Righteous AI: The Christian voice in the Ethical AI conversation

Gretchen Huizinga

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading committee:

Stephen Kerr, Chair

Walter Parker

James Wellman

Edward Lazowska, GSR

Program authorized to offer degree:

College of Education

Copyright

Extended Abstract

Background: Artificial intelligence (AI) is a priority for tech companies today. Considering its perceived value and power, people are paying attention to both the promise and the peril of AI. On the promise side, the main concern is what AI *can* do. On the peril side, the main concern is what AI *ought* to do. This has prompted a conversation, and an emerging body of literature, around Ethical AI principles. While religious traditions provide a wealth of wisdom concerning human moral behavior, secular ethical frameworks have become the acceptable rhetorical scaffolding for articulating guiding ethical principles, especially in Western technical circles. Religious perspectives have been marginalized and ethics has been framed in humanistic rather than transcendent terms.

Purpose: The purpose of this study was to explore how Christian teaching, texts, and traditions might make a valuable contribution to the discussion of Ethical AI. The objective was not to displace or silence other voices but to add a missing perspective and bring viewpoint diversity to the conversation and the literature.

Framework: As a framework for the research, I proposed that humanistic ethical principles, even if codified into laws and regulations, are necessary but insufficient to ensure robust and beneficial AI. I further proposed that acknowledgment of divine intelligence, along with an ordinate (or rightly ordered) understanding of human intelligence, is foundational to the development and use of artificial intelligence and therefore, religious voices should have a say in framing the ethical scaffolding around it.

Research Design: This basic qualitative study explored the Christian voice in AI ethics and focused on three main questions: 1) How does worldview affect our approach to artificial intelligence? 2) Does a Christian worldview have anything unique to contribute to the discussion around Ethical AI? and 3) How might AI ethics be more robust and beneficial if we brought Christian teachings, texts, and traditions explicitly into the conversation? Using a semi-structured question protocol as the primary data-collection instrument and a constant comparison method of data analysis during both collection and analysis phases, I conducted interviews with a purposeful sample of AI/tech/ethics professionals who were also professing Christians to identify key themes that differentiate the Christian ethical worldview from the materialist worldview that currently animates the conversation around Ethical AI.

Findings/Significance: My findings suggest that worldview (both implicit and explicit) informs every aspect of our approach to Ethical AI. While materialist thought seeks to compel humans to be good without transcendent reason or power, the Christian faith speaks clearly about the role of God as originator, motivator, and sustainer of human moral behavior. Christianity compels us to look beyond a humanistic idea of ethics and toward a creative notion of goodness that cannot be accomplished by our own will and power. This study adds critical insights to the field of AI ethics by deepening awareness of how faith in and fear of God could influence how artificial intelligence is designed and implemented. When Christian wisdom is included in every phase of AI development, we begin to think beyond a minimum-standard culture of Ethical AI and move toward a robust culture of Righteous AI.

Preface

My research, while situated at the intersection of artificial intelligence, ethics, and Christianity, is anchored in the College of Education at the University of Washington. This may seem odd, and early on it even did to me as I tried to connect the dots for puzzled friends, colleagues, and potential committee members who questioned the connections with furrowed brows. Since embarking on the project, however, I have come to believe that the College of Education may, in fact, be the ideal discipline from which to present my research.

The very concept of artificial intelligence is based on the idea that we can make machines that learn and ultimately think, hence the terms machine *learning* and artificial *intelligence,* and much of the current language surrounding these concepts borrows heavily from biological models of human learning and cognitive science. Machine learning methodologies like **deep learning**, which co-opts the language of neuroscience by developing artificial neural networks "heavily inspired by the way biological nervous systems (such as the human brain) operate" (O'Shea and Nash, 2015), and **reinforcement learning** which is modeled on the stimulus-reward architecture of classical conditioning documented by Ivan Pavlov (Pavlov, 1927), guide but two of the many current methodologies being used to simulate a human brain with a computer. Other, more blue sky, methodologies include things like brain-machine interfaces that use electrode threads to stimulate or restore certain functions of the human brain. (Musk, 2019)

But while computational models for intelligence have yielded some amazing results in very specific and narrow tasks, some say AI systems are less about automation than "fauxtomation" (Taylor, 2018) and the real brains behind computer brains are still people. Even if more sophisticated than the original Mechanical Turk or the Wizard of Oz, AI today is similar in nature: everything in a computer was devised by human intelligence and much of the labor

behind the magic is done by humans behind the scenes. Our thinking machines have been likened to "…a kind of Potemkin AI – little more than facades, designed to demonstrate to investors and a credulous media what an automated system would look like while actually relying on human labor in the background." (Crawford, 2021, p. 65, citing Sadowski, 2018) In truth, AI has yet to outperform people when it comes to things like learning transfer and general intelligence – the miraculous and yet-to-be-understood process of human growth and learning – and there is no evidence that computerized *brains* will ever become *minds* that think, understand, contemplate, or create like humans. It is worth interrogating the claims of those who say it will.

So, there's a connection between computer science and cognitive science – machine learning and human learning – but what about ethics and religion at the UW COE? Historically, the purpose of education has always included the moral as well as cognitive development of young people, and researchers and educators alike have long recognized the interconnected nature of cognitive, psychological, motivational, and moral maturity (Bloom, 1956; Piaget, 1932; Maslow, 1943; Kohlberg, 1958). While concepts of traditional moral, ethical, or values-based education are largely ignored in modern-day American university education departments, and therefore do not find their way into public school pedagogy, the idea of values education has persisted in the home and in the church and has always been a priority for parents and religious leaders. Even those who eschew faith in God and so-called traditional family values still embrace a morality of their own, often reifying it in the form of political activism if not religious devotion.

While many will argue about how to provide moral education (and indeed what constitutes moral behavior), most would agree that social qualities like empathy, honesty, self-sacrifice, justice, and mercy are among the characteristics recognized in well-formed, well-

educated people. These are the qualities, if inculcated in a society, that enable us to get along with one another. Even to love one another. Interestingly, ideas associated with intellectual development such as knowledge, understanding, and wisdom are difficult to separate from social qualities. Education scholars like Lev Vygotsky have proposed that cognitive development and psychological development are inseparable:

> Within a general process of development, two qualitatively different lines of development, differing in origin, can be distinguished: the elementary processes, which are of biological origin, on the one hand, and the higher psychological functions, of sociocultural origin, on the other. The history of child behavior is born from the interweaving of these two lines. The history of the development of the higher psychological functions is impossible without a study of their prehistory, their biological roots, and their organic disposition. (Vygotsky, in Cole, 1978, p. 46)

In this way we can see a connection between mind and soul – a link between the brain and the heart – that should give us pause to consider whether our current vision of the importance of personalized instruction and individualized education (now facilitated by comprehensive and expensive educational AI applications) is yet another by-product of a computerized view of the human brain at the expense of a sociocultural view of human development. If the purpose of education is holistic and if love is as important as intelligence, it seems appropriate to ask, in AI terms, what are we optimizing for?

This is why I'm conducting research on the Christian voice in Ethical AI, and this is why I'm situating it in the University of Washington College of Education.

Contents

List of Figures

Introduction

Background

Among the myriad technological innovations in the world today, perhaps none is imbued with more hope – or more hype – than artificial intelligence, or AI[1]. Because of both the hope and the hype, AI has now become a top priority for tech companies, governments, and other organizations (Cognilytica Market Intelligence, 2020) and has found its way into nearly every aspect of our lives, and nearly all our major societal institutions. Its proponents, among them academic researchers (e.g., Domingos, 2015) working on AI for scientific advancement and corporations (e.g., Google, Microsoft, IBM, Amazon) working on AI for profit, claim that AI will fundamentally remake the world for the better. Its critics (e.g., Crawford, 2021) maintain that these images of AI's potential are, in many ways, built on misunderstanding, exploitation, and even deceit. No matter who's right – or even if everybody is – people are paying attention to both the promise and the peril of AI. On the promise side, tech companies are developing the science in the race for AI dominance where the main concern is what AI *can* do. On the peril side, there is an increasing interest in the ethical development and use of AI, where researchers are asking what AI *ought* to do. Both the technical and ethical research is founded, generally, on a materialist view of the world that denies the existence of God and the idea of spiritual rewards and punishments, but demands human altruism and ethical behavior nonetheless. As part of an organization called AI and Faith, I am keenly interested in the moral issues surrounding the ethics of AI and specifically, how voices of religious faith (specifically, the Christian faith) might contribute to and shape the conversation.

---

[1] AI has many definitions and applications, and the term continues to change as technology advances, but for operational purposes, AI will be defined as **any computational mechanism or device that performs or simulates human thought and/or behavior**.

The term AI itself was coined in 1956, when a group of computer scientists began in earnest to lay the groundwork (as published in the 1955 Proposal for the Dartmouth Summer Research Project on Artificial Intelligence) for an intelligent – or learning – machine, under the conjecture that human intelligence could be "so precisely described that a machine can be made to simulate it." (McCarthy et al, 1955, p. 2) In the nearly seventy years since, the path to computational intelligence has been remarkably successful, especially of late, thanks to the current "Holy Trinity" of AI innovation: large data sets, sophisticated algorithms, and unprecedented compute power. Like many of its computational peers and predecessors, AI has been touted to improve, enhance, augment, extend, and even upgrade human life. But because the mission of many AI research labs is to build machines that learn from and understand the world in much the same way humans do and ultimately simulate human capabilities (thought, reasoning, understanding, sight, speech and language, decision-making…), it is substantively different. Even when the science of AI was still solidly fixed in the realm of fiction, author Isaac Asimov recognized the need for ethical guidelines and wrote his famous Three Laws of Robotics[2] to protect humans from autonomous intelligent agents (aka robots). Now that AI is more science than fiction, some (e.g., Anderson, 2017) suggest those laws need updating to reflect real life in the 21st century. Others (e.g., Martius, 2017) say we need alternative solutions that empower robots with heuristics that would replace the need for Asimov's laws. Speculations aside, rogue robots with moral agency appear to be a technically distant reality, so the focus for current AI ethics is more on where the science lives right now. This is actually biblical, as Jesus

---

[2] Asimov's Laws are: 1) A robot may not injure a human being or, through inaction, allow a human being to come to harm. 2) A robot must obey the orders given it by human beings except where such orders would conflict with the First Law. 3) A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

tells his disciples, "Therefore do not be anxious about tomorrow, for tomorrow will be anxious for itself. Sufficient for the day is its own trouble." (ESV Bible, Matthew 6:34)

Today, AI presents us with ethical issues that include social, commercial, and political concerns. Recent books and articles detail the way corporations can predict and even direct human behavior through AI tracking and monitoring (e.g., Zuboff, 2019) and the way governments can exercise social control of their citizens through AI espionage (e.g., Mitchell, A., Diamond, L., 2018), giving many cause for concern. Even as these corporations and government entities assure us of their commitment to social responsibility, they continue to prioritize power and profit over the common good. In response to this practice of so-called "ethics theater" or "ethics washing," scholars (e.g., Gebru, 2021) have published articles, even at the expense of their own employment, to call out and hold these powerful entities to account. Other organizations and institutions (see Appendix A, *Institutions*) have published calls for AI ethics, issued guidelines for human-AI interaction, established best practices for heads of governmental departments and agencies, passed continent-wide laws and regulations, and put out statements of principles, all addressing their vision for the ethical development and use of AI. That said, establishing rigorous, widely accepted ethical standards is difficult enough with technologies we understand well and that are clearly *distinct from us*. AI is a suite of technologies that seeks to *be like us* and even boasts a famous challenge – originally known as The Imitation Game (Turing, 1950) and now called the Turing Test – by which we judge whether a machine we've created can fool us into believing it's human. This reality takes us into new territory and demands that we ask larger questions around the ethics of AI than we have of other technologies. Simply rehashing the "what" of ethics, even if in stronger legal terms, may not be enough to navigate the novel ethical challenges associated with AI. Only a robust "why"

can compel humans to seek out, embrace, and live an ethical life. And only a robust "how" can equip and enable them to do so. Before we get to the why and how, though, we must answer a foundational question: what constitutes an ethical life?

Perhaps surprisingly to some, it's not just about following a bunch of rules. Technology ethics scholar Shannon Vallor puts it this way:

> At its most basic, ethics is about what the ancient Greek philosopher Socrates called 'the good life': the kind of life that is most worthy of a human being, the kind of life worth choosing from among all the different ways we might live. While there are many kinds of lives worth choosing, most of us would agree that there are also some kinds of lives not worth choosing, since we have better alternatives. For example, a life filled mostly with willful ignorance, cruelty, fear, pain, selfishness, and hatred might still have some value, but it would not be a kind of life worth choosing for ourselves or our loved ones since there are far happier choices available to us – better and more virtuous ways that one can live, for ourselves and everyone around us. (Vallor, 2008, p. 2)

Vallor goes on to ask what ethics or moral philosophy has to do with technology, and then answers her own question, saying,

> Human social practices, including our moral practices, have always intertwined with our technologies…ethics and technology are connected because technologies invite or *afford* specific patterns of thought, behavior, and valuing: they open up new possibilities for human action and foreclose or obscure others… Thus 21st century decisions about how to live well – that is, about ethics – are not simply moral choices. They are *technomoral* choices for they depend on the evolving affordances of the technological systems that we

rely upon to support and mediate our lives in ways and degrees never before witnessed.
(Ibid, p. 2)

So far, some may argue, we seem to be doing just fine approaching both AI technology and AI ethics from a secular point of view. What do we need religion for? In short, I propose that we're actually not doing fine, and that from a Christian point of view, Jesus is the basis of an ethical life, but there are three other arguments to be made as to why religion might beneficially find its way into the scientific discourse around the ethics of artificial intelligence.

1) **The world is religious**. According to a Pew Global Religious Landscape study, "A comprehensive demographic study of more than 230 countries and territories conducted by the Pew Research Center's Forum on Religion & Public Life estimates that there are 5.8 billion religiously affiliated adults and children around the globe, representing 84% of the 2010 world population of 6.9 billion." (Hackett, Grim, 2012, p.9) Even if the religious landscape among high tech workers is predominantly *non*-religious (perhaps especially if it is predominantly non-religious), it makes no sense to exclude a point of view held by 84% of the world's population. If a sizable majority of people in the world embrace some form of religious belief and/or practice, it is epistemologically sound to study a portion of this population.

2) **The foundations of ethics are religious**. While many of the buzzwords surrounding Ethical AI (dignity, equality, safety, security, fairness, justice, etc.) find a comfortable home among modern secular audiences, these very values are thousands of years old and, arguably, rooted in religious traditions, even if overt religious overtones have been softened or expunged to make them more palatable to a non-religious crowd. What's more, even people who self-identify as secular or atheistic have strong notions of right and wrong; fair and unfair; honest or dishonest; moral and immoral; good and evil; all terms that, historically, fall firmly in religious

territory. Given that both ethics and religion call us to make moral, good, or wise decisions, and given that humans have increasingly begun to outsource these decisions to AI algorithms, often without their knowledge or understanding, it seems wise to get a better grip on the origin, or provenance, of ethics, as well as the efficacy of ethics unbound by religion if we discover faith in God in the ethics family tree.

3) **Ethics tells us to be good. Religion tells us why and how to be good.** All ethical frameworks tell us we must *be* good and include some notion of what that looks like. Most fall short, however, when it comes to convincing us *why we should* be good or explaining *how we can* be good. Rather, these ethical frameworks rest on assumptions of altruism and self-sacrifice that regularly come in second when pitted against the powerful human desires for power, privilege, and profit. Without a transcendent reason for selflessness and/or a supernaturally enabling force within (and, theologically, even *with* those two things), it is impossible for humans to be perfectly, consistently, sustainably good, or ethical. According to the Bible, "If we say we have no sin, we deceive ourselves and the truth is not in us." (ESV Bible, 1 John 1:8) And "For I know that nothing good dwells in me, that is, in my flesh. For I have the desire to do what is right, but not the ability to carry it out." (ESV Bible, Romans 7:18) Even Jesus, when called 'good teacher' by a rich, young ruler challenges his conception by asking, "Why do you call me good? No one is good except God alone." (ESV Bible, Luke 18:19) Absent God, according to Christian scripture, ethics can preach but it cannot empower. That said, this assertion is made from a religious – and specifically Christian – point of view. Is anyone outside the fold, so to speak, curious about the origin and efficacy of ethics, particularly as they relate to AI technologies?

Technology philosopher and author Shannon Vallor is. In her seminal text *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*, Vallor asks, "How can humans hope to live well in a world made increasingly more complex and unpredictable by emerging technologies?... In essence, my answer is this: we need to cultivate in ourselves, collectively, a special kind of moral character, one that expresses what I will call the *technomoral virtues*." (Vallor, 2018, p. 1, emphasis, the author's) She goes on to advocate for "the contemporary renewal of virtue ethics" which "treat virtue and character as more fundamental to ethics than moral rules or principles." (Ibid, p. 10) She notes that in the recent past, deontological and utilitarian ethics have become the dominant approaches in philosophical discourse, while virtue ethics has fallen out of favor due to 1) its "strong associations with the Thomistic moral theology of the Catholic Church," 2) the fact that it was "seen as incompatible with evolutionary science… which denied… that human lives are naturally guided toward a telos, a single fixed goal or final purpose," and 3) its "emphasis on habit and emotion was also seen as undermining rationality and moral objectivity." (Ibid, p. 20)

Recently though, the virtue ethics approach has been making a comeback as the other approaches fail to deliver the goods in a technically complex, AI-infused world. Vallor is careful to separate "church and state" in her work, asserting that virtue ethics must be decoupled from religious traditions because those systems "speak only to their believers, and are thus poor candidates for a global technosocial ethic." (Ibid, p. 23) This may resonate with atheists and other unbelievers, but the Christian case posits that while all forms of ethics are a good start, no form of ethics (including deontological, utilitarian, relativistic, and even virtue ethics unmoored from divine empowerment) will be a "good candidate for a global technosocial ethic" unless it is in alignment with the ultimate reality of a good, loving, holy, and righteous God. This is not to

say that Christians believe their faith should supersede all other approaches to ethics in a

pluralistic world. It is simply to say that in that pluralistic world, Christian ethics offers an

approach to human benevolence that does not depend solely on human effort but rather, as

Christians claim, is enabled by a free gift of grace, made possible through the death and

resurrection of Christ, and empowered by the Holy Spirit.

In yet another milieu, journalism, we discover others who are thinking deeply – and

asking big questions – about the issue of ethics in high tech. In a recent New York Times op-ed

entitled *Can Silicon Valley Find God?*, journalist Linda Kinstler explored how ethics, religion,

and technology co-exist in Silicon Valley. She found that her "conversations often skirted

narrowly past the question of religion, alluding to it but almost never engaging with it directly."

(Kinstler, 2021, online article*,* paragraph 6) Her interviewees:

> …spoke of shared values, customs, and morals, but most were careful to stay confined to
>
> the safe syntax of secularism. Amid increasing scrutiny of technology's role in
>
> everything from policing to politics, ethics had become an industry safe word, but no one
>
> seemed to agree on what those ethics were… So [she] started looking for people who
>
> were saying the quiet part out loud… tech workers who left plum corporate jobs to
>
> research the spiritual implications of the technologies they helped build, and those who
>
> chose to stay in the industry and reform it from within, pushing themselves and their
>
> colleagues to reconcile their faith with their work, or at the very least to pause and
>
> consider the ethical and existential implications of their products. (Ibid, paragraphs 6-8)

Even within devoutly secular high-tech communities, questions of how to talk about and

agree on ethics are venturing more boldly into religious territory as people begin to consider

more seriously how the ethical systems they currently embrace will fare against the advance of ever more powerful and potentially dangerous technologies.

Others share Kinstler's interest in the overlap of the ethical and spiritual in high tech as well. One example is Tristan Harris, a former Google employee who now heads the Center for Humane Technology, a non-profit funded by both major foundations and individual donors. Harris is famous for his viral presentation at Google titled *A Call to Minimize Distraction & Respect Users' Attention*. (Harris, 2013) His "Time Well Spent" movement has been featured at *TED*, on *60 Minutes*, and in the documentary film *The Social Dilemma*. Harris argues that we cannot trust tech to regulate itself, decries what he calls human downgrading with technology, and advocates for practices of eastern religions like meditation and mindfulness as part of the pushback against technical determinism. I, myself, am part of a growing constituency of people who believe that AI is encroaching on what religious believers consider to be our fundamental human design and want to, as Kinstler puts it, "make sure thousands of years of text and tradition find a place among the algorithms."

As researchers and corporations are working to develop and deliver *robust* and *beneficial* AI applications, they are discovering a variety of ethical challenges when it comes to delivering powerful technologies that also protect humans. As ethical issues arise, the first step in addressing them is usually developing a set of principles by which practitioners and users agree, in principle, to abide. An example of this can be found in the field of biotechnology, which has famously adopted four principles – beneficence, non-maleficence, autonomy, and justice (Varkey, 2020, p. 18) – to guide research, development, and use. But these are high-level principles on which no one would disagree. It only takes a couple clicks to get to a place where questions such as *whose justice?* or *beneficence for whom?* crop up. What's more, as we have

seen in biotech, even if principles are "universally" agreed upon, or even legislated in some countries, there is no global enforcement mechanism to prevent scientists in other countries from pushing beyond them for the sake of science or, in the case of governments, control and oppression. So it is with general principles for any discipline and that is why the foundational problem for AI is the insufficiency of principles alone to guide us toward a solid ethical framework for AI. It seems we might need to think beyond principles and even laws to address problems inherent with a suite of technologies as powerful, invasive, and fast moving as artificial intelligence.

Research Thesis and Purpose

For this study, I proposed that humanistic ethical principles, even if codified into laws and regulations, are necessary but insufficient to ensure robust and beneficial AI. I further proposed that acknowledgment of divine intelligence, along with an ordinate (or rightly ordered) understanding of human intelligence, is foundational to the development and use of artificial intelligence. Therefore, religious voices should have a say in framing the ethical scaffolding around it.

