We Can Trust*. New York: Knopf Doubleday Publishing Group.
- McCutcheon, Russell T. 1997. ‘My Theory of the Brontosaurus’: Postmodernism and ‘Theory’ of Religion. *Studies in Religion. Sciences Religieuses* 26: 3–23. [CrossRef]
- McCutcheon, Russell T. 2001. *Critics Not Caretakers: Redefining the Public Study of Religion (Sunny Series, Issues in the Study of Religion)*, 1st ed. Albany: State University of New York Press.
- Mitchell, Melanie. 2020. *Artificial Intelligence: A Guide for Thinking Humans*. Stuttgart: Picador.
- Moore, Stephen D. 1994. *Poststructuralism and the New Testament: Derrida and Foucault at the Foot of the Cross*. Minneapolis: Augsburg Fortress.
- Noble, Safiya. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*, 1st ed. New York: NYU Press.
- O’Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, reprint ed. Portland: Broadway Books.
- Open-Sourcing Captum: A Model Interpretability Library for PyTorch. n.d. Available online: <https://ai.facebook.com/blog/open-sourcing-captum-a-model-interpretability-library-for-pytorch> (accessed on 19 October 2019).
- Raji, Inioluwa Deborah, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the A.I. Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. In *FAT* ’20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. New York: Association for Computing Machinery, pp. 33–44.
- Russell, Stuart. 2019. *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking.
- Saikou, Y. Diallo, Wesley J. Wildman, F. LeRon Shults, and Andreas Tolk, eds. 2019. *Human Simulation: Perspectives, Insights, and Applications*. Berlin and Heidelberg: Springer.
- Shults, F. Leron, Ross Gore, Wesley J. Wildman, Justin E. Lane, Chris Lynch, and Monica Toft. 2017. Mutually Escalating Religious Violence: A Generative Model. In *Proceedings of the Social Simulation Conference*. Dublin: Ireland, pp. 1–12. Available online: https://www.researchgate.net/publication/321170531_Mutually_Escalating_Religious_Violence_A_Generative_Model (accessed on 25 October 2019).

- Shults, F. Leron, Ross Gore, Carlos Lemos, and Wesley J. Wildman. 2018a. Why Do the Godless Prosper? Modeling the Cognitive and Coalitional Mechanisms That Promote Atheism. *Psychology of Religion and Spirituality* 10: 218–28. [CrossRef]
- Shults, F. Leron, Ross Gore, Wesley J. Wildman, Christopher J. Lynch, Justin E. Lane, and Monica D. Toft. 2018b. A Generative Model of the Mutual Escalation of Anxiety between Religious Groups. *Journal of Artificial Societies and Social Simulation* 21: 1–25. Available online: https://digitalcommons.odu.edu/vmasc_pubs/38/ (accessed on 25 October 2019). [CrossRef]
- Smith, Jonathan Z. 1992. *To Take Place: Toward Theory in Ritual*. Chicago: University of Chicago Press.
- Smith, Jonathan Z. 1994. *Drudgery Divine: On the Comparison of Early Christianities and the Religions of Late Antiquity*. Chicago: University of Chicago Press.
- Smith, Jonathan Z. 1998. Religion, Religions, Religious. *Critical Terms for Religious Studies* 1998: 269–84.
- Smith, Jonathan Z. 2000. Classification. In *Guide to the Study of Religion*. Edited by Willi Braun and Russel T. McCutcheon. Continuum: Bloomsbury Publishing, pp. 35–44.
- Smith, Jonathan Z. 2004. *Relating Religion: Essays in the Study of Religion*, 1st ed. Chicago: University of Chicago Press.
- Studstill, Randall. 2000. Eliade, Phenomenology, and the Sacred. *Religious Studies Review* 36: 177–94. [CrossRef]
- Tabor, James D. 2012. *Paul and Jesus: How the Apostle Transformed Christianity*. New York: Simon & Schuster.
- Tegmark, Max. 2017. *Life 3.0: Being Human in the Age of Artificial Intelligence*. Vintage: Knopf.
- Thompson, Clive. 2019. *Coders: The Making of a New Tribe and the Remaking of the World*. Kindle and London: Penguin Press.
- Vincent, James. 2018. Google ‘fixed’ Its Racist Algorithm by Removing Gorillas from Its Image-Labeling Tech. *The Verge*. January 12. Available online: <https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai> (accessed on 12 January 2018).
- Zunger, Yonatan. 2017. Asking the Right Questions about AI. Medium. Medium. Available online: <https://medium.com/@yonatanzunger/asking-the-right-questions-about-ai-7ed2d9820c48> (accessed on 12 October 2017).