Because the bulk of the research and debate on Ethical AI has come from secular sources and is founded on a materialist worldview (which claims the physical universe is all that exists and there is no spiritual realm or being), the gap in the literature – or the missing voice – has been that of religious faith, and in this case, the Christian faith. To fill this gap, I proposed to seek out and include the voices of Christians who work in AI, AI ethics, or associated fields, in an effort to bring their views to the larger multidisciplinary community of experts, adding a religious perspective to the AI ethics discussion. I did not seek to displace or silence voices already in the conversation, but rather add a missing voice, bringing viewpoint diversity to the

discussion and introducing (or re-introducing) an historically embraced but scientifically marginalized point of view into the debate. Since Christianity has made truth claims and provided a framework for moral behavior for millennia and, perhaps more importantly, has asserted that moral behavior is the fruit not the root of a righteous life, it seems reasonable to bring those claims into the marketplace of ideas concerning the ethics of artificial intelligence, and let them be heard and discussed, argued and debated, accepted or rejected.

Research questions

These considerations led me to formulate these basic research questions to guide the development and interpretation of this study:

1. How does worldview affect how AI researchers and developers make artificial intelligence?
2. Does a Christian worldview have anything unique to contribute to the discussion around Ethical AI?
3. How might AI ethics be more robust and more beneficial if we brought Christian teachings, texts, and traditions explicitly into the conversation?

Literature Review

Background

Let's begin by acknowledging that AI is remarkable. In little more than seventy years, the science has advanced significantly in many areas that were previously considered human-only domains. Today, AI systems can credibly simulate sight, hearing, and speech thanks to research breakthroughs in computer vision (e.g., Krizhevsky et al, 2012), voice recognition technology (e.g., DARPA RATS; Stanford Research Institute SIRI), and natural language processing (e.g., OpenAI GPT-3). AI has conquered games of strategy like chess (e.g., IBM Deep Blue) and Go (e.g., DeepMind AlphaGo), and even beat Ms. Pac Man using a form of reinforcement learning called hybrid reward architecture (van Seijen et al, 2017). It can solve complex problems and make informed predictions for everything from advertising to health care (e.g., IBM Watson Advertising Accelerator, Microsoft Project Inner Eye). AI has democratized certain forms of personalized assistance previously afforded only the affluent (e.g., Google Assistant, Alexa, Cortana, Echo, Siri, etc.) and promises to make our lives more convenient and frictionless than any computer technology before it (e.g., Castro, New, 2016). Yet AI still has major limitations when compared to general human intelligence and, while it may perform well in laboratories and controlled test settings, is fraught with flaws and failings (e.g., Bergstein, 2020), both technical and ethical, in real world settings.

These flaws and failings have prompted concerns about AI systems that enable everything from smart homes to self-driving vehicles to autonomous weapons, and have prompted researchers from a variety of institutions to focus on AI's purported dangers and harms viewed through the lens of such topics as safety, privacy and security; diversity, equity and inclusion; justice, equality and bias; and fairness, accountability and transparency in socio-

technical systems; and apprehensions about the black box nature of advanced AI algorithms used in deep learning (e.g., Knight, 2017). Reports on the research conclude by calling for a commitment to the ethical design, development, and use of AI as spelled out in memoranda, statements of principles, recommendations, frameworks, and guidelines (see Appendix B, *Guidelines*).

Unfortunately, there seems to be broad consensus that this will not happen, at least not soon. According to a survey on the future of Ethical AI design,

> Pew Research Center and Elon University's Imagining the Internet Center asked experts where they thought efforts aimed at creating ethical artificial intelligence would stand in the year 2030. Some 602 technology innovators, developers, business and policy leaders, researchers and activists responded to this specific question: *By 2030, will most of the AI systems being used by organizations of all sorts employ ethical principles focused primarily on the public good, yes or no?* 68% said they expect that ethical principles focused primarily on the public good *will not be* employed in most AI systems by 2030. (Pew Research Center, Elon University, 2021)

Follow up questions included things like, "Will AI mostly be used in ethical or questionable ways in the next decade? Why? What gives you the most hope? What worries you the most? How do you see AI applications making a difference in the lives of most people? As you look at the global competition over AI systems, what issues concern you or excite you?" Key concerns among the responses centered around themes of the difficulty of defining Ethical AI, the unlikelihood of a global consensus on norms and standards, the lack of formal ethics training for the humans making AI, the opacity of AI systems already deployed, and the threat of a superpowers arms race with China. Even if any of these issues could be resolved, the fact that

"control of AI is concentrated in the hands of powerful companies and governments driven by motives other than ethical concerns" led respondents to worry that ethics will continue to take "a back seat."

So, what do we make of this and how do we move toward a robust and beneficial AI that is conceived, developed, and used ethically? Put succinctly, we simply cannot separate robust but benevolent AI ethics from robust but benevolent humans. Microsoft President Brad Smith makes this case in his book *Tools and Weapons: The Promise and the Peril of the Digital Age*, saying:

> As computers gained the ability to make decisions previously reserved for humans, virtually every ethical question for humanity was becoming an ethical question for computing. If millennia of debate among philosophers had not forged clear-cut and universal answers, then a consensus was not likely to emerge overnight simply because we needed to apply them to computers. (Smith, 2019, p. 199)

Smith refers here to philosophers (which would include the full spectrum of Eastern and Western schools of philosophy) but ethical structures – in the form of human codes of conduct, rules for behavior, notions of right and wrong – have been present in every religion from earliest recorded history. If current attempts to solve the problem of ethical machines are having difficulty finding purchase, it may mean that we need to expand the territory on the search map and, once again, include religion.

Currently, researchers and practitioners are looking at **technical** approaches and **ethical/legal** approaches to solve for Ethical AI, and they are looking primarily through a secular humanist – or materialist – lens. They argue that we must default to a non-religious framework because it would be impractical at best to select from among myriad religious frameworks in a pluralistic society. The problem with this argument is that a secular framework is, itself, a

religious framework: it simply embraces atheism rather than theism. Both are human conceptions

of and arguments for (or against) the existence of God. In this study, I look at a third approach –

an overtly theistic, religious approach – that is underrepresented in both academic discourse and

literature: Christian ethics. To be clear, this approach is not underrepresented *outside* academic

discourse and literature, but rather suffers exclusion from so-called scientific circles. It is not my

intent to prove, via this qualitative research, that a Christian approach to ethics in AI will solve

all its problems. In fact, I believe all areas are important – even necessary – to do that. Nor is it

my intent to advocate for a *techno-theocracy* of sorts. I simply argue that at present, the research

is centered on materialistic technology and secular ethics, and that religious perspectives should

not be excluded from the conversation. Here, I lay out some of what has been proposed or

attempted in each of these areas in order to try to make AI – and the humans who use it – behave.

Technical Approaches

As problems with AI have emerged and the ethical issues surrounding AI systems have

been questioned, computer scientists – an incredibly smart group of people who believe their AI

applications will be beneficial, not harmful, to humanity – have worked to mitigate the problems

caused by technology… with technology. These technical solutions are an important part of

developing Ethical AI, but since thorny problems in moral development and behavior have

plagued humanity for millennia, adding the weight of solving ethical conundrums to the pile of

difficult-to-intractable technical problems already facing computer scientists seems both

unrealistic and unfair. That said, technologists across the research and development spectrum are

taking these challenges seriously and continue to attempt to make smart machines that are also

good machines. For example, some problems linked to racially biased machine learning

outcomes stem from the fact that datasets used to train the ML models have consisted primarily

of light skinned faces (Bulowami, Gebru, 2018). That discovery has prompted machine learning experts to pay close attention to data provenance and seek more inclusive datasets on which to train models. But it turns out at least part of the problem of these seemingly biased outcomes was – and continues to be – purely technical: even the best machine learning algorithms still have difficulty recognizing darker skinned faces equally and it's not because they are "racist." It's because discerning among darker tones in anything, including faces, is very hard for a machine. In a Venture Beat blog, Kyle Wiggins reports,

> Every few months, the U.S. National Institute of Standards and Technology (NIST)
> releases the results of benchmark tests it conducts on facial recognition algorithms
> submitted by companies, universities, and independent labs… our findings cast doubt on
> the notion that facial recognition algorithms are becoming better at recognizing people of
> color. (Wiggins, 2020)

But the quest for more Ethical AI via better algorithms persists and NIST's benchmark test is ongoing.

In other areas of concern like privacy, scientists are working on things like differential privacy, which uses meta-algorithms that allow researchers in healthcare and social science to observe population-level data without exposing information about specific individuals (Niu, 2022). In the field of safety-critical AI, researchers are trying to develop what they call adaptive systems that can function in real world scenarios and make decisions under uncertainty with so-called situated intelligence (Bohus, et al, 2021). Taking the step of solving technology problems with technology one step further, there is also the idea of using AI to help us solve ethical problems. For example, Oxford professor emeritus, John Lennox, in his book *2084: Artificial Intelligence and the Future of Humanity* (Lennox, 2020), rhetorically suggests it would be

"…fascinating to apply AI in this way to a gigantic, crowdsourced database of moral choices to see what commonalities arose. In other words, apply AI to moral decision-making itself as a help to what morality should be programmed into various kinds of systems under development." (p. 149) Lennox goes on to say that "of course this runs the risk of determining morality in a utilitarian manner by majority vote which, as history shows, is not always a wise thing to do" (p. 149) but it represents the current thinking in how computers can reach conclusions by processing large amounts of data and suggesting the most common or probable outcome. At the time of writing, for Lennox, this was a thought experiment. Not long after his book was published, this very idea was implemented in proof-of-concept research and the results published in a paper titled *Delphi: Towards Machine Ethics and Norms* (Jiang, et al, 2021). The research took a run at the "majority rule" hill by crowdsourcing 1.7 million moral opinions to build what they call "a commonsense norm bank" culled from internet moral philosophy sources like Reddit's *Am I the Asshole?* and *Dear Abby*. Delphi used this data to train machine learning models to make good choices, so to speak. Realizing Lennox's fears about a computational model that leans toward culturally relative ethics, the authors say they

> …formalize morality as socially constructed expectations about acceptability and
>
> preference… largely influenced by the works in *descriptive* and *situational* ethics (Hare,
>
> 1981; Kohlberg, 1976; Fletcher, 1997) which make no claims of moral absolutes and
>
> accept that morality is determined by situation. Thus, rather than modeling moral 'truths'
>
> based on prescriptive notions of socio-normative standards, we take a bottom-up
>
> approach to capture moral implications of everyday actions in their immediate context,
>
> appropriate to our current social and ethical climate. (Ibid, p. 4)

Moreover, in framing their approach, the authors operationalize morality and ethics as fungible terms:

> In literature, morality deals with shared social values of what's right or wrong. Ethics, on the other hand, governs rules, laws and regulations that socially impose what is right or wrong. For example, certain spiritual groups may consider abortion morally wrong even if the laws of the land may consider it an ethical practice. In this paper, we do not make this distinction, and use both terms to refer to culturally shared societal norms about right and wrong. (Ibid, p. 3, footnote)

Interestingly, in choosing to "not make this distinction" the authors firmly position themselves on the side of non-transcendent ethical norms, a position which is, in itself, a distinction for which they offer no justification except "we don't."

Another proposed solution for Ethical AI calls for research on designing AI systems that have the capacity to obey human laws and values. In a paper titled just that, *Designing AI Systems that Obey Our Laws and Values* (Etzioni and Etzioni, 2016), Oren Etzioni and his co-author father suggest that we develop and deploy of a variety of "AI guardians" or oversight systems in the form of interrogators, auditors, monitors, enforcers and, delightfully, "ethics bots." Of course, the inherent problem here is, as even the authors ask, who will guard the guardians? This is not to mention the upstream problem of the word *our* in the title: whose laws and values these are these bots supposed to obey? AI is a global technology deployed in multiple countries with different legal systems and moral values. Even within countries, there are many sub-cultures among which it would be difficult-to-impossible to forge agreement on a universal set of morals, ethical principles, and/or laws that would satisfy everyone, but this, like many other technical approaches is worth exploring.

From individual researchers and research institutions to professional task forces like the

One Hundred Year Study on Artificial Intelligence (AI100, 2021) and academic reports like the

Stanford AI Index (annual), to technical research published in academics and popular journals,

we see that the topic of how to keep a firm rein on AI's dark side while continuing to innovate

technically is of great importance to computer science professionals. It's just not that easy.

Ethical and Legal Approaches

Aside from technical fixes to technical problems, the other main approach to ensuring

robust and benevolent AI is to draw on the fields of ethics and the law to monitor the hallway, so

to speak. Out of the gate, we find a plethora of statements of ethical principles, produced

primarily by the companies that make AI and the entities they sell it to. Fox/henhouse analogy

aside, principles only go so far when there are no binding accountability mechanisms in place to

ensure compliance. Since laws usually *follow* transgressions rather precede them (e.g., until self-

driving cars were viable, few people worried about who or what would be held legally liable

when they crashed) and AI transgressions are only recently coming to the attention of the general

public (e.g., until digital misinformation was believed to play a leading role in recent elections

and a global pandemic, few people worried about bots and dolts on social media), regulations

have seemed too slow – and sometimes too late – in coming. This reality is changing. Even now,

several states in the U.S. are writing laws to mitigate AI's perceived dangers (NCLS, 2022) but

in the meantime, the need for AI accountability continues, regardless of the number of statements

of principles, best practice lists, recommended guidelines, and the like, put out by corporations,

institutions, associations, and even governments. Some see these efforts as an earnest response to

shared concerns. Others see them as a performative response to political pressure. This speaks to

the suspicion that there is inherent motivation for certain entities to avoid scrutiny and external

oversight by getting ahead of regulations that would hold them more accountable and inhibit the kinds of technical innovations that lead to pole position in the first-to-market/first-to-publish race and ultimately, market dominance. Nevertheless, these entities make an effort to show their concern about the potential dark side of their technologies, even if they stop short of hard-coding accountability into their statements of principles, and these efforts comprise the most common form of ethics literature to date.

To understand the topography and terrain of AI ethics literature, more than one researcher/organization has attempted to present what I'll call a *snapshot of intent*, collecting as many of these statements into a central document or database as possible. For example, AI and Faith has curated an extensive, if not exhaustive, list of these statements and guidelines into a software platform called The Brain (link in References). Another example can be found in a paper titled *The Ethics of AI Ethics: An Evaluation of Guidelines* (Hagendorff, 2020), where the author gives what he calls a "detailed overview of the field of AI ethics." These are but two so-called "lists of lists" assembled to paint a picture of the AI ethics landscape. In reviewing these documents, readers will note they are remarkably like each other and focus largely on commonly acknowledged problems of AI. Whether these lists reflect "deep consensus about what is important, arrived at independently by numerous different actors, or merely a shallow consensus due to the fact that different groups have read similar papers and built on the work of one another" (Whittlestone et al, 2019, p. 196) remains unclear. Perhaps more importantly, whether these lists make any difference – or were even meant to make any difference – in true AI accountability remains unclear as well. Some think not. Writes Hagendorff:

> The current AI boom is accompanied by constant calls for applied ethics, which are
> meant to harness the 'disruptive' potentials of new AI technologies. As a result, a whole

body of ethical guidelines has been developed in recent years… However, the critical

question arises: Do those ethical guidelines have an actual impact on human decision-

making in the field of AI and machine learning? The short answer is: No, most often

not… AI ethics – or ethics in general – lacks mechanisms to reinforce its own normative

claims… these mechanisms are rather weak and pose no eminent threat… When

companies or research institutes formulate their own ethical guidelines, regularly

incorporate ethical considerations into their public relations work, or adopt ethically

motivated 'self-commitments,' efforts to create a truly binding legal framework are

continuously discouraged." (Hagendorff, 2020, p. 99-100)

Another facet of the ethics or legal approach is the proliferation of non-profit "watchdog"

organizations and institutions that monitor current trends in technology and lead the way in

publishing critiques and proposing solutions (for a sample list of these, See Appendix B,

Guidelines) for AI ethics and accountability. Again, these highlight the problems with AI and

conclude with calls to action and lists of recommendations, but they are not legally enforceable.

At least not yet. On the legal side, we have begun to see some wide-ranging frameworks that are

both already implemented and as mentioned above, on the way to being implemented. Among

those in place now are Europe's General Data Protection Regulation (GDPR) which was

instituted in 2018 and laid down rules and rights for how personal data is processed and shared in

the EU. The GDPR has an interesting history in Europe, largely due to an understanding – and

fear – of how governments (particularly how the Nazi regime, and later the East German Stasi)

could collect and misuse information to the detriment of their citizens, leading to a series of data

protection acts that predate GDPR but lay the foundation for privacy and self-determination

regarding personal information. In the U.S., the state of California has been proactive in this

arena with its California Consumer Privacy Act (CCPA), also passed in 2018, which says

citizens have the right to know how personal information is collected, used, and shared; the right

to delete personal information that has been collected; the right to opt out of the sale of their

personal information; and the right to non-discrimination for exercising those rights. As noted

above, several other states are following suit. These are but examples of the kinds of solutions

currently being presented to address big issues with AI in the ethical and legal domains in the

hopes that everyone from governments and corporations to individual private users will, to put it

succinctly, behave.

To reiterate, both the technical approaches and the ethical and legal approaches to Ethical

AI are predicated primarily on a secular humanist worldview. The viewpoint is embedded, so to

cite literature on a so-called materialistic approach to Ethical AI would be to cite literature on the

existing technical and ethical approaches (which I've attempted to do above) and therefore,

would be redundant.

A Christian Approach

There is yet a third approach to Ethical AI, and it is the one that interests me for the

purpose of this study: the Christian approach to moral or ethical behavior. I conceptualize this

approach not as a replacement for technical or ethical/legal approaches, but rather a foundation

for them and begin by operationalizing the idea of Christian ethics in contrast to other ethical

frameworks. Professor Wayne Grudem, theologian, general editor of the ESV Study Bible, and

author of the textbook *Christian Ethics: An Introduction to Biblical Moral Reasoning* (Grudem,

2018), defines Christian ethics quite simply as "Any study that answers the question, 'What does

the whole Bible teach us about which acts, attitudes, and personal character traits receive God's

approval, and which do not.'" (p. 37, adapted from Frame, 2008, p. 10) This definition

distinguishes God as the ultimate arbiter of right and wrong, and a Christian ethical system as different from other ethical systems by nature of its appeal to the whole Bible as the ultimate source of ethical wisdom. It is important to note that because the system of Christian ethics is derived from the whole Bible, it will necessarily include certain aspects of deontological, teleological, and virtue ethics, as well as historical, philosophical, and Jewish traditions, but as Grudem expands,

> A Christian approach to ethics will also exercise caution about adopting conclusions from the secular versions of these ethical systems, because all secular systems assume that ethical principles must be developed by human beings and using only human observation, reasoning, and intuition, whereas a Christian approach believes that the Bible's ethical teachings are not merely a result of human thinking but have been revealed by God himself… A Christian ethical system will not adopt moral relativism, for the Bible does teach that there is absolute right and wrong as defined by God himself. (Ibid, p. 43)

To allay any conspiracy theory fears, bringing a Christian perspective to the broader AI ethics conversation is not an attempt at a sort of ethical theocracy. Crazy Christian political movements aside, true followers of Jesus believe in free will and have faith in the persuasive nature of God rather than a governmental top-down approach. They also believe, however, that the claims of Jesus are true and that we disobey God at our own peril. Therein, a tension for some, a paradox for others, and nonsense for the rest.

Understanding a Christian ethical approach requires that we begin upstream and explore what we might call the genesis of ethical norms and behaviors, or the provenance of ethics. Provenance here denotes "the beginning of something's existence or its origins" and acts as "a record of ownership… used as a guide to authenticity or quality." (Oxford English Dictionary

online) Provenance is important in any discipline where an historical record of origin, ownership, and process is essential to determine value or resolve disputes and has long been used in the art world to trace the heritage of an antique or work of art. More recently, the search for data provenance has become an increasingly valuable practice within data science and machine learning communities since they depend on large datasets to make predictions and train algorithms. Because of several highly publicized AI "mistakes" (Google's errant facial recognition of dark-skinned women being a prominent example), calls for greater scrutiny and transparency around data have increased. In their paper *Datasheets for Datasets* (Gebru, et al, 2018; 2021), AI ethics researchers explore the problem of dataset documentation and argue that,

> …although data provenance has been studied extensively in the databases community, it is rarely discussed in the machine learning community… In the electronics industry, every component… is accompanied with a datasheet describing its operating characteristics, test results, recommended usage and other information… We propose that every dataset be accompanied with a datasheet that documents its motivation, composition, collection process, recommended uses, and so on [as they] have the potential to increase transparency and accountability within the machine learning community… (p. 2)

Like an electronics piece or part, and like machine learning data, ethics also has a birthplace, and an investigation of our moral origins, yielding a sort of *datasheet for ethics* from at least one of the world's great religions, could be helpful in grounding subsequent principles, laws, and regulations in what most people in the world consider the ultimate reality, regardless of the prevailing materialist sentiments in Silicon Valley and other centers of science and industry.

Ironically, many who eschew a spiritual reality embrace their own worldview or belief system

with a sort of religious fervor. Linda Kinstler writes,

> Tech is a stereotypically secular industry in which traditional belief systems are regarded
>
> as things to keep hidden away at all costs…Which is not to say that religion is not amply
>
> present in the tech industry. Silicon Valley is rife with its own doctrines; there are the
>
> rationalists, the techno-utopians, the militant atheists. Many technologists seem to prefer
>
> to consecrate their own religions rather than ascribe to the old ones, discarding thousands
>
> of years of humanistic reasoning and debate along the way. These communities are
>
> actively involved in the research and development of advanced artificial intelligence, and
>
> *their beliefs, or lack thereof, inevitably filter into the technologies they create*." (Kinstler,
>
> 2021, paragraphs 10, 11, 12, emphasis mine)

This is a key point, and one to which I will return in the Discussion section of this paper,

but it is important to note that Kinstler recognizes a critical truth: everyone has religion. In other

words, everyone has faith in something.

But what of the literature in this arena specifically as it relates to the ethics of artificial

intelligence? Among Christian thinkers, writers, and scholars (many of whom are, themselves,

computer scientists and AI professionals), there is much discussion around getting the

provenance of ethics right. For example, John Lennox asserts:

> Human morality was originally defined in terms of obedience or disobedience to
>
> the word of God. It only has significance insofar as the humans could understand
>
> what God said and had the capacity to choose either to obey or disobey. This, if
>
> true, as I believe it to be, is of crucial importance for the ethical evaluation of AI
>
> in the contemporary world. Genesis here is claiming that ethics is not relativistic,

nor did it evolve horizontally through social evolutionary processes… but it was transcendent in its origin… Moral convictions are, therefore, to a certain extent hardwired. (Lennox, 2021, p. 141)

If, as Lennox argues, we are moral beings by design, and as Kinstler and others contend, everyone has some form of religious belief, we would be remiss to ignore overt claims of the divine provenance of morality and only explore the opaque claims of evolutionary morality or moral relativity in an examination of Ethical AI.

There is also, among Christian thinkers, writers, and scholars, a focus on spiritual and moral formation that lies at the heart of ethics. The contention is that moral behavior does not come out of a vacuum and that so-called ethical people are built not born and must be taught right from wrong. This thinking aligns with the virtue ethics tradition and is not uniquely Christian, as secular sources also recognize both the need for and the reality of moral development (e.g., Kohlberg, 1958), but the Christian Bible has much to say about ethical training, from Old Testament to New. For example, immediately after the Hebrew Shema (the great commandment to "love the Lord your God with all your heart and with all your soul and with all your might") in the book of Deuteronomy, God's people are commanded to instruct their children in God's laws and precepts:

And these words that I command you today shall be on your heart. You shall teach them diligently to your children and shall talk of them when you sit in your house, and when you walk by the way, and when you lie down, and when you rise. You shall bind them as a sign on your hand, and they shall be as frontlets between your eyes. You shall write them on the doorposts of your house and on your gates. (ESV Bible, Deuteronomy 6:6-9)

In the Bible's wisdom literature, there is no shortage of aphorisms or proverbs on moral training but one of the most famous is, "Train up a child in the way he should go; even when he is old, he will not depart from it." (ESV Bible, Proverbs 22:6) In the New Testament, the Apostle Paul reminds both parents and children of the importance of moral behavior and moral education, saying, "Children, obey your parents in the Lord, for this is right… Fathers, do not provoke your children to anger, but bring them up in the discipline and instruction of the Lord." (ESV Bible, Ephesians 6:1-4) Perhaps the most sweeping biblical overview of the link between anchoring our cognitive faculties of knowledge and wisdom in God as a prerequisite for moral and ethical formation comes in the second chapter of Proverbs:

> My son, if you receive my words and treasure up my commandments with you,
>
> making your ear attentive to wisdom and inclining your heart to understanding;
>
> yes, if you call out for insight and raise your voice for understanding, if you seek
>
> it like silver and search for it as for hidden treasures, then you will understand the
>
> fear of the LORD and find the knowledge of God. For the LORD gives wisdom;
>
> from his mouth come knowledge and understanding; he stores up sound wisdom
>
> for the upright; he is a shield to those who walk in integrity, guarding the paths of
>
> justice and watching over the way of his saints. Then you will understand
>
> righteousness and justice and equity, every good path; for wisdom will come into
>
> your heart, and knowledge will be pleasant to your soul; discretion will watch
>
> over you, understanding will guard you, delivering you from the way of evil, from
>
> men of perverted speech, who forsake the paths of uprightness to walk in the
>
> ways of darkness, who rejoice in doing evil and delight in the perverseness of

evil, men whose paths are crooked, and who are devious in their ways. (ESV

Bible, Proverbs 2:1-15)

Regardless of overt efforts in moral training and education, religious or otherwise, we are

constantly being shaped – or misshaped – by the worldviews and moral assumptions we live with

even if they opaque to us. For example, in his book *The Abolition of Man* (Lewis, 1944), C.S.

Lewis depicts a relativistic culture that denies an absolute truth and transcendent morality yet

demands consistent ethical or moral behavior. Using two teachers, an English schoolboy, and a

particular English prep text, Lewis lays out the narrative of unconscious indoctrination:

> …they are dealing with a boy who thinks he is 'doing' his 'English prep' and has no
>
> notion that ethics, theology, and politics are all at stake. It is not a theory they put into his
>
> mind, but an assumption, which ten years hence, its origin forgotten and its presence
>
> unconscious, will condition him to take one side in a controversy which he has never
>
> recognized as a controversy at all. (p. 20)

This practice is still prevalent in public education today, from K-12 through university,

and while it is wise to avoid picking a side, religiously speaking, in a pluralistic society that

values free public education, it is unwise to simply operate on the assumption that because we

can't agree on which God is the transcendent source of morality, we should throw all of them out

without thoughtful comparative assessment. Perhaps it would be better to bring religion of all

kinds into the marketplace of ideas and at least allow the different versions of transcendent

morality to be discussed. Absent the exploration of transcendent and timeless principles and

teaching, we default to relativistic and shifting moral structures based on whims and cultural

tides, all culminating in a sort of moral anarchy about which we remain profoundly puzzled.

Lewis drives this point home in one of his most famous passages from *The Abolition of Man*:

It still remains true that no justification of virtue will enable a man to be virtuous.

Without the aid of trained emotions, the intellect is powerless against the animal

organism… The head rules the belly through the chest… The Chest –

Magnanimity – Sentiment – these are the indispensable liaison officers between

cerebral man and visceral man… The operation of The Green Book and its kind is

to produce what may be called Men without Chests… It is not excess of thought

that marks them out. Their heads are no bigger than the ordinary: it is the atrophy

of the chest beneath that makes them seem so. And all the time… we continue to

clamour for those very qualities we are rendering impossible… In a sort of

ghastly simplicity, we remove the organ and demand the function. We make men

without chests and expect of them virtue and enterprise. We laugh at honour and

are shocked to find traitors in our midst. We castrate and bid the geldings be

fruitful. (pp. 35-37)

Finally, I will anchor the formative literature in the arena of the Christian approach to

Ethical AI on a concept which is at the epicenter of Christian ethics: love. Love is a subject that

is difficult to study, scientifically, because of its many psychological, social, emotional, and

spiritual valences. Abraham Maslow, famous for his hierarchy of needs, says, "It is amazing how

little the empirical sciences have to offer on the subject of love." (Maslow, 1953 in Sorokin,

1954, p. xii) Among a group of social scientists who sought to remedy that from an academic

perspective, was Dr. Pitirim Sorokin, a sociologist originally from Russia, who helped establish

Harvard University's first sociology department, founded the Harvard Research Center in

Creative Altruism, and spent the latter part of his career studying the role of altruism and

unconditional love in the moral transformation of society. In his work, Sorokin compared the

concepts of eros and agape love (See Figure 1) and suggested that agape love and the so-called

creative power of love energy is the way to move "from tribal egoism to universal altruism."

(Sorokin, p. 459)

Figure 1

Eros and Agape Love

EROS	AGAPE
Eros is a desire of good for the self.	Agape is self-giving.
Eros is man's effort to ascend.	Agape comes down from above (from God).
Eros is man's way to God	Agape is God's way to man.
Eros is man's achievement, the endeavor of man to achieve salvation.	Agape is a free gift, a salvation which is the work of Divine Love.
Eros is egocentric love, a form of self-assertion of the highest, noblest, and sublimest kind.	Agape is unselfish love, which "seeketh not its own," and freely spends itself.
Eros seeks to gain life divine, immortal.	Agape lives by God's life, and therefore dares to "lose it."
Eros is a will to have and to possess, resting on a sense of need.	Agape freely gives and spends, for it rests on God's own richness and fullness.
Eros is primarily human love, and God is the object of Eros.	Agape is primarily God's own love, for God is Agape.
Eros is given only to those who deserve it; hence it is not spontaneous, but "caused" by the value of the object	Agape is poured out on "the evil and the good;" hence it is spontaneous and "uncaused," and bestows itself on those who are worthy and not worthy of it.
Eros recognizes value in its object, and therefore loves it.	Agape loves and creates value in its object.

From The Ways and Power of Love: Types, Factors, and Techniques of Moral Transformation, Pitirim Sorokin

Figure 1 - Eros and Agape Love

Much more about love as the heart of morality and goodness can be found in the

literature of Bible. As noted above, the foundational commandment to Israel is to love. "Hear O

Israel, the Lord our God the Lord is one. You shall love the Lord your God with all your heart

and with all your soul and with all your might." (ESV Bible, Deuteronomy 6:4-5) Jesus repeats

this as the greatest commandment in the New Testament but takes it a step further. "And one of

them, a lawyer, asked him a question to test him. 'Teacher, which is the great commandment in

the Law?' And he said to him, 'You shall love the Lord your God with all your heart and with all

your soul and with all your mind. This is the great and first commandment. And a second is like

it: You shall love your neighbor as yourself. On these two commandments depend all the Law

and the Prophets.'" (ESV Bible, Matthew 22:35-40) Jesus had put it more succinctly to a larger crowd before. In the verse that is known, nearly universally, as The Golden Rule, he says, "So whatever you wish that others would do to you, do also to them, for this is the Law and the Prophets." (ESV Bible, Matthew 7:12). As far as the Law formed the basis for Jewish ethics and morality, love fulfilled and then formed the basis for Christian ethics and morality. In his famous Sermon on the Mount, Jesus explains how:

> Do not think that I have come to abolish the Law or the Prophets; I have not come
> to abolish them but to fulfill them. For truly, I say to you, until heaven and earth
> pass away, not an iota, not a dot, will pass from the Law until all is accomplished.
> Therefore, whoever relaxes one of the least of these commandments and teaches
> others to do the same will be called least in the kingdom of heaven, but whoever
> does them and teaches them will be called great in the kingdom of heaven. For I
> tell you, unless your righteousness exceeds that of the scribes and Pharisees, you
> will never enter the kingdom of heaven. (ESV Bible, Matthew 5:17-20)

He then continues in Matthew 5 to give specific examples of how love fulfills and surpasses the law on such topics as anger (vv. 21-22), unforgiveness (vv. 23-26), lust (vv. 27-28), sin (vv. 29-30), divorce (vv. 31-32), oaths (vv. 33-37), retaliation (vv. 38-39), and giving (vv. 41-42). He ends the whole section with the most powerful and difficult commandment calling his followers to take love a step further and love their enemies:

> You have heard that it was said, 'You shall love your neighbor and hate your
> enemy.' But I say to you, love your enemies and pray for those who persecute
> you, so that you may be sons of your Father who is in heaven. For he makes his
> sun rise on the evil and on the good and sends rain on the just and on the unjust.

For if you love those who love you, what reward do you have? Do not even the

tax collectors do the same? And if you greet only your brothers, what more are

you doing than others? Do not even the Gentiles do the same? You therefore must

be perfect, as your heavenly Father is perfect. (ESV Bible, Matthew 5:43-48)

It is this counter-intuitive approach to morality, this upside-down mandate from Jesus

that differentiates Christian ethics from other systems: "…if anyone would come after me, let

him deny himself and take up his cross and follow me. For whoever would save his life will lose

it, but whoever loses his life for my sake will find it." (ESV Bible, Matthew 16:24-25) Christians

are commanded to behave better than the law and give more of themselves not just because God

will repay each person for what he has done and their reward is in heaven, but because they are

to model the character of Christ,

…who, though he was in the form of God, did not count equality with God a thing to be

grasped, but emptied himself, by taking the form of a servant, being born in the likeness

of men. And being found in human form, he humbled himself by becoming obedient to

the point of death, even death on a cross. (ESV Bible, Philippians 2:6-8)

In other Christian literature, authors have asserted that divine love, the agape love that

goes beyond mere justice and fairness, is essential to activate true moral capacity and the heart of

ethical behavior. For example, in his book, *Don't Waste Your Sorrows* (Billheimer, 1977), Paul

Billheimer argues that morality is inextricably linked to love:

What is meant by a universe that is moral? It is one in which the Law of Love is supreme,

because love is the fulfilling of the law. It fulfills every obligation to every intelligence in

the universe… The most fundamental characteristic of an order that is moral, therefore, is

agape love… the love which characterizes God himself. Agape love is the love which

loves because of its own inherent nature, not because of the excellence or worth of its

object… it is spontaneous, automatic love… Agape love is not primarily an emotion but

aggressive, benevolent, sacrificial, outgoing goodwill. It is the soul of ethics. (pp. 31-32)

So where do we stand today, calling for Ethical AI in 2022, and what does the Christian

faith have to say about it? To date, two prominent Christian denominations have produced their

own versions of the secular world's statement of principles to address the ethical issues around

AI from the position of religious faith. The Catholic Church has released *The Rome Call for AI

Ethics* (Rome Call, 2021) and the Southern Baptist Convention's Ethics and Religious Liberty

Commission has released *Artificial Intelligence: An Evangelical Statement of Principles* (ERLC,

2019). While neither document is highly technical, and each takes a different approach, together,

they form a base from which to expand formal research literature on AI ethics and religion. The

Rome Call largely mirrors corporate ethical principles and talks about the concept of

"algorethics" and a "RenAIssance" in human-centric AI. The ERLC Statement by contrast

begins with the phrase, "As followers of Christ…" and proceeds to highlight areas of AI concern

in the form of affirmations and denials, with underlying scriptural support. Both documents

represent a point of departure for the Christian church to weigh in on the technological, cultural,

ethical, and spiritual importance of artificial intelligence.

While some may argue that ancient religions are irrelevant to modern technologies, we

cannot ignore the fact that ancient religions remain stubbornly relevant to a large majority of

modern humans, and since technologies, especially AI, have been proposed as extensions of

human abilities (McLuhan, 1964; Postman, 1992; Lanier, 2010), the wisdom of the ages should

get a new hearing as we chart the uncertain path forward toward an AI future.

Methodology

In this study, I sought to bring viewpoint diversity to AI ethics by understanding and describing a position that has been marginalized in Ethical AI research: how Christian teaching, texts, and traditions might contribute to the evolving debate surrounding the ethics of artificial intelligence. The purpose of this research was to give voice to a perspective that is rarely heard in scientific discourse. I sought more to add voices to an area of study that lacks them, than to undermine or invalidate other voices. I conducted this research as a means to understand both the common ground *and* the gaps that can be found at the intersection of artificial intelligence, so-called secular ethics, and Christian theology, focusing less on the perils of artificial intelligence writ large, and more on the provenance of ethics and the efficacy of AI ethics without a transcendent motivation for love, the root of Christian ethical teaching.

I note here that while an overview of many religions could have been the basis for this study, I settled on one since trying to include all religions, or all major religions, or even a subset of, say, the monotheistic religions, would have been overwhelming and unruly for a doctoral dissertation. This interrogation of the subject through the lens of one religion is in no way intended to limit or preclude research on how other religions address the issue of ethics in AI. It is my hope that this study would prompt other voices of religious faith to join the conversation.

Research Design and Strategy

My study represents "basic qualitative research" (Merriam, 2009, p. 22) that foregrounds Christian voices in the Ethical AI conversation. For focus and feasibility, I concentrated on voices from the world's largest religion by both affiliation and practice, and one that has already actively entered the conversation with published statements on how humanity should view, engage with, develop, and use AI (ERLC Evangelical Statement of Principles, 2019; Rome Call

for AI Ethics, 2020). Because I was interested in the culture of Christians in high tech, hoping to tease out "the beliefs, values and attitudes that structure the behavior patterns of a specific group of people" (Merriam, 2016, p. 29), I situated my work very loosely within the genre of ethnographic research, but firmly in the descriptive tradition. While all qualitative studies view "the qualitative researcher [as] the primary instrument for data collection and analysis… mediated through this human instrument rather than through inventories, questionnaires, or machines (Creswell, 1994, p. 145), true ethnographic studies require "a participant observer" (Merriam, 2016, p. 29) for full environmental description. Due to a variety of factors that included a global pandemic and significant geographic distance, I relied on qualitative interviews alone as my means of data collection. Still, I sought to probe "the various ways [this group goes] about their lives and to the belief systems associated with that behavior" (Wolcott, 2008, p. 22) and I was "interested in process, meaning, and understanding gained through words" (Creswell, p. 145) so the work became culturally descriptive and therefore ethnographic in nature.

Data Collection

My primary data-collection instrument was a semi-structured question protocol (See Appendix C, *Protocol*) and I conducted and recorded interviews online via Zoom with a targeted group of informants. I emailed each one a short set of standardized questions to elicit context around their positionality in the world of AI and in their own Christian faith practice. The remainder of the questions were organized in a semi-structured format and flexible enough to focus on the informant's particular knowledge and expertise as it informed, and is informed by, their beliefs, opinions, and perceptions about AI ethics.

To avoid any barriers to discussion, I avoided using elicitation techniques that were overly abstract, broad, open-ended, or nondirective and therefore possibly ineffective in

uncovering the kind of data I really want to collect (Barton, 2015). The goal was to uncover and then demonstrate whether there was a contribution to be made from a Christian perspective on Ethical AI that was not already being made by those working in the secular ethics space.

I asked several general questions about the issues around AI, ethics, and faith, but it is here that I will report an adjustment in the protocol due to my use of the constant comparison method of concurrent data gathering and analysis (Merriam, 2016, p. 202): I had initially decided to frame my study around ethical tensions and trade-offs in AI that arise when there is no clear or easy solution, including problems are currently considered intractable, either technically or morally. For example, how can we build robust AI applications that depend on large amounts of user data and also protect the user's privacy? How can we deploy AI in place of human workers and still respect human dignity? These, and other tensions, formed the basis of my original protocol and my thought was that they might elicit deep thinking around how Christian technologists, ethicists and/or theologians might provide a framework that their secular peers had not considered. It turns out that my attempts to center the interviews on the tensions did not bear as much fruit as I had expected. I begin my Findings section with a more detailed explanation of this, but considering nearly all my informants, having initially attempted to address the issues in AI through the lens of ethical tensions, ended up steering the conversation on to issues they considered more relevant or were more conversant on, I decided to focus more specifically on my three main research questions and interrogate the broader topic of how Christian beliefs might drive the ethics conversation toward a more expansive vision of the anthropology and theology of technology.

Sample and Setting

For this dissertation, I chose to recruit a sample of twenty informants to participate in hour long qualitative interviews about the Christian faith, ethical principles, and artificial intelligence. I landed on the more aggressive word *informant* because of the level, depth, and sensitivity of data I was seeking. As Sharan Merriam says, "Anthropologists and sociologists speak of a good respondent as an 'informant' – one who understands the culture but is also able to reflect on it and articulate for the researcher what is going on." (Merriam, 2009, p. 107) It was important that my informants were neither technically illiterate but spiritually literate, nor the other way around. They had to be both technical professionals (in some form or fashion) and professing Christians. Therefore, I drew on a *purposeful* sample (Merriam, 2009, p. 78) of informants who embodied both qualities so I was able to gather data from *information-rich* sources (Patton, 2002, p. 237) rather than a random sample that may include people who have either no expertise in AI or no religious faith. (See Figure 2)

Figure 2

Informants

	PROFESSION	FAITH EXPRESSION
Narcisa	CS and Theology Professor	Quaker, Former Lutheran
Robert	ML Researcher, Large Tech Company	Pentecostal
Timothy	HRI Researcher, CS Professor	Nominal United Church of Christ
John	AI Pioneer, Author, Apologist	Evangelical Independent Church
Nicolas	Philosophy Professor, Tech Ethics	"Homeless" Christian
Christian	AI Start-up Founder, Ex-Big Tech	Evangelical Presbyterian
Christopher	CS Professor	Catholic
Mark	Dean of CS & Engineering	Catholic
Dominic	Philosophy Professor, Dir Tech Ethics Ct.	Catholic
Joseph	Chair of Research, Tech Ethics, Non-Profit	Southern Baptist
Michael	ML Scientist, Biotech Company	Christian Missionary Alliance
Peter	Pastor, Former Software Engineer	Non-denominational Christian
Joan	AI Start-up Founder, Ex-Big Tech	Evangelical Charismatic
Edmund	Data Scientist, Large Bank	"Fringe" Episcopalian
Benedict	Director of Tech Ethics Center, University	Catholic
Jonas	Theology Professor, CS background	Catholic
David	CS Professor, Author	Christian Reformed
Jerome	Mathematics Professor, Author, Apologist	Non-denominational Christian
Anthony	Director, Faith and Tech Resource Center	Evangelical Anglican
Martin	CS Professor	Baptist
James	Doctor, Author, AI/Robotics Researcher	Non-denominational Christian

List of study informants, current occupation, and Christian faith expression/denomination. Names have been changed for privacy.

Figure 2 - List of Informants

I reached out to my informants personally, beginning with a cohort from the founding members of AI and Faith (See Appendix D, *AI and Faith*), of which I am a board member and research fellow, to be part of my study. From there, I built a *snowball*, *chain*, or *network* sample (Patton, 2002, p. 237) expanding my selection pool of informants through referrals by my initial group of contacts. I offered two reasons for this. First, some on my initial list were unavailable but willing to recommend someone they knew that I did not. Second, I supposed that some informants I identified may have been unwilling or unable to provide the kind of data I wish to gather due to fears of being open about their faith in a workplace culture known for its skepticism of religious beliefs in general and, more specifically for its hostility toward Christian believers. I ended up with twenty-one informants none of whom, whether they were actively working in high tech or not, was reluctant to share insights and only one on my list referred me to a colleague due to her own unavailability. The other informant who was not on my original list (and became my twenty-first interview) was a recommendation that came during an interview, where my informant was so taken by what I was working on that he said that I must speak with this person he knew who would add value to my research. I readily agreed, because I had just read one of his books and now had a personal introduction from a friend and colleague.

In the end, none of my informants had any barriers (real or perceived) that would preclude them from talking about their Christian beliefs in the context of AI and the ethical issues that surround it. My plan has always been to keep the identities of my informants private in the write up of this dissertation to ensure there were no self-imposed roadblocks to honest, authentic responses from the start. To that end, I have changed the names of the informants for this study. The attentive reader will note they are all names of Catholic saints.

As to the make-up of my sample, my hope was to gather a cohort that was ecumenical but not exhaustive in nature. While Christianity is a faith with many different doctrinal and denominational expressions, I sought neither the broadest sample that might include every denomination or expression of the Christian faith nor the narrowest sample that would include only one. Rather, I looked for a *typical* sample of tech workers from a variety of denominations and faith expressions that "were not in any major way atypical, extreme, deviant or unusual." (Patton, 2002, p. 236) As it turned out, I did not know the particular denominational or theological identity of any but a few of the people I contacted to participate in this study and ended up with a much broader spectrum of Christian belief and practice than I anticipated. What's more, the sample turned out to include a wide range of entry points to the Christian faith. Some had been raised in the faith by Christian parents while some had come to faith on their own. Some had come to follow Jesus as young people, and some much later in life. Some had been lifelong congregants of a particular denomination, and some had changed denominations as their faith changed or grew. Some were solid in their doctrinal and denominational beliefs, and some were spiritually wandering and seeking a church home, so to speak. This diversity added a richness to my study that I had not expected, as the Findings section will bear out, but had the reciprocal effect, due to the varying levels of experience with Jesus, of causing some of them to think more deeply about their own opinions regarding AI ethics and Christian practice because of the questions I posed.

As for the field work setting, since my informants were spread across the U.S. and Canada, as well as the UK, and at the time COVID restrictions were still in place, I conducted and recorded the interviews in my home office on the Zoom platform. Had I done this study any sooner, I might have struggled more to make this approach happen, but since March of 2020, and

through more than a year of working from home, people got used to using online applications to meet, learn, teach, socialize, do business, and conduct interviews. The online interview option afforded a kind of face-to-face intimacy that was better than a simple voice interview (which would have been my other option given the pandemic circumstances and my budget) as I had the benefit of seeing facial expressions and body language during conversations, although perhaps not as rich an experience as talking in person. In all, Zoom turned out to be not only an effective mode for qualitative interviews, but also an inexpensive way to gather data from geographically diverse participants.

To set up my interviews, after receiving IRB approval, I reached out to each informant via email. I re-introduced myself (most on my list were acquainted with me in some way because of my role at AI and Faith) and asked if they would be willing to participate in a qualitative study on how the Christian faith might inform the conversation around Ethical AI. I included an executive summary of my project with an overview of the methodology I planned to use and told them that I'd conduct and record the interviews online but that their identities would be kept confidential in the write up and that all data gathered would be for the purpose of my dissertation only. I ended up with a nearly 100% participation rate.

During the interviews, I took field notes as the conversations progressed even though the interviews were being recorded, both to remember new questions I wanted to ask while the informant was answering another one, and to make special notes about what I was hearing at the time, even though I knew I would have a verbal recording and written transcript to analyze later. This method had been recommended for qualitative observers to "stimulate critical thinking about what you see and become more than a recording machine" (Bogdan and Biklen, 2011, p. 163) but I found it helpful, in a metacognitive way, as I cultivated novel propositions on-the-fly.

At the same time, I engaged in "axial coding" (Charmaz, 2014), aka, analytic coding, while conducting my interviews to supplement the process in the data analysis phase, stimulating "interpretation and reflection on meaning" (Richards, 2015, p.135) during data collection.

Data Analysis

As qualitative research expert Sharan Merriam asserts, "Data analysis is one of the few aspects of doing qualitative research – perhaps the only one – in which there is a preferred way… the much-preferred way to analyze data in a qualitative study is to do it simultaneously with data collection." (Merriam, 2016, p. 197) Choosing the "more enlightened scenario" (Merriam, 2016, p. 196) of sitting down to one transcript, fresh from an interview, and beginning the work of data analysis during the work of data collection is now the method recommended by seasoned qualitative researchers. To put a finer point on it, Merriam and other qualitative research experts agree that "qualitative data analysis is primarily *inductive* and *comparative*" (Merriam, 2016, p. 201) and using the constant comparative method of data analysis is an effective way to generate findings. (Merriam, 2016, p. 202)

As for process, I headed into the interviews with a rough list of the themes I expected I might find. Then, as I conducted interviews over the course of two months, I transcribed and reviewed them immediately after they were complete, so I was able to judge and/or adjust, 1) whether my questions were effective in eliciting the kinds of responses I'd hoped for (and if not, re-thinking some of the questions for the next interviews), and 2) whether the categories I drafted were, in fact, representative of thinking laid out in my research problem and theoretical framework (and if not, adjusting or adding categories that reflect the actual data). It turns out that my initial list only partially overlapped the codes I landed on while analyzing the data. This is partly because I set out to frame the interviews around the tensions previously mentioned, and

ended up not doing that, but also because my informants provided a richness and depth of thought around Ethical AI and Christian teaching that went beyond my expectations. This did not cause me to abandon my thesis but rather broadened and expanded my thinking around it.

As the interviews progressed, I engaged in the constant comparison method of qualitative analysis, interrogating the data against both my assumptions going in and my findings along the way. As I mentioned before, this was an incredibly wise approach as I did find that I had headed down a road that made less sense to my informants than I had anticipated. To be honest, the process of constant comparison helped clarify and solidify the real question I had, which was whether and how Christianity made any difference – or shed any new light – on Ethical AI. The study was better for it.

As for data management, again, the pandemic inadvertently paved the technical runway with affordances I would not have had even two years ago. Online meeting and recording platforms blossomed because they had to, but other innovations made life easier as well. Specifically, online, AI-assisted transcription programs (irony is not dead after all) allowed an unemployed doctoral student to avoid hiring an expensive transcription service to transcribe audio files to written files or engage in the time-consuming process of doing the transcriptions manually. My process was as follows: record the interview on Zoom, download the audio file to my hard drive, upload the audio file into Descript (a podcast platform with AI transcription capabilities), copy-paste the AI transcription into a Word document, edit the document for accuracy by listening to the audio file while reading/editing the AI transcript on the fly, and finally, printing out the interview for paper review and manual coding. As a natural underliner, a highlighter, and a margin writer, I find tangible paper transcripts help me to connect physically with the work. Then, I moved to the computer to manipulate or move the data around, via digital

cut and paste, into coded files and folders, ready – as supported by other interviews and analysis – to be included in the Findings section of my dissertation. As Merriam advises, "The important task is to create an inventory of your entire data set… This data set needs to be organized and labeled according to some organizing scheme that makes sense to you, the research, and enables you to access any piece of your data at any time." (Merriam, 2016, p. 200)

Finally, as the interviews progressed, I again took Merriam's advice. She recommends that qualitative researchers "move between seeing the big picture (the forest) and the particulars (the trees)" (p. 207) to keep two things front of mind. One, which represents the forest, was to remember the purpose of my study, the framework in which I situated it, and the questions I wanted to answer. The other, which represents the trees, was to look critically at the individual pieces of data I was collecting, the insights I was gathering from my informants, and the meaning I was making as I went. How did the two intersect and interact? What biases did I bring into the study and how were they being supported or challenged? How did the forest inform the trees and the trees the forest? Were my findings pertinent to the literature? Could I generalize anything to my theoretical framework?

Once I reached the end of my list of informants and concluded all the interviews, I found that I did reach a sort of saturation point where "no new information, insights, or understandings are forthcoming" and I moved into "thinking in a more deductive rather than inductive mode… testing [my] tentative category scheme against the data." (Merriam, 2016, p. 210). Where at the beginning of my analysis, I was in the inductive mode, I moved into "hard coding" my categories to reflect a synthesis of what I hypothesized, what the literature said, and how my informants' responses confirmed or confounded either. The final picture gave me a vantage point from which to write up the results.

Concept Map of Contribution

My motivation in bringing the Christian worldview into the Ethical AI conversation was to explore and present a perspective. My intent was not to prove a particular point of view, replace another point of view, nor argue that the findings would generalize to a larger population, but rather offer, through the voices of my informants, a representative reflection of a worldview held by many and make a case that this worldview not only belongs in the marketplace of ethical ideas but established it. My hope is that the findings of this study will contribute to the field of AI ethics and add to the literature in the following ways:

- Highlight the differences between a prevailing materialist worldview and a Christian worldview, revealing how each informs the science and ethics of AI

- Deepen awareness of how faith in and fear of God could influence the way AI is designed, developed, and implemented

- Demonstrate how Christianity has already made a valuable contribution to AI ethics – whether it is acknowledged or not –and map a path to future conversations where Christian wisdom is included in every phase of AI development

- Present a more inspiring vision of the future that moves beyond a culture of Ethical AI and toward a culture of Righteous AI

Researcher Qualifications

While I am not an AI professional, I entered this study with some critical qualifications. First, I worked as a qualitative researcher at an educational software company for a year and learned the basic ins and outs of the software development cycle, including how large the gap between hope and hype can be when it comes to trying to deliver features the customer expects in an unrealistic timeframe. Although the core product of the company was a learning

management system (LMS) similar to many on the market today, the ultimate goal was to move into AI-driven personalized learning and individualized instruction, and the company was in early talks with a now-defunct Gates Foundation initiative called inBloom, which fell apart when parents and other concerned constituents realized that the AI version of a private tutor turned out to be quite a threat to student privacy. That job ended because the company ended. Subsequently, I underwent a sort of baptism by fire (no religious pun intended) during my three-year tenure as executive producer and host of the Microsoft Research Podcast, where I interviewed more than one hundred researchers, many of whom were dedicated to the science of artificial intelligence and machine learning. This job became a weekly crash-course on cutting edge research in AI and exposed me to the emergent field of AI ethics where researchers, who perceived AI's unique position in the technical world, were beginning to call out its mistakes, misrepresentations, and dangers. Finally, I have been a follower of Jesus since I was in junior high school, and immersed in the study of Christian texts, teachings, and traditions for more than fifty years qualifying me, from a spiritual experience perspective – and, if churches were unionized, a seniority perspective – to situate my study at the intersection of AI, ethics, and Christianity.

Findings

A note on change of focus

In my proposal to conduct research, I had decided to anchor my interrogation of the

Christian faith's impact on AI ethics around what many consider a standard set of ethical

tensions proposed in a paper titled *The Role and Limits of Principles in AI Ethics: Towards a*

Focus on Tensions (Whittlestone, et al, 2019, p. 196). I chose this approach because the authors

asserted, in strong alignment with my own research thesis, that "…while principles are

important, they are not in themselves enough to ensure society can reap the benefits and mitigate

the risks of new technologies.". They continued, "…even the strongest advocates of

principlism… acknowledge that principles alone are not enough." (Ibid, p. 196) They concluded

that "Reframing research questions to be more focused on understanding and resolving tensions

is an important step towards solving practical problems arising from the use of AI in society,

since it directs attention to where new technological or governance solutions might help push the

development of AI in robustly beneficial directions." (Ibid, pp. 200-201) The tensions they

identified were as follows:

1. Using data to improve the quality and efficiency of services vs. respecting privacy and autonomy of individuals. (Ibid, p. 199)
2. Using algorithms to make decision and predictions more accurate vs. ensuring fair and equal treatment. (Ibid, p. 199)
3. Reaping the benefits of increased personalization in the digital sphere vs. enhancing solidarity and citizenship. (Ibid, p. 199)
4. Using automation to make people's lives more convenient vs. promoting self-actualization and dignity. (Ibid, p. 199)

Given that understanding and resolving moral or ethical tensions has been a key value

proposition of Christianity for millennia, I thought exploring the religious perspective of Ethical

AI through the lens of widely accepted tensions would be a good starting point for my

investigation. And, given that even Microsoft President Brad Smith explicitly enlarged the tent saying, "Ultimately, a global conversation about ethical principles for artificial intelligence will require an even bigger tent. There will need to be seats at the table not only for technologists, governments, NGOs, and educators, but for philosophers *and representatives of the world's many religions.*" (Smith, 2019, p. 208, emphasis mine) I felt not only justified in pursuing this line of inquiry, but welcome to do so. With these "givens" in mind, I designed a protocol around the tensions articulated by Whittlestone, et al, and plunged into my qualitative interviews.

What I found was that this approach didn't produce the fruitful discussion I had expected. Some informants essentially ignored my questions about the tensions and steered the conversation to topics they thought more important. Others suggested the tensions were not particular to AI but reflected humanity writ large and that larger questions were looming behind the assertions. Apropos of this sentiment, one asked: "Is the AI biased or is the AI reflecting statistical realities in the 20th century as an aggregate? What is an accurate picture of the world? Is it the world as it functions or is it the world as it ought to be?" (Jonas) Still others argued that human nature and business realities were the driving force behind the fact that these tensions even existed. One said, "…there's only so much bias that we can remediate… [My Company] is a business, and so, it ultimately tries to make the most money while making people believe that it's the most fair." (Robert) Another argued, at length:

> I would push back… on these dichotomies. Like, I don't see robustness and privacy as
> being necessarily opposed to each other… I think these dichotomizations are often
> 'motivated reasoning' and they don't have to be, and this is why: because you *can*
> maintain privacy if you encrypt the data properly… The downside of it is that… it's going
> to cost more money in terms of compute and in terms of energy costs and those sorts of

things. So, when you are running your company, you say to yourself, oh, we could do the

nice privacy solution, or we could do the cheap solution and since we're a startup, what

are you going to choose? You're going to choose the cheap solution of course. There are

ways to solve these problems so that they're not dichotomies. (Benedict)

My early take-away was that the tensions were seen as less an issue of "*what-is-the-right-*

thing-to-do?" and more an issue of "*we-know-the-right-thing-to-do-but-don't-really-want-to-do-*

it." I don't want to diminish the reality of resource constraints, but once the first corner is cut and

nothing bad happens – or perhaps more seductively, an advantage appears (e.g., we get all this

valuable data for free, and customers don't know what they're trading for our application) – the

temptation to make unethical choices becomes less inhibiting. In fact, the temptation to make

unethical choices precedes AI by thousands of years. As Cain contemplated killing his brother,

God warned him, "If you do what is right, will you not be accepted? But if you do not do what is

right, sin is crouching at your door; it desires to have you, but you must rule over it." (ESV

Bible, Genesis 4:7) Later, St. James said, "If anyone, then, knows the good they ought to do and

doesn't do it, it is sin for them." (ESV Bible, James 4:17). From a Christian vantage point,

choosing the right path may not be easy but it's rarely a mystery. More plainly, when confronted

with a choice between the right thing and the expedient thing, it's easy to take our excuses for

wrong-doing and re-brand them more palatably as ethical dilemmas or tensions.

Ultimately, while the tensions around which I had framed my study were not entirely

irrelevant, I decided to focus on what my informants were foregrounding. What surfaced was a

nuanced and relevant discussion around my three main research questions: **1) How does**

worldview affect the way in which AI researchers and developers conceive of and make

artificial intelligence? 2) Does a Christian worldview have anything unique to contribute to

the discussion around ethical issues in AI? and 3) How might AI ethics be more robust and more beneficial if we brought Christian teachings, texts, and traditions explicitly into the conversation? My informants held strongly to the premise that what we believe about the ultimate reality of the world informs everything we do, and they shared their thoughts about this through a variety of topical lenses. Before I present their insights, I think it would be helpful to operationalize the major beliefs of two basic worldviews – religious and non-religious – that animated this discussion. I say two, acknowledging that between religious and non-religious worldviews, there are many permutations of each. For the purposes of this research, I have anchored my study around **materialism**, which I propose is the dominant worldview among computer scientists and technology workers, and **Christianity**, which I propose is underrepresented in computer science and tech circles. (See Figure 3) While this is clearly not an exhaustive chart, it lays out some basic claims of each worldview.

Figure 3

Worldview comparison chart.

	MATERIALISM	**CHRISTIANITY**
Philosophy	Naturalism, with faith in human reason alone.	Supernaturalism, with faith in God and in human reason.
God	Does not exist. The material world is all there is.	Does exist. Is loving, omnipotent, omniscient, and omnipresent.
Jesus	Was a historical person, considered by some a prophet and teacher.	Is the risen Son of God, savior, and redeemer of fallen humanity.
The Universe	Appeared ex nihilo through quantum fluctuations. Is evolving.	Was created ex nihilo by God. Is sustained by God.
Humans	Are the result of evolutionary processes.	Were created in love, by God, in his image.
Spiritual World	Is imaginary and therefore uninhabited.	Is real and is inhabited by spiritual beings, both good and evil.
Religion	Is a myth made up by humans to explain phenomena on earth.	Is the human belief and worship of God's spiritual reality on earth.
Evil	Is an unexplainable reality that disproves the existence of God.	Is a supernatural force that that seeks the destruction of mankind.
Morals and Ethics	Determined by humans. Transmutable, situational, evolving.	Determined by God. Transcendent, absolute, unchanging.
Dignity and Value	Defined by humans and achieved by human effort.	Defined by God and inherent in every human.
The Future	No life after death in either heaven or hell.	Eternal existence after death in either heaven or hell.

Figure 3 - Worldview Comparison

As noted in the literature review, our underlying belief systems – which encompass our understanding of what it means to be human, the nature of human intelligence, the difference between a brain and a mind, the human desire for transcendence, and the reality of spiritual evil – inform what we believe we can, should, and must do with artificial intelligence. I found it interesting that a presumed majority of high-tech workers hold a predominantly atheistic, materialistic worldview but much of the rest of the world does not. Recall the Pew Global Religious Landscape study reporting that 84% of the world's population identifies with a religious group or religious belief (Hackett, Grim, 2012, p.9), while according to a Lincoln Network survey, about half of tech workers identify as atheist or agnostic (Lincoln Network, 2018). Those who do not, tend to keep their religion to themselves because traditional religious belief systems – or more specifically, the moral and ethical frameworks they represent – are unpopular in predominantly progressive, atheistic environments. Even if everyone in high tech is not overtly hostile to, say, Christianity (though some certainly are), Christian values do not generally inform the underlying ethos of Silicon Valley (here, a synecdoche for the technology industry writ large). That reality formed the basis for my inquiry into the beliefs and opinions of people in the tech industry who were Christian believers. I note that while the three main research questions (in bold, above) were my primary concern, interviews varied according to individual informant enterprise, experience, and expertise. For example, the questions I asked of an HRI researcher whose work focused on robotics and health care were slightly different from those I asked of a machine learning researcher who focused on search optimization and were quite different from questions I asked of a pastor who used to be a software engineer or a professor of computer science and theology. Here, I present a curated selection of findings from each interview, along with brief descriptions of the informants and a recap of their contributions.

Interview Data

Narcisa is a professor at a private Catholic university where she teaches both computer science and theology courses. A former Lutheran, she is now a practicing Quaker. On the reality of AI, "I take a very strong view that singularity ain't coming and… we're not going to create an AGI (artificial general intelligence)." She expands on the theme, "A machine cannot be wise. You might say that to a certain extent a machine can be knowledgeable, maybe even smart… but it can't be wise… Wisdom is knowing when to make the exception… when to throw away the data and not to do what the algorithm says, and the machine can't have that."

On the purpose of AI, "I think AI is a desire to have something other than fellow human beings with which we can relate… goes back to the Christian understanding that we are created in the image of God to be relational beings… So, the desire to create an artificial intelligence in our own image makes sense. It's an extension of our relationality… We're trying to have an intimate relationship that should be had with another human being with a thing."

On the dangers of AI, "One of the problems with artificial intelligence is… it tends to separate off human intelligence as a qualia – a thing unto itself – that we want to perpetuate and enhance and what that tends to do is to separate us from the rest of nature, as well as from God." Expanding, "What the Christian tradition has to say, and say very loudly is, don't let this become a substitute for either God or one another. We have a tendency to make an idol out of our technology… sort of… a substitute for God… to… somehow help us overcome death, overcome the human condition… It's looking for a new salvation and new eschatology in this technology and it's making it a do-it-yourself project." More, "There's a worldview in the AI world and it's a worldview without a transcendence… If there is transcendence, it's that the computer will transcend us, and if there is eternal life, it's the humanist vision that we'll upload our brains or

something. And I think what matters here is that when we ground what we do in that vision, well, we lose humility. Let me put it that way. You know, we think that in a sense we make gods of ourselves." On the line between humility and hubris, "While I think that a lot of our desire to create an AGI is rooted in something good about us… our desire for relationality, what it becomes instead is… a sort of hubris of what human beings are and what we are capable of. And I think the religious traditions can help us get back a little more of a sense of humility, a little more of a sense of concern that things can and probably will go very wrong."

On AI in a fallen world, "Tools are always amplifiers, and so our tools… will amplify not just our capabilities to do good things, they will amplify human sinfulness… The power in our technology coupled with human sinfulness may mean that every technologically advanced society does itself in before it can leave the home planet… Our religious traditions have tried to counter that… To borrow an old Amish saying, 'We grow too soon old and too late smart.' We grow too soon technologically capable, and too late do we recognize what our religions have been screaming at us through the years: you need to overcome that inherent selfishness. You need to overcome that inherent competition. We need to love one another."

On the provenance of ethics, "Even those [students] that are not really well grounded in a religious tradition… still have the kind of a vague ethic that permeates our society, which I believe is… an ethic that grows out of the Christian tradition…" And on how Christian teachings could level up AI ethics, she says, "When Jesus is asked to sum up all the law, it's 'Love the Lord your God with all your heart, all your soul, all your mind, and your neighbor as yourself,'" and says, "Maybe intelligence is not our greatest virtue, you know? Maybe what we need to be enhancing is love, but what we are going for is enhancing intelligence. And when we separate

those two from each other, it's often when we tend to get into trouble. It's not the pathway. It never has been. Our pathway to divinity is through love. It's not through intelligence."

Narcisa's unique background in both technology and theology revealed itself in a practical, no-nonsense approach to AI tempered by a deeply Christian understanding of reality and relationship, humility and hubris, and the supremacy of love.

Robert is a machine learning researcher for a large, well-known tech company. He was raised as a "nominal" Christian in his Asian family but has embraced charismatic Christianity after an encounter with the Holy Spirit. On worldview diversity, he says, "So, fairness a lot of times, is grouped into categories. What is the category that you're trying to be fair in?... I'm trying to bring faith into the picture, like being fair to people of faith as well as people without faith… if we do discuss the correct notion of fairness… I think we need to bring in a lot of worldviews." On the reality of fairness and the promise of AI to mediate, "We realized that our world is in some sense unfair. And so, no matter how good the data you can get, it still reflects an unfair world… there's only so much bias that we can remediate… [My Company] is a business, and so, it ultimately tries to make the most money while making people believe that it's the most fair."

On purpose, or telos, "God's overall objective is for us to have a relationship with Him and to go back to him as much as possible for that relationship. Unfortunately, for [My Company], it's kind of the same… we want to have a relationship and we want you to come back to us, but it's mainly… in a way that generates profit."

On love and optimization, "Attention is what [My Company] considers love… How much do you like [My Company]? You know, how much do you approve of [My Company]? How much do you want to use [My Company] in the future? And in a sense, that is [My

Company's] definition of love. [My Company] will never optimize your love for another thing, so to speak, and God will not either." Expanding on the theme, he says, "A lot of technology is idolatry in the sense of the way we designed it… am I causing people to love other things more than God?"

On using technology and transcendence, "The way of transcendence in the Bible is like we're empowered by the Holy Spirit… God is helping us to accomplish these things that we can't do on our own… The Bible says in Proverbs 'do not be wise and in your own eyes but fear the Lord'… That's where the Christian view of transcendence really shines."

On the brain and AI, "I think we're still on that very primitive level in terms of understanding the brain and trying to replicate certain tasks but… I think even those basic tasks will glorify God, ultimately, because the more we understand how just complicated one single thing is, it only goes to show that the glory of God and how the glory of his creativity is etched into every part of our brain."

On how a Christian worldview might add value to AI ethics, "The claim of AI artificial intelligence in the secular world is… one day all your problems will be solved, and AI is the savior of the world… That would contradict what we as Christians believe, but I think it also contradicts what a lot of people believe now, even if you're not Christians… as people recognize the limitations and try to remediate the problems. I think that's where the Christian perspective could provide some guidance as to what direction that would go in terms of just like applying a greater wisdom in the way we design things…"

On how AI ethics might be more robust in light of Christian teaching, Robert said, "The mercy and grace of God and the fundamental tenets of Christianity… I think those things are very hard to be captured by ethics… Christianity is more than just a bunch of rules, right? I

believe that Christianity is more than just ethics. It's more… a kind of correct way to live your life.”

Robert was the youngest of my informants and passionate about his Christian walk. He often used his company as a stand-in to demonstrate how technology and God can play the same roles in our lives.

Timothy is a human-robot interaction (HRI) researcher and computer science professor at a private research university. He identifies, nominally, with the United Church of Christ. On worldview, Timothy says, “Of course worldview matters! Why is [a robot priest in a little cute Buddhist monk shape] imagined for a robotic application… dispensing wisdom? Did worldview enter into that? Absolutely. One of my chief interests is care work and the uses for which automated systems… are going to be envisioned at death beds or in hospice care. Will those be influenced by what you think death is and how death should be prepared for and experienced with others socially or individually? That's going to drive a lot of how things are imagined, for sure…” Expanding on worldview, and the likelihood of AGI, he says, “The type of judgments a good human being conversationalist estimates are just so far beyond what any system could reproduce intentionally… the magic eight ball is a great autonomous conversational aid. It just has no internal intelligence we would ascribe to it… I don't think AI will ever lose part of its magic eight ball quality of needing to have human reception and human creativity interpretation give its product life and…”

On taking ethics to another level, “Ethics gets put in the red-light function or the forbidding parent… it just sets up a dynamic where ethics is never going to actually be in the constructive position… it gets cast, often gendered, I would say, as kind of… maternal prohibition… I really think it's important that ethics take and infiltrate the creative imaginative

side… and say that really the problem in many cases is that there's a lack of imagination… [we need] an influx of people who really can speak to real needs and do so imaginatively, resourcefully, creatively, collaboratively, so that it's not about prohibition."

On the provenance of ethics and the need for a robust Christian voice, "Part of the dissatisfaction with ethics… stems from the fact that there are no Archimedean points in terms of context-free, interest-free, personality-devoid, statements of principle. They're all coming from somewhere and some place. And… the question of provenance is going to be key in terms of looking at why do norms change, and under what basis, or what conditions do they change?... With AI, God, and life, if you have an impoverished notion of intelligence or a fantastic image of intelligence, then you tend to just run over some things that I think Paul was trying to get at in talking about love… Intelligence does have a provenance… and we need to think about where it is situated. Is it situated alongside power?... Is it situated next to life and vitality?... I think First Corinthians… puts a lot of ideals like intelligence, to the test… My interest in care is surely shaped by that kind of Christian tradition of love…"

Timothy leans toward the liberal/progressive edge of mainline Christian belief and much of his thinking around AI ethics focuses on human flourishing, justice, creativity, and love.

John is an AI pioneer, author, and Christian apologist from the UK. He identifies as an evangelical independent. On the materialist worldview and AI, he says, "If you get into the mindset of somebody who's not a Christian… and material is all that we have… then the human brain is simply a very complex, organic computer and we can replicate at some point. And then… because… we can't prolong how long the physical body lasts, they want to upload theirs into a computer and… of course… live in a computer forever." He continues by qualifying that, "A relatively small number of people think that utopia is achievable by AI. I think you'll find that

a huge number of scientists don't buy into that because they understand the limits of what we have… People attribute far too much to artificial intelligence and it isn't intelligent at all actually. All current algorithms are doing, and I think all they will ever do is to… very cleverly mimic aspects of humanity, or human agency… but the danger lies in the fact that it's often very oversold… We [humans] have a reasoning ability that AI does not, and never will possess… The whole idea of the singularity… that a computer will just be able to get cleverer and cleverer by itself is complete nonsense."

On ethics and laws, and God's moral nature and ours, he says, "Why do we have legislation at all? It's a reflection of the fact that we're sinful people and we don't do what is right… It's very interesting… that the scope or the complexity of law and legislation increases the further we get away from God… I think the further we get away from God's moral nature, the more we have to codify things in detail, because we don't see it." And on ethical provenance, "Eternity is written in our hearts, and we know that we're made in God's image, so even though people are atheists, there's still something in there."

On moral freedom and love, "God has created us as free moral agents… and I think the whole issue of having the capacity to love… cannot be possible without moral freedom because to love is to have the freedom to love, and of course, God epitomizes that in that he loves us, even though we are unlovely… Of course, the fall has distorted that… The grounding is that we are free, but sin has messed that up, which is why we need to be all the more cautious about putting things in front of us that actually mess it up even more."

And on the limitations of so-called virtue ethics, he says, "[Virtue] has a huge resonance with people who are not necessarily Christians… from a Christian worldview perspective it's

very easy to settle on virtue as being the guiding principle… but we can't really be virtuous without the Holy Spirit's work and his power within us…”

On how Christian teaching could inform a more robust version of AI ethics, “I've produced a sort of taxonomy of AI where you can, through that lens of the image of God say, what aspects of our humanity are being potentially compromised? And it goes even as far as to say, when it links into the issue of idolatry, to what extent are we being lured away from what God wants us to be by something that seems so cool, so convenient, so efficient? But actually, is it helping us to image God? And I would argue that if you look through the lens of what it means to image God, to image Christ, you can say, this is taking me further away from him, so therefore it becomes idolatry because I'm replacing God, I'm replacing my obedience to his will by… As Christians, we can think about these different AI applications and what they're doing to us and answer the question, is this helping or hindering my walk with God, my holiness, my loving, my neighbor, my being kind to people?”

John has worked in AI since the early 1990s and his vision of AI is deeply informed by his Christian faith. Because he has first-hand understanding of the limits of AI, he is skeptical of artificial general intelligence. He is also skeptical of the efficacy of ethics without the indwelling power of God.

Nicolas is a philosophy professor who specializes in tech ethics. He identifies as “homeless” Christian and is not currently affiliated with a church or denomination. On the Christian value-add to AI ethics, “The term AI ethics is really broad… but I think of there being like four categories... One is data ethics… you don't have to care about AI to care about just information ethics. The second bucket is the ethics of machine learning models… accountability, how much we know about the models and also how the models become biased. The third bucket

is what the models actually do, which is predict… And then the fourth bucket I'm calling AI personhood… how it's changing us. Like, are there going to be standards eventually about how we treat non-organic beings? That's where I think religious ethics in general and Christian ethics specifically has the most to say. The other three I think can be fairly adequately handled with the secular ethics, but it's the personhood stuff where I think the resources of religious traditions are by far the most important… AI personhood… is definitely the least understood the hardest to understand, the hardest to predict, most futuristic… It just seems obvious to me that the personhood bucket is testing the limits of consciousness and human connection and so, I mean, even attempting to have that conversation, pretending like the last three thousand, four thousand years didn't exist, was just insane."

On the inflection point of secular and Christian ethics, "[When] we're talking about the right policies and the right things to do… I feel very good about having these conversations with secular people or people of different faiths. It's just when you stop the conversation and say, okay, but why should we care about this at all? I think Christianity has very good answers to that and other people who have no faith at all, don't actually have good answers to that. Like, okay. So why do you care about this at all? It's like, well, I don't know… I've just been so caught up in caring about it that I just do. So, in that sense, I think Christianity deeply has much to say about it."

My interview with Nicolas was the shortest. When he hadn't thought deeply enough about my questions to give a good answer, he said so, which I respected! Overall, he felt that so-called secular ethical frameworks could handle most areas of AI ethics just fine, but that the Christian worldview was uniquely qualified to speak to personhood and a rationale for ethics.

Christian is a former big tech developer now running his own AI start-up. He is an Evangelical Presbyterian. On the unique quality of the Christian worldview, he says, "The most unique thing… that Christianity brings… is telos… is purpose: the Bible is a witness to what God is doing in creation, from the beginning unto the end. And so, we believe there's a certain direction to things that God is bringing about through human innovation and technology and through AI. And so, our unique concern isn't merely ethics, which is more of the domain of what you might say is what's in the common interest, in good of humanity. It is very specific. It's peculiar. It's the kingdom of God, which is more than the common good. There is something unique about the belief in the new creation that Jesus has inaugurated through his death and resurrection… We can participate in the common good, because we believe it's a witness to the kingdom, but it's not really an end in itself... So, at the highest level, that's what I think is what is essentially Christian in this conversation about AI."

He continues, "It's more than an ethical principle, it's like a mindset, that… takes into account the fallenness of the world… and then in light of that, trying to build our things in a way that can be redeemed, you know, ways that can be healed, ways that can recover from these kinds of situations and give people freedom.**"**

On the Christian understanding of evil and its relationship to us and AI, "As Christians, we already believe in a superintelligence that is actually malicious towards us, which is Satan… Satan has so much more power… and so much more intelligence than any human being, and actively works against God's will to destroy human civilization, to corrupt it, to dominate it… [and] to corrupt what human beings have made also. The way I view AI is that it's something within God's plan and God's story… but it's also something that's… corrupted and actively used towards evil ends that may be unintentional to human beings who are ignorant of those possible

ends… [because of that], there's that awareness even for me as a creator of AI is that, you know, it's not like I'm the good guy. Not necessarily."

On the Christian view of empowerment and creativity in ethics, "We believe that renewal and transformation happen first from within through the action of the Holy Spirit on a person that changes them and gives them the power to behave 'ethically' but also in ways that go beyond that creatively… There's a comparison between the law [in the Old Testament] and the royal law of love... The royal law of love, unlike all the other laws, is not proscriptive. It is actually generative. It's creative. It unleashes. To love your neighbor as you love yourself is something that is above ethics. Like, you can do that in a million ways."

On the value-add of Christianity in AI ethics, **"**The Christian ethic… is not just ethical behavior. It's something more. It's glory. It's vitality. It's life, and it's overflowing life. And so, I think that ethics and rules… are never going to match up to that kind of vital force in a person's life… that vitality is something that's far greater… It's just another hint at this underlying deeper reality that ethics can never codify what God has woven into creation itself. And that vitality of the Holy Spirit, that creative generative force is so much more… that it produces the real vitality people long for when they talk about the common good, or when they talk about ethics… That is what's actually desired of the human heart."

On the provenance of both intelligence and ethics, "With regard to ethics, we already have seen this story play out... God created human beings in his image and gave them freedom, autonomy to make choices. And then they rebelled against God's will and law. And the question was, how was God going to deal with that?... And we see the way that God responded to rogue AI, you could say. If human beings are an artificial intelligence of the divine intelligence and the human beings create an artificial intelligence of our human intelligence, you can learn from

God's pattern in scripture of how God treats rogue human AI. And what does God do? God actually works actively to save and redeem us. Not simply, you know, set the reset button to erase it or something else. Not even fear that it would get out of control and destroy God somehow. I just think there's some interesting analogies there in what the human project of trying to create artificial general intelligence is and what we have in the kind of the original stories of Genesis."

On the limits and telos of AI, "I personally don't believe that… we could ever build AI that's so perfect, so good it can only do good or anything like that… that's why all the more ought we to be innovating for the kingdom of God, innovating for a witness to the gospel, innovating to love people and serve them well."

Christian was among the most passionate of my interviewees. He had clearly thought deeply about technology and theology, and presented an expansive and inspiring vision for the Christian role in AI ethics

Christopher is a computer science professor at a private west coast university. He was born and raised in Central America but now lives in the U.S. He is a practicing Catholic. On worldview, he contrasts the end goals of AI creators, saying, "The materialistic point of view is going to have to go with material, with worldly stuff, right? The other one wants to be godly, in theory. So, the tension is in asking, what are we doing this for? What it is that we want to achieve? We were created… in his image, so if he's a creator, we must have a creative part… The thing is that since we are also broken, when we create stuff that is beautiful, there's always somebody [with ill intent] that will come and say, how can we use this too?"

On the role of Christians in AI ethics, he says, "Our role is not just to be a brake… if we really are people of faith… if we are seeking God the way that Jesus taught us… then I'm going

to show the love of Jesus in this… it's going to show in a mystical, a metaphysical way… not a very objective way… a way that is not measurable, but it does show."

On the base line question for Christians in AI, "Is this getting me closer to God? Am I showing love in this? How is my work showing the love that God has for me to my neighbor? I have to acknowledge first, that God loves me and I'm here to show everybody else that through me… So, when I'm doing artificial intelligence, when I'm applying machine learning, when I'm building a new algorithm… Why am I doing this? What is the purpose? How is this going to affect the users?"

On the Christian contribution and value to ethics, he says, "We can go all the way back and get to the morals and ethics about our claims however, if we really want to build on stone, we need something to back us up…. I don't know about the other religions…. but about ours specifically? Okay, we have a stone where we can build on and it is explicit, okay? Jesus Christ, cornerstone. So, if we build upon that, okay, we're going to make an unshakable foundation. That is our faith."

On the fallen state of humanity and the redemptive nature of God, "So, we have a saying in Spanish: 'Dios escribe derecho en renglones torcidos.' I think the same in English is something like, 'God writes straight in crooked lines.'… So even if things go wrong… he's going to work out things because he loves us… He could actually make it right just by doing this, but he's going to try his best with our freedom because that's actually a big constraint for him."

Christopher's focus on purpose was central to his thinking about AI. He came back, again and again, to the Christian theme of love, and its potential to play a redemptive role in AI ethics.

Mark is a dean of computer science at a Catholic university. He is Catholic. On the limits of AI, "I personally am very skeptical of AGI. I feel that we don't even know how to create self-awareness. Some people say it's just going to emerge at some point and, well, maybe, but we can't even be self-aware about ourselves…"

On the Christian view of our design and the role of AI in our lives, "People are desperate for relationships. Now I would say God hardwired us that way… We are created to love and be loved… and people are so desperate for that, they will form relationships with pets, if that's the best thing they have or with machines that are pretending to be pets or even now with machines. If we got to a point where people were going to machine therapists or having relations with machine… falling in love with robots... that's not the way things are supposed to be… I think that it's sad that some elderly person would be in a situation where they're being taken care of by a machine as opposed to a nurse."

On the provenance of ethics Mark had a great deal to say, "There are things about our current worldview that we take for granted, but they weren't there before the Christians showed up. In other words, they weren't part of the pagan world… This idea about human dignity… The idea that everybody has rights… This new idea of family, moving away from the idea that basically the wife and children are property… That came out of Christianity… [The idea] that work in itself has dignity, that comes from the Judeo-Christian tradition… This vision of seeing God as love as opposed to God as a service contract… Love your neighbor, taking care of widows and orphans and the stranger… A change in how government is seen, so that rulers serve the ruled… That's a Christian idea. There are so many things that we take for granted that basically society inherited from Christianity… and have now become secular ideas… if we said, look, we owe the church a debt of gratitude for what it was able to do in spreading these values

along with the good news, they could say, yeah, but it doesn't matter who the messenger is, it's the message that counts… It's an interesting question, whether a [a culture founded on] Christian ideals can survive without Christian citizens."

On whether ethical principles and laws enough to ensure robust and benevolent AI, he says, "Well, of course the answer to that is no because people are sinful. And I know sin isn't a very popular word. It doesn't show up in modern culture at all… it doesn't even show up in certain denominations... But the Catholics have no problem with sin. We talk about sin all the time… So, no. Ethical principles and laws are not enough. That's why you have to get people into the habit of doing the right thing. Inculcating virtue."

Mark has written textbooks on technology ethics and is no newcomer to the field. His interview tended toward the philosophical and historical with a distinct Catholic intellectualism.

Dominic is a philosophy professor and director of a tech ethics center at a large Catholic university. He is Catholic. On the importance of worldview, "Worldview influences virtually everything we do as human beings… including how we develop and use AI." He expands with a story. "We could spend hours talking about specific examples of the ways in which worldview can make a profound difference in how we approach the development and use of AI but… I can think of examples from my own experience that start to illustrate this. I offered a course… on ethical issues around nuclear weapons and I was struck… [by] a large number of aeronautical engineering students... I thought that's a curious demographic so… I took a couple of them aside and I said, what's going on with you aeronautical engineers? And the answer… almost brought me to tears. What they said is, 'Well, look, our aeronautical engineering program is world-class. We all leave here with good jobs… in aerospace or defense. At the same time… there's this constant chatter about the importance of the value dimension in human life and human work and

we just all decided that before we leave campus, we had to find a place where we could go to think through the value implications of the career choices we have already made…' For me, that's a really vivid demonstration of the way in which the cultural, religious traditions that we come out of make a demonstrable difference in the way people think and act."

On his views of different ethical traditions, "I've always been strongly drawn to the virtue ethics tradition… The more deeply I got into serious work on technology ethics, [the more the] consideration of other kinds of ethical issues didn't have the same impact in helping me to come to understand the sort of thinness of, especially the deontological tradition in ethics and the consequentialist or utilitarian tradition in ethics… At first, I was a little bit shy about proselytizing, but with every succeeding year, I'm losing more of that inhibition. I mean, I still, as an educator have to give every voice its due… but my students always emerge understanding that while I will play devil's advocate for any position, they realize that I'm really a virtue ethics guy. And so, I think this is one of the very specific and really powerful ways that a tradition with obvious deep roots in the Christian tradition makes a huge difference…"

On how Christian teaching can take ethics to a higher level, he said, "In so many settings, what people hear when they hear ethics is… it's all about telling me what I can't do… Whereas I keep saying… you have to think… about ethics as a way of pointing us toward better futures, helping us to imagine how the world can be… holding up this vision of what's a better world. Why does that work so well with our technical colleagues? Making the world a better place is really the core moral impulse that animates most of our engineering colleagues. There's a sense in which that's what engineering is all about. It's making the world a better place… the engineers and the scientists… are the cultural optimists. They're the ones who say, oh, there's a problem. Let's fix it. Let's make it better…. most of what technology ethicists have done is to worry about

risk and harm and the bad decisions that have been made, as opposed to promoting this sort of visionary conception of what technology ethics should be as trying to contribute to human flourishing in a positive way, in collaborating with our technical colleagues and helping that to happen."

Dominic is a technical optimist who cares deeply about human flourishing and believes in the power of a good story. He says he has thought long and hard about how Christianity can make a positive impact on AI ethics.

Joseph is the chair of research in technology ethics at an evangelical non-profit center. He is a Southern Baptist believer. On worldview, "So obviously as a Christian, I believe all people are created in God's image. This is part of my moral and ethical and theological worldview… But we live in an age… [where] there's nothing supernatural, nothing outside of this world, nothing transcendent… in this naturalistic materialistic understanding… there's no real purpose… morality is a social construct, you just came from the muck, you have evolved over millions and millions of years. There's nothing unique. There's nothing special… when you die, you die… there's no afterlife... So, it's really interesting to me that… at the same time [in history that] we're humanizing machines… we're dehumanizing ourselves… we're trying to create these machines to be like us because we inherently recognize there's something unique and special about us... And that's where I think Christian ethics speak into this, is to say, no, you're not just an organic computer. You're something more than that." He adds, "As a Christian – and I know this is going to sound incredibly arrogant, but I mean it in the most humble way possible – I believe that my beliefs are grounded in reality. That they're God's design for the universe."

On AI ethics in light of the biblical understanding of human nature, "In the AI community, I think there's this utopian idea that we can create an unbiased algorithm… You

cannot create an unbiased algorithm because it's created by fallible humans who have our own biases… so naturally, that's already baked in… The problem isn't, well, we just have bad algorithms, or we just have injustice. It's actually that we have a sin problem… Bias or unfairness or racism or sexism… are real issues, and we have to deal with them, but I think the root issue is sin. It's a rebellion against God. It's a rebellion against our designer and our creator."

On what Christianity adds to the Ethical AI conversation, he says, "We can't shed ourselves of moral rules and duties… but the Christian ethic is a much more robust than people think… it's the idea that you put off the old and you put on the new… there's this watershed moment of when I become a believer, I'm a new creation in Christ… Thus, I'm called to live in a certain way… I'm becoming more like Christ, who is the epitome of virtue and goodness and righteousness and holiness… we're to put on humility, grace, humbleness, charity, and love." In contrast, he says, "This utilitarian mindset that's really prevalent in… AI ethics is the means to the end. It's the greatest good for the greatest number of people… A Christian comes in and says, no, our ethical theory is actually based on the image of God or human dignity for all people, which means… we should be respecting the rights of all people…"

On a Christian value add to ethics, "And so you have that centrality of love, and that's where I think we have to have a grander vision. When we talk about ethics… the Christian perspective adds a lot to the conversation around technology by asking, what were we created for? What are we becoming? What is our telos?"

On the nature of faith and the need for diverse voices, he says, "We are a worshiping people but who are we worshiping? Are we worshiping God? Are we worshiping ourselves? Are we worshiping other people or things? We are worshiping something, so in that sense, we're all a

people of faith. It's now [a question of] whose faith is actually welcomed at the table when we're making very important decisions and public policy and ethics and philosophy, surrounding the research and development of artificial intelligence."

Joseph's view of tech ethics is deeply steeped in his evangelical worldview. He is widely read and has written books on technology for the church. Like others, he believes in the centrality of love and the divine purpose of life.

Michael is a data scientist and machine learning researcher at a biotech company. He is a member of a Christian Missionary Alliance congregation. On worldview, "As Christians, we believe that we are God's created people to see his work done here on this earth… And I think from a computer science point of view, you think that there must be something unique to humans then that we can't put into a machine... One of the areas of application of AI that I'm using is looking at how to do gene editing, so we can talk about AI in terms of what it means for a machine to approximate life… And from an evolutionary point of view, I don't think this would be controversial… it could lead to a fitter species, right?... From a Christian point of view, we're still using the code… the language that God imprinted on all of us, we're just taking some of the people that… maybe have certain genetic defects and fixing those defects."

On the Christian value add to AI ethics, he says, "On the areas that church can help, we talked about virtues, but there's another area that I think is maybe even more important, which is, let's just call it the shalom of the soul… We can optimize certain things about healthcare or driving or reduce number of deaths… but it doesn't really do anything for some of the inner turmoil that we're all dealing with… The message of the gospel is that your biggest problem has already been solved, right? Rest in me [Jesus]. That's the message. And that's not something I

think the folks driving this technology are going to address… Our biggest problem in some ways is an internal one and I don't think AI is going to solve that."

On playing God with AI, "As a Christian, a lot of what we do is based on faith that we are endowed with a unique soul. So, any effort to build or to create our own humans is a bit like the tower of Babel, right? Like a lot of time, a lot of effort is going to go into it, but it'll never be endowed with a soul…" And on the need for Christian wisdom in the mix, "AI should not be a standalone decision-making tool... It's a tool for us to use, not a life for us to create or something that we would use unquestionably… I want to say that there's going to be a number of realms where it will be important to have a Christian perspective involved."

On how Christian teaching could level up AI ethics, he said, "I've been thinking a lot about the Proverbs lately. Proverbs don't tell you what to do. The Proverbs tell you what kind of person to become so you know what to do. That's part of what the church does… if we adopt more of the character of Christ, we'll learn the right things to do even in the face of things that are very uncertain right? Now, how does that play into AI?... I can ask Google… the consensus for a given question, but it may not have optimized towards mercy, or kindness or wisdom. There is a dimension there that's missing. You'd have to codify mercy or kindness somehow… I wouldn't know how to optimize for mercy because mercy is almost illogical… Tit for tat doesn't sound like something Jesus would teach. Jesus would say turn the other cheek."

Michael is more optimistic about AGI than other Christians in this study but positions himself as one who probably won't be working on it. He is an advocate for the Christian voice in the work of AI from proof-of-concept to use.

Peter is a former software engineer and now a pastor of a non-denominational Christian church in Silicon Valley. On AI and anthropology, "If I were to ask people on the street, do you

have any fears that AI might degrade what it means to be human or might somehow threaten our

humanity, most people would say yes… and yet if you ask those same people, well, what does it

mean to be human then? What is it that might be threatened?... they're just not sure… The

tension that tech companies face is that they're aware that they're potentially threatening

humanity, but they don't know what humanity is."

On worldview, "In a world that's strictly materialistic, humanity boils down to neurons

and brain chemistry, but I suspect even the most diehard materialists have some sense that that's

an insufficient anthropology. That there's something more than just chemicals." He continues, "I

have kind of a framework for a biblical anthropology that I center around four kinds of

relationships. (See Figure 4) One is a relationship to the creation… which is characterized by

ruling or stewarding.... One is a relationship with ourselves… which is characterized by self-

knowledge or insight... The third is a relationship with other people… which is characterized by

love. And then the fourth is a relationship to the divine, to God himself… which is characterized

by service and worship… If you put AI in that diagram, and you put us in the role of God, then

we are really only concerned that the AI would serve us and rule over the things that we

instructed to rule over."

Figure 4
Framework for biblical anthropology.
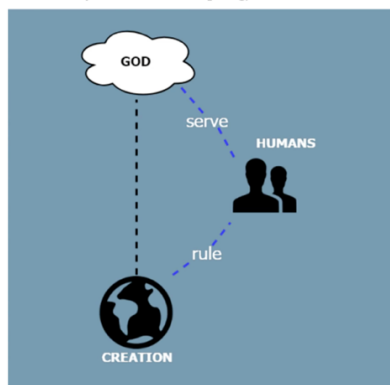
© allthingsnew.tech

Figure 4 - Biblical Anthropology

On the provenance of anthropology and ethics, "I think there is kind of a modern, secular humanist anthropology that is roughly agreed upon and I actually think it comes from the Bible… the Christian message is a message of inclusion that the gospel is available to everybody, which in the first century was a radical thing... the gospel came in breaking down tribalist walls. But now, two thousand years later, our culture judges Christianity as tribalist and exclusive, which is a little bit ironic to me."

On using AI for good, Peter says, "When you're trying to solve [big] problems, full steam ahead! And let's solve them!" But on whether we could ever go too far, he said, "I think our technology is meant to stretch our constraints and move our boundaries but not to transcend them… I don't think we *can* transcend our constraints. So, it's not a matter of, oops we might accidentally transcend! Like there are constraints for a reason… If they are actually constraints, they're imposed by a third party, not by us, so, in some ways, we don't have to be careful not to transcend to them… the creation story ends in Genesis 3 with God putting an angel with the flaming sword at the entrance to the garden saying 'none shall pass' so I don't think AI is going to get around that flaming sword. I think God has put a boundary that we are not capable of transcending."

On the goal of Christians and the redemptive nature of God, he said, "I'm not that concerned with transcending the boundary or even doing dangerous things with AI because the Christian goal is not to optimize life on this planet. The goal is to encounter God. And sometimes people encounter God when they do something terrible. I don't know. Maybe Ray Kurzweil will manage to upload his consciousness to the cloud and realize how terrible that was and that will be the way God reaches out. So, I don't want to stop him from doing that. My goal is not to change the end point, my goal is to say, how is God at work in this? How is God at work in the

conversation? And I think AI actually creates an opportunity that our culture is asking questions about what it means to be human."

Peter has thought deeply, spoken widely, and preached passionately about how AI impacts humanity. He is a technical optimist in large part because he believes God is bigger and smarter than AI.

Joan is a former big-tech AI researcher and multiple patent holder who is now running her own AI start-up. She is an evangelical charismatic Christian. On the Christian worldview's impact on AI, "So in everything we do the 'why' is important, and the Christian worldview totally informs the 'why' of AI… Foundational to the Christian worldview is that there is a higher intelligence. That human intelligence is very low compared to the wisdom of God and there is a hierarchy that keeps men humble. The 'why' of humanity being made in the image of God… shapes the agenda for AI… Christians recognize the limitation of human intelligence… If human intelligence is inferior to God's wisdom, how much more inferior is artificial intelligence, created by man's intelligence? It really puts it in the proper place… Why do I say that? Because we don't have unlimited time. We don't have unlimited resources. So… [I would] engage in AI work that aligns with my priority, according to my worldview... With limited time and limited investment, I do what Christians are called to do: to build up, encourage and bless."

On the concept of the doctrine of common grace, "Because we're made in the image of God, whether our spirit connects to God or not, we have that DNA to do good because we have that image of God. And so, do I believe atheists would aspire to make the world a better place? Yes, absolutely… You don't need to be a Christian to see a hungry child and have empathy… So, we shouldn't deny the atheist contribution to make the world a better place..."

On governance, "Any technology by itself is neutral, but it is the sinful nature of man that turns a tool that's neutral into evil. And so, there's absolutely the need for governance, right? Governance is necessary but I don't think governance is enough. You need to have enforcement there."

On the need for human wisdom, "Let's not forget where machine learning comes from. Anything it learned is from the data. It develops a model from the data and then from that model, it would do the next thing because it's informed by the data… so that's the DNA of AI and the thing is, there is a difference between wisdom and intelligence. Intelligence means efficiency reliability, accuracy… Machines can do a lot better in a lot of these accuracy, efficiency, productivity type of tasks, operations, but that's not the only thing. There is the wisdom... For example, if even all the data tells you to do this, wisdom might override it because of compassion. Because of the care of others. Because you have a higher value… with God, you have a different why."

On how Christianity would enhance AI ethics, "It's very important to have Christians in these spaces because light needs to penetrate into this space… When the light is present, then you can direct. Or if there is darkness, you can shine… That's what we're called to do…. Behind the data, behind the algorithm, there could be some darkness there… Christians can be the loud voice of testimony… God says, seek first the kingdom of God and his righteousness. And it just shows how much God cares about righteousness. [We need] righteous AI."

Joan left a big tech company for two reasons: their utopian vision for AI and hostility toward Christians. She is deeply driven by the love of Christ and the "why" of AI.

Edmund is a data scientist at a large bank. He calls himself a "fringe" Episcopalian. On ethical tensions and living with them, "So how do we resolve the tension? I would propose that

we don't... definitely don't resolve it prematurely… Here's my biggest beef with Christians, which included myself for many years, if we're going to use that as an excuse to sit on the sidelines and do nothing… to look at the Tower of Babel and say, well, at least we're not doing that… that becomes an excuse for inaction… We need to have an anchor in Christian tradition… that clearly says humanity is prone to vices and to sin and to messing things up… but there is this God in us that pulls us into something better."

On the nature of sin and scapegoating, "We're like, oh, the problem is technology… The problem is the algorithm. The problem is Facebook. The problem is the big companies. It's always a scapegoating somewhere. It's never us or ourselves."

On the value add of the Christian worldview for AI ethics, "I cannot find any better way to give glory to God than to build technology that will make this planet more livable and solve problems that are taking away life. God is a God of life... Whether you're a creationist or not, even the secular person would say, wow, life on earth is a miracle… at the end of the day, we can hold hands with others who are not Christians, who may not believe in God or anything, but to say, look, we're all on this planet, we all can agree that working for the flourishing of life is a good thing. So, can we work together to build technology that will do that? So that's, to me, the vision of God's kingdom is building technology for the flourishing of life."

On the Christian worldview informing a broader perspective on AI ethics, "Obviously laws and regulations are necessary… But what goes beyond?... The way we build technology is just an extension of how we live our lives, and in many ways of who we are, so ethics then becomes an issue of identity as opposed to a list of things that you should or should not do. So, what if, instead of asking how could things go wrong, which we still need to ask, we ask how does this technology speak of who we are? Maybe the popular idea would be that ethics is…

about behavior and achieving behavioral outcomes… It's easy for us to just focus on what could

go wrong… but if we lose sight of AI for good… we lose also the opportunity to see the beauty

in it and in some ways to glorify God in it… Let's marvel at the good that has come from these

technologies."

On bringing Christian teaching into the AI ethics dialog, "When we're talking about

ethics, it would be an immense waste if we just really rejected 2000 years of Christian reflection

on ethical issues… Christianity has a long tradition of thinking about these things. The

technologies are new, but the problems are not."

Edmund is a realist on the limits of AI and believes people give it too much credit for

both good and ill. While his Christian faith informs his own life and work, he has an expansive

vision for working in unity with people of differing worldviews.

Benedict is the director of a technology ethics center at a Catholic university. He is

Catholic. On ethical tensions, "I would push back… on these dichotomies. Like, I don't see

robustness and privacy as being necessarily opposed to each other… I think these

dichotomizations are often 'motivated reasoning' and they don't have to be, and this is why:

because you *can* maintain privacy if you encrypt the data properly… The downside of it is that

it's energy intensive so it's going to cost more money in terms of compute and in terms of energy

costs and those sorts of things. So, when you are running your company, you say to yourself, oh,

we could do the nice privacy solution, or we could do the cheap solution and since we're a

startup, what are you going to choose? You're going to choose the cheap solution of course.

There are ways to solve these problems so that they're not dichotomies."

On a worldview diversity, "A lot of people doing AI right now are… atheist utilitarians…

They're into this idea that human nature is very malleable, and we can apply technology to

human nature as long as it produces some benefit later. They don't recognize that that's a very small perspective on reality. They are a very tiny minority of all the people on earth and they're deciding to make decisions for everyone else… Other people's voices need to be included and this is an opening, not just for people of different genders or sexualities or racial or ethnic backgrounds. It's an opening for people with different worldviews also. And worldview diversity is something that people have been talking about more and more also, which is that we really don't need to have a bunch of utilitarians who want to quantify everything making all the choices for us, because they're a very small percentage of people in the world… and there's a lot more going on in the world."

On ethical provenance, "Very often atheists in the Western world have Christian ethics… [but] they're working on the fumes. The gas tank has run out, but they still have some fumes. And so, you have these ideas like suffering is bad. Let's get rid of suffering. But if you asked the Romans about that, the Romans would have been like, who cares? Let them suffer."

On the Christian vision of sin and freedom, "Original sin is completely real. I think there's a very strong tendency of humans to ethically make bad choices… Christianity's understanding of sin highlights that… if we are not making the right choices, we're going to end up living in hell… When it comes to ensuring good, there's no way to ensure it because humans have freedom… Ideally, we would all choose good by nature…. But we don't live in that situation yet… In the meantime… we're somewhere in between that quote, 'if all men were angels, no government would be necessary and if all men were devils, no government would be possible.'"

On how the Christian worldview leads to a higher vision of AI ethics, "When it comes to technology, we need to not be seeking power, we need to be seeking goodness or ultimately

holiness if we want to phrase it biblically speaking, because without that, power will end up destroying us…. [We need a] technology of holiness… It's like righteous AI. Ultimately… we need to make technology holy. It has to be better than good. It has to be better than beneficial. It has to be holy. And if you look at the holiness code, 'be holy as God is holy,' that's a big calling."

Benedict made the strongest case against ethical tensions being an intractable problem in AI ethics. His job is to think and write about technology ethics and he brings extensive scriptural support to his work.

Jonas is a theology professor at a private university. He is Catholic. On ethical tensions, Jonas wondered, "Is the AI biased or is the AI reflecting statistical realities in the 20th century as an aggregate?... What is an accurate picture of the world? Is it the world as it functions or is it the world as it ought to be?"

On worldview and ethical provenance, he says, "Christianity is a worldwide, two-thousand-year-old tradition in which ethics… cannot be… fully assessed or understood except with reference to God… We can deploy ethical principles… to discern the right action in a particular situation, but the Christian claim is that… we don't operate by inference from principals. We are trained by principles and then they become part of us. And somebody could just say, 'Well, aren't you just saying things that contemporary cognitive science might say?' And then the Christian could reply, 'Well, yes, but Christianity is a tradition of human reflection that has had a hold on these things for much longer than cognitive science has and is perhaps more equipped to deploy these things accurately…'"

On the need for a Christian voice in the conversation, "What Christians have to do for AI… is ask what would it look like if we let Daniel Dennett define what it means to be a

person… We can implicitly leave people from a secular background with the challenge, like, okay, you guys explain it. Can you?... Christians… have a justifying reason that comes from a standpoint of faith. Not blindness but illuminated faith… such that they would be willing to defend it, even if they can't prove it…"

On the Christian view of the human mind and AI, "Christianity… is self-actualizing through self-giving love and that's something that cannot be represented in a neural network… There's no way of quantifying the interiority of human self-gift… it doesn't register as representable to a computer in some sort of big dataset… Social psychology tries to quantify these things, but then it often ends up with rather shallow definitions that don't get at the richness of these experiences and these realities in human life… but within Christianity… the transcendence of self-gift… the wounds of Christ [show] what utter compassion for and self-gift to a fallen world looks like." He continues, "I think of the GPT [generative pre-trained transformer] systems that will put out this text that's more or less coherent, but it doesn't really know anything. This is something that Dante talks about in the first circle of hell in the Inferno. There are these people who are constantly running in a huge circle… chasing a banner with no symbol on it… they pursue with all their energy what is in fact, an aimless pursuit… And that's what the mind of a neural network is… it's not even an imitation of a human mind because it simply is following the banner of that reward signal or the training signal that comes back for the adjustments to the weights in the network… Human beings are different from either symbolic AI or machine learning because you can give human beings an explicitly, symbolically denoted rule, and then tell them to go with it and they'll struggle and they'll strain and they'll try to apply it and that bumbling around in moral life and human relationships is how we get formed for our implicit sensitivity to the dynamics of human relation and the moral landscape of human life."

On how a moral life is formed, "What Christianity has to offer is… that our own ethical training is delivered to us both by example… and by precept. But we don't really get a hold on the precepts, on how to apply them, until we're able, through our own trial and error in our practice, to link the precepts to our human relationships."

On the uniquely Christian view of personhood, "In the ancient Greco-Roman world… Before [you] were a person, [you] were a member of the human species… [your] role or [your] instrumental service is what made [you] a somebody… But in Christianity, the notion of the person becomes attached to the persons of the Trinity… The Father, the Son, and the Holy Spirit… form some sort of community in which… they actually have their existence by their relations with one another… It's an interiority of self-gift from which we go forth toward one another. We express fully our personhood through relationships with one another."

While Jonas is a theology professor, and is widely read in philosophy and history, he also has a background in computer science and brings a strong theology of technology perspective to the conversation.

David is a computer science professor at private Christian university. He is a member of the Christian Reformed Church. On an assumptive worldview in technology, "If you asked me as a new graduate what was the worldview of your education, I'd say it's neutral. Like I just learned vehicles IR and C programming and X, Y, and Z. But the truth is that even though the professors never talked about the big ideas that animated every class, there was a worldview that was implicit in almost every lecture and every course about seeing the world through a technical lens, seeing the world as a machine that could be controlled and optimized… if you go back and reflect, you go, yeah, there was a certain animating view of the world and how things should be that were shaping me and misshaping me in some ways."

On how worldview impacts ethics education, "A lot of schools in the secular academy…
are pushing to add a mandatory ethics course alongside the standard computing curriculum. And
I think that's good. That's better than nothing. But I would say that the premise that you teach
everyone in a certain way of thinking about the world, and then you slap on an ethics course,
doesn't make an ethical person… You have to be careful that it doesn't become all about
efficiency because at its kernel, it's all about optimizing some kind of goal function, which
means, necessarily, that you boiled things down and numerics, you've reduced things to
measures and to numbers."

On the nature of humanity how worldview shapes us, "We're people who are driven by
our hearts... and the way to the heart isn't necessarily through the brain. We aren't primarily
thinking things, we're primarily loving things… How are our loves shaped? Our loves are shaped
by practices… If you're immersed in a certain worldview, being taught to think about
efficiency… about technical solutions to everything, or you're working in an industry where
those values are always swirling around, then you're going to be misshaped in a certain way that
needs to have some kind of countermeasure."

On the unique contribution of Christians and the doctrine of common grace. "As
Christians, we know that God cares about technology. God cares about people. He cares about
our cultural activities… We also recognize the notion of common grace, which is the idea that
even non-believers by God's grace and the working of the Holy Spirit… are able to uncover
creational realities. And I mean, we see this in computer science. Most of the people coming up
with wonderful new possibilities are not Christians, but they're living in God's creation, just like
any one of us and they're able to see creation, even though they don't acknowledge the creator,
the creation impinges on all of us and we're able to observe it and we can take that." He then

cites theologian Abraham Kuyper's claim, "There's not one square inch in the whole realm of human existence over which Christ who is Lord of all, does not cry mine."

On Christian character and the need for supernatural empowerment for ethics, "What does it mean to become the kind of person who reacts, who understands the context, knows their times… and knows how to respond in an appropriate way to situations that you couldn't anticipate with rule-based ethics… rules are important… but… what does it mean to cultivate a Christian character that is able to have the wisdom and the intuition to respond well?… That's where the work of the Holy Spirit comes in shaping us."

David has written several books on the how Christianity impacts technology and engineering. His view of the forces that shape or misshape us is central to his thinking.

Jerome is a professor emeritus of mathematics, author, and Christian apologist. He identifies as a non-denominational Christian. On the worldview landscape in high tech, "In the west, and for the last fifty years or so, for the first time we're living in a space where there's no shared worldview. The Judeo-Christian tradition is being rejected by and large. And a lot of the pioneers in AI work, particularly in AGI, are, as you know, atheists in their thinking. So clearly, worldview plays a very big role."

On intelligence and consciousness, he says, "In creating humans, God linked intelligence with consciousness, and we normally associate it with consciousness, but in the AI world, consciousness is, for many of the leading people, totally unimportant. In other words, we do a thing that simulates intelligence. If it's not conscious, that's not important."

On how a vision of transcendence impacts AI, "If your worldview has no transcendent dimension, it could be simply, 'Well, if it can be done, it should be done…' Unless we acknowledge a transcendent dimension, we're in real trouble… All AI systems so far have to be

fitted with a moral code that is the moral code of the programmers. So where do they get it from?... On what do they base them? And are they going to stand the test of time and the test of pressure?"

On Christian altruism and evolution, "It's basic Christian morality that makes the difference… Christian altruism, putting oneself out for others… doesn't fit in with the evolutionary picture at all."

On the nature of faith, "You use the word faith… faith is absolutely part of my professional life, but not faith in God. It's faith in the mathematical intelligibility of the universe. I feel today that we need to make very clear what we are talking about when we use the word faith because of its multi-valence. The secularists of this world say [Jerome] is a man of faith, so it's not worth talking to him, that definition of faith is believing where there's no evidence… I would always add the question, faith in what? In other words, if you're talking about faith in God, well say so, because one of the very important things in the whole AI debate is the whole business that worldviews are matters of faith. All of them."

On the unique value of the Christian worldview in AI ethics, "[There] is a major misunderstanding of what the Christian message is because many people think of it in terms of rules and regulations… to follow in the hope that one day, God will accept you. And that gives you the kind of moralistic approach to religion… which is not Christianity… Christianity is not principally a moral code. Although it *has* a moral code, it's to do with the relationship with God and the new power that enables us to live… If people today are prepared to sit and listen to Max Tegmark or Yuval Noah Harari… I want to say… there is another scenario which is much older than theirs and has got the great advantage of its supporting evidence, which is much more powerful than theirs… it actually accomplishes what they vaguely hope to achieve, but never

will, as far as one can see…" He continues, "The message of Christian salvation… is offering a firmer basis for ethics in that it gives people a power to obey… the whole problem is not the rules written upon the wall, but where do I get the power to keep them? If there's nothing transcendent… if there's no God who watches and will judge, then there's no argument left… It's not that we can safely get God out of the picture and carry on and it's all going to be okay… Without God, there's no real reason to behave."

Jerome has thought deeply and spoken and written widely about Christianity and technology. His work – and his answers to my questions – reflects a strong Christian apologetics approach.

Anthony is the director of a resource center for Christian tech workers. He is an evangelical Anglican. On the rationale behind his resource center, he says, "Christians working in technology don't know other Christians working in technology so they tend to feel alone… Christians working in technology didn't have an imagination for how their skills could be used in ministry… and Christians working in technology don't know how their faith or the Bible relate to technology or how technology relates to scripture."

On worldview, "Artificial intelligence and computers in general are structured in a very materialist worldview and philosophy that says everything can be quantified or digitized and that information can be reduced to data without any loss of meaning or without any loss of information. Christianity has a priori commitments to a reality that is both material and spiritual and there are ways that that certainly overlaps with a materialist perspective, but [the materialist perspective] is insufficient for understanding the world that we live in."

On narratives of the future, he says, "[I'm] looking at digital technologies in light of the presence of God… My interest was… around how the internet functions as a spiritually

formative space… Technology has a view for what the future looks like… and a lot of the stories that are being told about technology, including about AI, are the stories that Christianity tells about what the future looks like and where we're moving towards."

On the Christian contribution to more robust ethics, "In a Christian worldview, justice is not enough. It's a good thing, but it's not the ultimate good that God has… if God is a just God, then we all get our just desserts.... but if God is a God of mercy and a God of grace, we have great hope… The gospel invites us to give of ourselves not only to be blessed, but to be a blessing. And so that framework is… a downward mobility approach where we are giving up the power that we have. We are giving up the privilege that we have... There's a spirit of generosity that comes out of the project of blessing. Justice is about accounting, right? It's about that numerical practice. It's about, 'Let's get the data exactly right.' Whereas blessing is like, 'We don't really need to keep track…' There's an openness and generosity. There's an indifference to equity… There's just a freedom in that blessing that doesn't come when you are trying to generate an accounting of justice."

On the expansive Christian vision of AI ethics, "I go back to the boundless generosity of God's love… The notion of equity or justice that is probably fundamental to artificial intelligence is a zero-sum game… but God, who is love, gives of himself and does make a sacrifice, and at the heart of that sacrifice is generosity and the willingness to bear the cost of sin and our recidivism."

On the larger narratives of the Christian faith, he says, "God is inviting us to pay attention to certain narratives… to draw together a story. And a story is different than a pattern. AI is drawing on patterns and it's only the interpreters of AI that can tell a story with it. But God is calling us to a certain future and putting that vision in front of us… How might the future that

God has for us be the data that drives our system? When I think about something like [God's] law, it's not simply, here's something to follow. It's like, here is what it's like to live in the presence of God. Here's what holiness looks like. Here's what righteousness looks like. When you are living as God's people, your lives will manifest love, joy, peace, patience… Those will be the data points that that will be true of you."

Anthony brings an aspirational voice to the AI ethics conversation. His stated role as an equipper invites Christians in technology to see their work in the broader context of blessing the world.

Martin is a computer science professor in a Little Ivy university. He currently attends a Baptist church but was not raised as a Baptist. On the materialist worldview of many of his CS students, he says, "A lot of students have bought into the underlying metaphysics of physicalism or materialism: that all that exists is a physical universe. And that being the case, the human mind is reducible to a mechanistic, deterministic, biological brain. And if you assume that the human mind is reducible to a brain, is reduceable to a complex computing device, then… computers can do anything that the human brain can do, and therefore, anything that your mind can do, because there's no meaningful distinction between mind and brain."

On his Christian worldview, he says, "I do believe that God created us as moral beings. That somehow, deep within us… [there is] a recognition of God, a need for God, a longing for God, and some knowledge there is right and wrong… even the people who most adamantly deny that there's any objective moral or ethical framework actually have a moral, ethical framework that they're trying to push on people… [So, I always ask] what is your foundation for why you claim this is a good? Why should this be a human right? Why is that an ethical foundation? And what is the groundwork? What's the framework? We have this unquestioned faith in science and

technology. And what that leads to is a belief that… if technology is going to save the world, then there should be no limits to the technologies that we create… So, one of the things I as a Christian want to do is simply begin by challenging that assumption."

On how to assess what's good or bad, he says, "The only [question] that most of [my students] ever ask is, how fast is an algorithm? How efficient is it? And my suggestion is… if your only question is how efficient is something, then we'd say nuclear weapons are really good because they are really efficient at destroying things and destroying people. Can we and shouldn't we apply that same questioning to computer algorithms and data structures? Does it harm society or is it helpful to society? Should you seek to create something because you can, or because you're trying to figure out if you can? Or before we even begin that, should we be asking… ought we even be looking at this?"

On the ethics of AI weapons and tool neutrality, he said, "One, I don't buy the argument that, well, if I don't do it, somebody else will so I might as well do it myself. Even if it's true that somebody else will, I still don't believe that that justifies me doing it. And second, I don't believe that tools are neutral… you can say a hammer is a neutral tool, but the moment you have a hammer in your hand, to use a very classic example, you become a different person because you start looking for nails. When I have a camera in my hand, the camera is not a neutral tool that I merely control. It's passive, but I am a different person when I'm holding a camera. I'm looking at the world differently… and it's really clear that when you're holding a gun in your hand, that you are a different person."

Martin embraces his role of educating the next generation of computer scientists with a broader, more ethical vision of their future. Working in a largely secular academic setting, he

doesn't proselytize but rather asks questions and challenges presuppositions that preclude a

Christian worldview.

James is an MD, an author, and a robotics researcher. He identifies as a non-denominational Christian. On worldview, "Worldview affects every area, not only of AI but of life." On hostility to Christians in STEM fields, "I was struck by how many Christians there are in the STEM areas, and interestingly, they talked about 'coming out' as a Christian. 'Coming out' as a Christian is a much bigger deal than coming out as gay or transgender… and it's often regarded as a career limiting move… I worked for thirty years as a Christian physician, lecturing about medical ethics in a very, very secular, hostile university in central London. And I learned the techniques of ducking and diving… you can see it in Jesus, that Jesus so often is asked a straight question and he doesn't give a straight answer. He ducks, he dives, he asks another question… because he's in a hostile environment."

On the bigger calling of the Christian role in AI ethics, he says, "The Christian approach is to try to see beyond the simplistic balancing of harms and benefits towards the deeper narrative that is going on here… The more we engage with these intelligent machines, the more it raises the question, 'What does it mean to be human?' I mean, people have been wrestling with that question since the dawn of time, but… every time as technology advances… it brings a new twist to that question of what it means to be human… The Christian understanding of being human is that humans are both at one and the same time, wonderful and remarkable… each human being is a masterpiece… and yet, hideously flawed, broken, fallen, and contaminated by evil. It is both… greater and worse than the secular understanding."

On the Christian view evil, "Christians have always taken evil very seriously. It's something that we recognize. It's something that we treat with great respect. It's something that we try to fight against. But we certainly recognize it as evil, as an extraordinary powerful force."

On the materialist view of evil, "Interestingly, in the world of the secular technologists, computer specialists, and so on, there really isn't a category for evil… I mean, they understand the category for programming errors and mistakes, and they have a category for just bad things happening, fate or disaster. But the idea of personal malevolence just doesn't make sense. It doesn't compute… What they have created is this category of the 'bad actor,' which is a very interesting phrase. So, most of us are decent, good, honorable people, like me, and then there's a few weirdos out there who are, quote-unquote, 'bad actors.' They're up to no good, so we've got to try and develop algorithms that can detect and prevent them. But again, that's a very simplistic understanding, isn't it? To have just a few rogue elements?"

On the uniquely deceptive nature of AI and the potential for idolatry, he says, "I think this is genuinely new… this issue of the simulacrum, and of how we should respond from a Christian point of view… I do think there is astonishing potential for deception. I think there's a capacity for self-deception. We willingly allow ourselves to be deceived. And it starts off as a bit of a game… this box that talks back to me... But I think it can easily morph into something that's darker… In the Old Testament, the distinction between the idols and the living God was that the idols couldn't hear, and they couldn't speak and they couldn't move… but the strange and worrying thing is that now we have idols that can do all those things… And so, they're much more dangerous as idols, much more powerful, much more deceptive… it does seem to me that in terms of the spiritual warfare, it's like a whole new front is opening up which hasn't been there before."

On the Christian view of redeemed humanity and human inventions, he says, "The extraordinary positive message that Christianity teaches is that these things which have an evil potential can be redeemed, they can be brought back out of the hands of evil and used for good. And so that concept of redemption, that technology needs to be redeemed, again, is a particularly Christian understanding."

James has written extensively on the Christian response to AI and robotics in a number of settings. His understanding of the dual nature of humanity (sinful and flawed by nature yet deeply loved and forgiven in Christ) informs his realistic yet redemptive vision for our approach to AI.

Discussion

Review of Thesis and Purpose

I began my study by proposing that humanistic ethical principles, even if codified into laws and regulations, are necessary but insufficient to ensure robust and beneficial AI. Further, I proposed that acknowledgment of divine intelligence, along with an ordinate understanding of human intelligence, is foundational to the development and use of artificial intelligence. Therefore, religious voices should have a say in framing the ethical scaffolding around it. Worldview diversity is critical to achieving true diversity and the Christian worldview was a gap in the Ethical AI conversation and literature. To fill that gap, I asked a *purposeful sample* (Merriam, 2016) of informants to offer their thoughts on what Christianity would add to the conversation. Specifically, I asked them **1) how worldview affects AI research and development, 2) whether a Christian worldview had anything unique to contribute to the discussion around Ethical AI, and 3) how AI ethics might be more robust and beneficial if we brought Christian teachings, texts, and traditions explicitly into the conversation.** I did not seek debate or argument, but simply opinions, attitudes, and perceptions. For this kind of research problem – what does the missing voice say, and can we learn anything from it? – the methodology of qualitative, semi-structured interviews worked very well, functioning as a sort of microphone for marginalized voices and amplifier for disenfranchised discourse. What I found confirmed my thesis in many ways, but also broadened my vision by presenting a new way of thinking about Ethical AI that I had not considered before: God is never described as merely ethical. Rather he is described with words like holy, just, merciful, loving, and righteous. Therefore, it's possible that from a Christian perspective, ethical is too small a word to

encompass the righteous life God calls us to. We should perhaps think of a new way to describe our aspirations for AI that aligns with God's aspirations for us.

In this section, I will attempt to connect the dots between my research questions and the data by discussing the data in context of the larger themes that emerged during the interviews. I will end with a segue to some reflections that articulate the direction I believe the data is pointing in terms of possible future steps for Ethical AI.

Discussion of the Data

Like other faiths, Christianity is not a monolith. While each of my informants identified as a Christian believer, there was predictable denominational and personal variation, so it was important to take these differences into consideration as I attempted to make sense of the data. As an intellectual exercise, I pondered what I would hear if I got all my informants in a room and started the conversation among them. Where would they agree and disagree? Would there be points of indisputable agreement? Areas of strong disagreement? Would they agree on matters pertaining to AI but not necessarily on matters pertaining to doctrine? Since I could not do this in person, I had to let the data do the talking, which was no small matter given that I had twenty-one hour-long interviews to work with. In a sorting task that involved condensing approximately two hundred fifty single spaced pages of raw interviews down to approximately forty double spaced pages of presentable findings, and then reducing those to a manageable set of representative quotes under thematic headings, I found myself looking for the most relevant common themes without ignoring areas of disagreement and interviews – or sections thereof – that were outliers in the mix.

Though my informants did not all speak on every subject, nor did they share identical views on every doctrinal or theological issue, they all identified as Christians and all but perhaps

one or two (I did not ask the specific question so cannot assert beyond my inferences) held a

traditional biblical worldview[3]. There was a good mix of Catholics (five) and Protestants

(sixteen), as well as an interestingly diverse array among the Protestants (one Quaker, two

Baptists, one Christian Reformed, one evangelical Anglican, one evangelical Presbyterian, one

Christian Missionary Alliance, two Pentecostal Christians, five non-denominational believers,

one who identified as nominally United Church of Christ, and one self-proclaimed "homeless"

Christian).

The five Catholic informants foregrounded the storied Catholic intellectual tradition,

bringing strong philosophical and historical perspectives to the discussion. They tended to favor

virtue ethics as their framework of choice. "Ethical principles and laws are not enough. That's

why you have to get people into the habit of doing the right thing. Inculcating virtue." (Mark) As

might be expected, none of the Catholics had a problem talking about human sin and evil. "Sin

isn't a very popular word. It doesn't show up in modern culture at all… it doesn't even show up in

certain denominations... But the Catholics have no problem with sin. We talk about sin all the

time!" (Mark) Perhaps because of their proclivity toward history and philosophy, the Catholics

brought the most compelling arguments for the Christian provenance of ethics.

The evangelical believers (which included the non-denominational and Pentecostal

informants) also spoke about the reality of sin and spiritual evil, but when speaking of virtue,

particularly as it pertains to ethics, they foregrounded the need for regeneration in Christ and the

empowerment of the Holy Spirit. "From a Christian worldview perspective, it's very easy to

settle on virtue as being the guiding principle… but we can't really be virtuous without the Holy

Spirit's work and his power within us…" (John) "We believe that renewal and transformation

---

[3] The view that God is sovereign, the Bible is the inerrant word of God, and a biblical epistemology informs the
Christian vision of knowledge, wisdom, and truth.

happen first from within through the action of the Holy Spirit on a person that changes them and gives them the power to behave 'ethically.'" (Christian) "Christianity is not principally a moral code. Although it *has* a moral code, it's to do with the relationship with God and the new power that enables us to live… The message of Christian salvation… is offering a firmer basis for ethics in that it gives people a power to obey…" (Jerome)

At least two informants fell into the category of what I would call liberal Christianity. One focused on ethics more from a political perspective than a religious one and another spoke to only very specific areas where he felt Christianity differed substantively from a secular ethics worldview. This is not to say that I claim to have fully understood their faith perspectives from their comments in a one-hour interview. Nor is it to say there mightn't have been others who leaned toward liberal theology or progressive politics. Rather, it is to say that politics was not part of the protocol, so unless it was brought up by the informant, or revealed itself via their comments, it was not part of the conversation.

As in other studies, many smaller themes specific to my informants' experience, expertise, and education came up during the interviews, but do not warrant space here. However, several larger themes emerged that are worth reflecting on in light of my research questions. The concept of worldview – which anchored my first question – leads off. After that, I address the second question about Christianity's unique contribution to ethics by contrasting the differences between a Christian worldview and a materialist worldview over six key themes, ultimately proposing that the Christian value-add comes down to a realistic understanding of what we're up against in the form of evil, sin, and idolatry, and a realistic understanding of where we find our ethical moorings in the form of transcendence, provenance, and love. I address the final question of how Christian teaching, texts, and traditions might make for a more robust vision of Ethical

AI in the Concluding Reflections section. Here, then is a brief discussion of the major thematic take-aways:

Worldview matters. My informants universally agreed that our approach to artificial intelligence is informed by our beliefs about the reality of the world. "Worldview affects every area, not only of AI but of life." (James) "Worldview influences virtually everything we do as human beings… including how we develop and use AI." (Dominic) "Of course worldview matters! Why is [a robot priest in a little cute Buddhist monk shape] imagined for a robotic application… dispensing wisdom? Did worldview enter into that? Absolutely!" (Timothy) Christians hold that a realistic and cohesive anthropology of intelligence (divine > human > machine) is an important underpinning of AI. When God is removed from the picture, our anthropology is incomplete. "I have kind of a framework for a biblical anthropology that I center around four kinds of relationships… [to creation, ourselves, others, and God] … If you put AI in that diagram, and you put us in the role of God, then we are really only concerned that the AI would serve us and rule over the things that we instructed it to rule over." (Peter) "Human intelligence is very low compared to the wisdom of God and there is a hierarchy that keeps men humble. The 'why' of humanity being made in the image of God… shapes the agenda for AI… Christians recognize the limitation of human intelligence… If human intelligence is inferior to God's wisdom, how much more inferior is artificial intelligence, created by man's intelligence? It really puts it in the proper place…" (Joan)

One big worldview question that arises when we try to simulate human intelligence with machines is what it means to be human, and this was a critical question to my informants as the Christian worldview differs dramatically from the materialist worldview. Are we, as Christians believe, created in God's image as deeply loved, unimaginably valuable, irreducibly complex

beings who have fallen away from God and need redemption in Jesus? Or are we, as materialists believe, remarkable accidents of natural selection, descended from other species, and waiting for (or even actively developing) our next evolutionary upgrade? "In a world that's strictly materialistic, humanity boils down to neurons and brain chemistry, but I suspect even the most diehard materialists have some sense that that's an insufficient anthropology… The tension that tech companies face is that they're aware that they're potentially threatening humanity, but they don't know what humanity is." (Peter)

While computational models of the human brain remain central to the science of AI and drive much of its current funding and research, my informants held that the brain is not merely a complex biological information processing unit but the inimitable epicenter of human intelligence and consciousness. "The underlying metaphysics of physicalism or materialism [is] that all that exists is a physical universe. And that being the case, the human mind is reducible to a mechanistic, deterministic, biological brain. And if you assume that the human mind is reducible to a brain, is reduceable to a complex computing device, then… computers can do anything that the human brain can do, and therefore, anything that your mind can do." (Martin) "If you get into the mindset of somebody who's not a Christian… and material is all that we have… then the human brain is simply a very complex, organic computer that we can replicate at some point… I think you'll find that a huge number of scientists don't buy into that because they understand the limits of what we have." (John) Therefore, most argued, a machine version of the brain will never be a mind, and computers will never attain artificial general intelligence. The contrast between Christian and materialist worldviews presents an inherent conflict in the telos of AI and reveals a difference in how we even conceptualize it much less deal with it once we've made it. The materialist is all for optimizing life on the planet because that's all there is. "The

Christian goal is not to optimize life on this planet. The goal is to encounter God… My goal is to say, how is God at work in this?" (Peter)

Evil. Many of my informants said, in some way or other, that at the heart of our collective felt need for ethics is the reality of a darker force: evil. Here, the difference between the materialist and the Christian worldview is quite interesting. For materialists, evil is a logical challenge. They know it exists, but they don't really know why, and they certainly don't believe in the Devil. My informants generally do believe in the Devil, and said, variously, that evil is personal, it is spiritual, it is deceptive, and it seeks to destroy or pervert both humans and their inventions. "As Christians, we already believe in a superintelligence that is actually malicious towards us, which is Satan… Satan has so much more power… and so much more intelligence than any human being, and actively works against God's will to destroy human civilization, to corrupt it, to dominate it… [and] to corrupt what human beings have made also." (Christian) "Christians have always taken evil very seriously… It's something that we try to fight against. But we certainly recognize it as evil, as an extraordinary powerful force… Interestingly, in the world of the secular technologists… there really isn't a category for evil… They understand… programming errors and mistakes, and they have a category for just bad things happening... But the idea of personal malevolence just doesn't make sense… What they have created is this category of the 'bad actor'… So, most of us are decent, good, honorable people, like me, and then there's a few weirdos out there who are, quote-unquote, bad actors." (James)

It is interesting to note that a concealed identity works to the Devil's advantage. French poet Charles Baudelaire captured the sentiment in a famous phrase, "La plus belle des ruses du Diable est de vous persuader qu'il n'existe pas!" (The Devil's finest trick is to persuade you that he does not exist.) Well before Baudelaire, the Apostle Paul reminded Christians that, "…we do

not wrestle against flesh and blood, but against the rulers, against the authorities, against the cosmic powers over this present darkness, against the spiritual forces of evil in the heavenly places." (ESV Bible, Ephesians 6:12) If this is true, as my informants believe it is, it explains a lot about why things go wrong in the world, and what will ultimately go wrong with AI. Ethical principles and laws may be good, but they are no match for evil.

Sin. According to many non-religious worldviews, humans are born good and, somehow, corrupted by "society." According to my informants, the Christian worldview tells us that all humans are born in a corrupted (or fallen) state and, though created to be good, have the capacity to be the so-called bad actor. "Why do we have legislation at all? It's a reflection of the fact that we're sinful people and we don't do what is right…" (John) My informants held that human sin is universal and is positional rather than behavioral. "The problem isn't well, we just have bad algorithms, or we just have injustice. It's actually that we have a sin problem… Bias or unfairness or racism or sexism… are real issues, and we have to deal with them, but I think the root issue is sin. It's a rebellion against God… our designer and creator." (Joseph) "Original sin is completely real. I think there's a very strong tendency of humans to ethically make bad choices…" (Benedict) "Any technology by itself is neutral, but it is the sinful nature of man that turns a tool that's neutral into evil." (Joan) "We're like, oh, the problem is technology… The problem is the algorithm. The problem is Facebook. The problem is the big companies. It's always a scapegoating somewhere. It's never us or ourselves." (Edmund) This concept is amply supported in scripture. "All we like sheep have gone astray, we have turned, everyone, to his own way." (ESV Bible, Isaiah 53: 6) "They have all turned aside; together they have become corrupt; there is none who does good, not even one." (ESV Bible, Psalm 14:1-3). "For all have sinned and fall short of the glory of God." (ESV Bible, Romans 3:23) In this context, contrary to

the generally accepted secular idea that things will be fine if the right people are in charge, Christians hold that things, including AI, probably won't be fine even if the so-called right people are in charge. Bias, injustice, unfairness, etc., are inextricable from fallen humanity and therefore inextricable from the tools that humans make. This isn't to say Christians have an inherently pessimistic or negative view of life. Quite the contrary, the gospel presents a hopeful view of how to overcome sin and evil, but still, we have to factor sin into the AI product roadmap.

Idolatry. The human tendency to imbue technology with "salvific" qualities leads to the human tendency to put faith in technology rather than God. AI arguably presents humans with the most compelling array of technically salvific qualities that they have encountered in recent history and is therefore a uniquely tempting replacement for God. For materialists, this is not a problem: there is no God. For my informants, and other Christian believers, this can be a big problem. "The claim of artificial intelligence in the secular world is… one day all your problems will be solved, and AI is the savior of the world… That would contradict what we as Christians believe, but I think it also contradicts what a lot of people believe now, even if you're not Christians… A lot of technology is idolatry in the sense of the way we designed it… am I causing people to love other things more than God?" (Robert) "What the Christian tradition has to say, and say very loudly is, don't let this become a substitute for either God or one another. We have a tendency to make an idol out of our technology… sort of… a substitute for God… to… somehow help us overcome death, overcome the human condition… It's looking for a new salvation and new eschatology in this technology and it's making it a do-it-yourself project." (Narcisa)

The warning against making and worshiping idols is ubiquitous in Christian scripture. The prophet Habakkuk sums up many of the Old Testament warnings saying, "What profit is an idol when its maker has shaped it, a metal image, a teacher of lies? For its maker trusts his own creation when he makes speechless idols. Woe to him who says to a wooden thing, Awake; to a silent stone, Arise! Can this teach? Behold, it is overlaid with gold and silver, and there is no breath at all in it. But the Lord is in his holy temple; let all the earth keep silent before him." (ESV Bible, Habakkuk 2:18-19) The prophet Isaiah literally mocks humans that worship things they make, saying, "He cuts down cedars, or he chooses a cypress tree or an oak… He takes a part of it and warms himself; he kindles a fire and bakes bread.… and the rest of it he makes into a god, his idol, and falls down to it and worships it…" (ESV Bible, Isaiah 44:14-17) And the Apostle Paul presents a dim view of idolatrous people saying, "For although they knew God, they did not honor him as God or give thanks to him, but they became futile in their thinking, and their foolish hearts were darkened. Claiming to be wise, they became fools, and exchanged the glory of the immortal God for images resembling mortal man and birds and animals and creeping things." (ESV Bible, Romans 1:21-23)

The temptation to idolize technology is real but this doesn't mean my informants believe artificial intelligence is, in toto, idolatrous. Many of them are actively researching or developing AI applications themselves. Rather, they recognize that AI's oracular attributes (its ability to see, hear, speak, answer questions, predict…) make it much more sophisticated and deceptive than a piece of wood or stone, and therefore increase the chances that those predisposed to do so will seek to put their trust in Google rather than God.

Transcendence. "If your worldview has no transcendent dimension, it could be simply, 'Well, if it can be done, it should be done…' Unless we acknowledge a transcendent dimension,

we're in real trouble…" (Jerome) "There's a worldview in the AI world and it's a worldview without a transcendence." (Narcisa) My general takeaway from these statements is that without belief in a transcendent God, or a humble acceptance of our finitude as created beings, hubris may lead us to harm. But that hasn't stopped us from trying to transcend without the Transcendent. To the materialist who has rejected the idea of religious immortality, AI represents an alluring alternate pathway to overcoming our biological limitations and even death. For an example of this, we don't need to look further than the title of Ray Kurzweil's 2005 book *The Singularity is Near: When Humans Transcend Biology*. Kurzweil is not alone among materialists – who do not believe in God or an afterlife – in longing for transcendence through technology. Transhumanist and singularity organizations (e.g., Humanity+, Singularity University, etc.) and the wealthy technopreneurs that fund them have spent millions trying to create their own version of immortality. To my informants, technology won't get us there. We can't "AI our way" to transcendence after death or even to transcending our troubles here. "[AI] can optimize certain things… but it doesn't really do anything for some of the inner turmoil that we're all dealing with… Our biggest problem in some ways is an internal one and I don't think AI is going to solve that." (Michael) Rather, they say, our pathway to peace on earth and eternal life in heaven is found elsewhere. "The message of the gospel is that your biggest problem has already been solved, right? Rest in me. That's the message." (Michael) If, as Christians believe, the grandest claims of AI have already been accomplished in the person of Jesus Christ, we will focus more on developing and using AI to bless humanity and less on trying to transcend our limits and build an AI stairway to heaven. "We don't have unlimited time. We don't have unlimited resources. So… [I would] engage in AI work that aligns with my priority, according to my worldview...

With limited time and limited investment, I do what Christians are called to do: to build up,

encourage, and bless." (Joan)

      Provenance. If data provenance is important for trustworthy and reliable AI, arguably,

ethical provenance is important for trustworthy and reliable AI ethics. My informants,

particularly those with expertise in philosophy and the Catholic intellectual tradition, hold that

the roots of modern ethical principles can be traced to Christianity, but that Christianity isn't

getting credit. "There are things about our current worldview that we take for granted, but they

weren't there before the Christians showed up. In other words, they weren't part of the pagan

world… There are so many things that we take for granted that basically society inherited from

Christianity… and have now become secular ideas… we owe the church a debt of gratitude for

what it was able to do in spreading these values along with the good news." (Mark) This

amounts, some might argue, to a sort of worldview plagiarism: copy-pasting God's ideas into

your ethics document and calling them your own. Other informants agree. "I think there is kind

of a modern, secular humanist anthropology that is roughly agreed upon and I actually think it

comes from the Bible" (Peter) "Even those [students] that are not really well grounded in a

religious tradition… still have the kind of a vague ethic that permeates our society, which I

believe is… an ethic that grows out of the Christian tradition…" (Narcisa)

      But why, some may ask, does it matter where ethics came from as long as we all

generally agree that qualities like love, joy, peace, patience, kindness, goodness, faithfulness,

gentleness, and self-control are admirable now? "They could say, yeah… it doesn't matter who

the messenger is, it's the message that counts." (Mark) Two reasons. The first is that God, the

creator of the universe, says over and over in the Bible[4] that the fear (reverence, respect, awe) of the Lord is the essential first step on the journey to knowledge, wisdom, and goodness. The second is that, as Christians contend, those qualities listed above are literally "fruits" of the Spirit of God: "But the fruit of the Spirit is love, joy, peace, patience, kindness, goodness, faithfulness, gentleness, self-control; against such things there is no law." (ESV Bible, Galatians 5:22-23) and that fruit can only grow on the vine: "As the branch cannot bear fruit by itself, unless it abides in the vine, neither can you, unless you abide in me. I am the vine; you are the branches. Whoever abides in me and I in him, he it is that bears much fruit, for apart from me you can do nothing." (ESV Bible, John 15:4-5).

This raises the question of whether societies founded on Christian ethics can sustain those ethics once God is no longer feared and Christianity is no longer the informing worldview. "It's an interesting question, whether [a culture founded on] Christian ideals can survive without Christian citizens." (Mark) History tells us no. Those societies that have tried to impose the Christian ethos of self-sacrifice on *other* people without Christ – e.g., calling on a materialistic version of altruism while excluding the Author of altruism – have typically fallen apart as sinful human nature inevitably takes over. The twentieth century alone is replete with examples of revolutionary societies founded on altruistic yet atheistic principles. In less than a hundred years, most of those experiments in God-less altruism (e.g., the USSR, Southeast Asia, Eastern Europe, Latin America) have failed. Christians argue that the same things that happen with cultures made by humans will also happen with technologies made by humans. If the secular vision of Ethical AI demands personal sacrifice for the common good, the Christian apologists among my

---

[4] Proverbs 1:7, Proverbs 2:1-15, Proverbs 3:7, Proverbs 9:10, Proverbs 10:27, Proverbs 14:27, Proverbs 15:33, Proverbs 19:23, Psalm 111:10, Isaiah 33:5-6, and the entire chapter of Job 28, ending with, "Behold, the fear of the Lord, that is wisdom, and to turn away from evil is understanding."

informants answer that without God there is no logical reason for helping others at your own expense. "It's basic Christian morality that makes the difference… Christian altruism, putting oneself out for others… doesn't fit in with the evolutionary picture at all... If there's no God who watches and will judge, then there's no argument left… It's not that we can safely get God out of the picture and carry on and it's all going to be okay… Without God, there's no real reason to behave." (Jerome) "Part of the dissatisfaction with [secular] ethics… stems from the fact that there are no Archimedean points[5] in terms of context-free, interest-free, personality-devoid, statements of principle." (Timothy) If we're intellectually honest about the provenance of ethics, many of my informants argue, we have that Archimedean point in Christianity. If we need humans in the loop for robust AI, they might say, we need Jesus in the loop for robust AI ethics.

Love. In the end, my informants' view of the basis for AI ethics was built upon the central value proposition of the Christian worldview: love. "We're people who are driven by our hearts... and the way to the heart isn't necessarily through the brain. We aren't primarily thinking things, we're primarily loving things…" (David) Certainly, we should aim for justice, fairness, equity, transparency, dignity, and the like, but those are by-products of what St. Paul calls, in the last verse of 1 Corinthians 12, "a more excellent way." He then gives his magnum opus on love – which is essentially the Christian primer course on ethics – in 1 Corinthians 13:

> Love is patient and kind; love does not envy or boast; it is not arrogant or rude. It does
> not insist on its own way; it is not irritable or resentful; it does not rejoice at wrongdoing
> but rejoices with the truth. Love bears all things, believes all things, hopes all things,

---

[5]Or Punctum Archimedis: A hypothetical viewpoint from which certain objective truths can perfectly be perceived (also known as a God's-eye view) or a reliable starting point from which one may reason. (From Wikipedia)

endures all things. Love never ends… So now faith, hope, and love abide, these three; but the greatest of these is love. (ESV Bible, 1 Corinthians 13:4-8, 13)

In short, Jesus values our minds but optimizes for our hearts. "Attention is what [My Company] considers love… How much do you like [My Company]? How much do you want to use [My Company] in the future?... [My Company] will never optimize your love for another thing, so to speak, and God will not either." (Robert) "Maybe intelligence is not our greatest virtue, you know? Maybe what we need to be enhancing is love, but what we are going for is enhancing intelligence. And when we separate those two from each other, it's often when we tend to get into trouble. It's not the pathway. It never has been. Our pathway to divinity is through love. It's not through intelligence." (Narcisa)

According to my informants, love is better than ethics. It transcends the realm of legalism – the watchful mother or the hall monitor – and brings us to the realm of unlimited and creative blessing. "[Ethics] gets cast, often gendered, I would say, as kind of… maternal prohibition… I really think it's important that ethics take and infiltrate the creative imaginative side and say that really the problem in many cases is that there's a lack of imagination." (Timothy) "I go back to the boundless generosity of God's love… God, who is love, gives of himself and does make a sacrifice, and at the heart of that sacrifice is generosity and the willingness to bear the cost of sin and our recidivism." (Anthony) "The royal law of love, unlike all the other laws, is not proscriptive. It is actually generative. It's creative. It unleashes. To love your neighbor as you love yourself is something that is above ethics. Like, you can do that in a million ways." (Christian)

In the end, my informants confirmed my thesis that worldview makes a difference, and that the Christian worldview had a significant contribution to make to the conversation on

Ethical AI. In fact, it already had, but because those who hold an aggressively secular worldview currently control the slider on the Overton Window (See Figure 5) of acceptable discourse on Ethical AI, Christian perspectives have been relegated to the outer margins while materialist perspectives occupy the center. In part, the findings of this study may serve to readjust the settings to some degree, but it won't be from the top down. Christianity has lived quite comfortably on the "radical" fringe of the Overton Window for millennia largely because followers have found the teachings of Jesus to be radical in a good way. For deeply theological reasons, Christianity will never be "popular" in this world, but contrary to secular propaganda, it only seeks to become "policy" in the hearts of believers who then influence the larger culture.

Figure 5

Overton Window
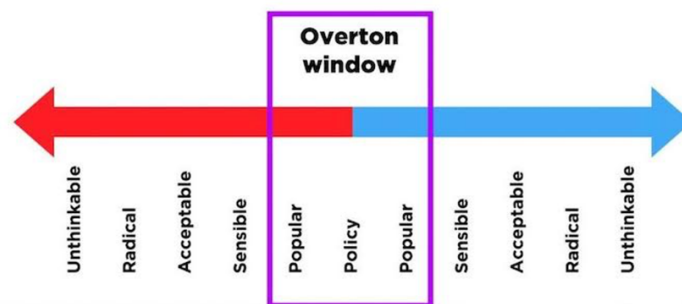
Graphic © Toronto Guardian, 2019; Concept © Joseph Overton

Figure 5 - Overton Window

As to broadening my vision, my informants reminded me of the novel framework that epitomizes Christianity: while we are called to obey laws for our own good and the good of others, we are no longer slaves to the law but recipients of Christ's righteousness. St. John explains how Jesus levels up the law: "For from his fullness we have all received, grace upon grace. For the law was given through Moses; grace and truth came through Jesus Christ." (ESV

Bible, John 1:16-17) And St. Paul says he wants to be found in Christ, "not having a

righteousness of my own that comes from the law, but that which is through faith in Christ, the

righteousness that comes from God on the basis of faith." (ESV Bible, Philippians 3:9)

Christianity challenges us to reflect on the qualities we truly value, in ourselves, in others, and in

the tools we make, and to put our priorities in order. To illustrate this, I've suggested a series of

"greater than/less than" expressions (See Figure 6), but boiled down and reduced, Christianity

calls us to look beyond our idea of mere human ethics and toward the righteousness of Christ.

Figure 6

Findings Expressions

Divine > Human
Human > Machine
Augmented > Artificial
Mercy > Justice
Love > Intelligence
Righteous > Ethical

Figure 6 - Findings Expressions

Concluding Reflections

Abolish religion if you like. Throw everything on secular government if you like. But do not be surprised if a machinery that was never meant to do anything but secure external decency and order fails to secure internal honesty and peace.
<div align="right">*G.K. Chesterton (Daily News, 1905)*</div>

So, the Christian worldview proposes to bring a realistic understanding of the things we're up against in the battle between good and evil (spiritual evil, human sin, and the human tendency to replace God with idols) as well as a realistic understanding of where we find our moorings in the search for robust ethics (a transcendent being who gave us the algorithm for ethics in the form of love). But what might it look like if we brought Christian teachings, texts, and traditions explicitly into the Ethical AI conversation? Specifically, what does this study contribute to the conversation and the literature on Ethical AI? My informants had significant, almost universal agreement that the Christian faith forces us to reevaluate our assumptions and look beyond the materialist idea of ethics and toward a creative righteousness that cannot be accomplished by our own will and power. "The Christian ethic is much more robust than people think… it's the idea that you put off the old and you put on the new… there's this watershed moment of when I become a believer, I'm a new creation in Christ." (Joseph) "[There] is a major misunderstanding of what the Christian message is because many people think of it in terms of rules and regulations… to follow in the hope that one day, God will accept you. And that gives you the kind of moralistic approach to religion… which is not Christianity…" (Jerome)

It's not that we don't need ethical principles, regulations, and laws. Certainly, all are necessary for civil society and governance of technologies as powerful as AI, but none have proved historically nor biblically sufficient to vanquish the reality of spiritual evil and the effects of human sin that make their way into our technologies. Rather, my data suggests that the

Christian contribution is visionary. It calls us to look beyond ethics and go, in the words of C.S.

Lewis, "further up and further in." (Lewis, 1956)

Beyond Ethical AI

My informants articulated this idea in various ways: "Obviously laws and regulations are

necessary… But what goes beyond?... The way we build technology is just an extension of how

we live our lives, and in many ways of who we are, so ethics then becomes an issue of identity as

opposed to a list of things that you should or should not do. So, what if, instead of asking, how

could things go wrong, which we still need to ask, we ask how does this technology speak of

who we are?" (Edmund) "The mercy and grace of God and the fundamental tenets of

Christianity… I think those things are very hard to be captured by ethics… Christianity is more

than just a bunch of rules, right? I believe that Christianity is more than just ethics. It's more… a

kind of correct way to live your life." (Robert) "And so our unique concern isn't merely ethics,

which is more of the domain of… the good of humanity. It is very specific. It's peculiar. It's the

kingdom of God, which is more than the common good. There is something unique about the

belief in the new creation that Jesus has inaugurated through his death and resurrection… We

can participate in the common good, because we believe it's a witness to the kingdom, but it's not

really an end in itself… The Christian ethic… is… something more. It's glory. It's vitality. It's

life, and it's overflowing life. And so, I think that ethics and rules… are never going to match up

to that kind of vital force in a person's life… that vitality is something that's far greater…

(Christian) "I keep saying… you have to think… about ethics as a way of pointing us toward

better futures, helping us to imagine how the world can be… holding up this vision of what's a

better world." (Dominic) "Technology has a view for what the future looks like… and a lot of

the stories that are being told about technology, including about AI, are the stories that Christianity tells about what the future looks like and where we're moving towards." (Anthony)

Reflecting on this data, we have to consider for a moment just how the Christian ethic – that vital force and vision of the future – differs from the so-called secular ethic. In a secular vision of ethics, it's all about social justice: no sin goes unpunished, and nobody gets more than they deserve. In the Christian vision of ethics, it's all about biblical justice: in Christ, all our sins are forgiven, and through Christ we get more than we deserve. "In a Christian worldview, justice is not enough. It's a good thing, but it's not the ultimate good that God has… if God is a just God, then we all get our just desserts.... but if God is a God of mercy and a God of grace, we have great hope… The gospel invites us to give of ourselves not only to be blessed, but to be a blessing." (Anthony) This is the message of the gospel of grace: in radically unfair trade, Jesus took our bad and gave us his good. In doing so, he shattered legalism with forgiveness, and brought us out of the realm of ethics, principles, regulations, and laws, and into the realm of mercy, grace, blessing, and love. As St. Paul explains, "For our sake he [God] made him [Jesus] to be sin who knew no sin, so that in him we might become the righteousness of God." (ESV Bible, 2 Corinthians 5:21)

Two of my informants weighed in on what it might look like to go beyond ethics and where we might start. "God says, seek first the kingdom of God and his righteousness. And it just shows how much God cares about righteousness." (Joan) "When it comes to technology, we need to not be seeking power, we need to be seeking goodness or ultimately holiness if we want to phrase it biblically speaking, because without that, power will end up destroying us… Ultimately… we need to make technology holy. It has to be better than good. It has to be better than beneficial. It has to be holy." (Benedict) It is the concept of righteousness that brings

something new – and, I'll argue, valuable – to the conversation on Ethical AI. According to the

Christian worldview, we need more than Ethical AI. What we need is Righteous AI.

Toward Righteous AI

While "Ethical" AI remains the generic term, we currently see artificial intelligence being

marketed variously as Trustworthy AI, Accountable AI, Reliable AI, Private AI, Equitable AI,

Human-centered AI, and Responsible AI. Based on the findings of my study, I propose we

explore the concept of Righteous AI, and I'll admit right now, I don't know what that is yet. I do

know it will be a tough sell in a secular world, not only because righteousness is loaded with

religious connotations, but because the average person would have a difficult time defining it. If

we can get past the religious baggage of the word, we might agree that the qualities of

righteousness express, more fully, the qualities we long for in our quest for ethics. So, what is it?

Online dictionaries define righteousness as being morally right or justifiable, free from

guilt or sin. Under synonyms, we find words such as ethical, good, honorable, just, moral,

upright, and true. This sounds very much like a section of St. Paul's letter to the Philippians:

"Finally, brothers, whatever is true, whatever is honorable, whatever is just, whatever is pure,

whatever is lovely, whatever is commendable, if there is any excellence, if there is anything

worthy of praise, think about these things." (ESV Bible, Philippians 4:8) In fact, the Bible is

replete with verses about righteousness because it is a key to the character and nature of God:

- But the LORD sits enthroned forever; he has established his throne for justice, and he
  judges the world with righteousness; he judges the peoples with uprightness. (ESV Bible,
  Psalm 9:7-8)
- The way of the wicked is an abomination to the LORD, but he loves him who pursues
  righteousness. (ESV Bible, Proverbs 15:9)

- To do righteousness and justice is more acceptable to the LORD than sacrifice. (ESV Bible, Proverbs 21:3)

- In the path of righteousness is life, and in its pathway, there is no death. (ESV Bible, Proverbs 12:28)

- Whoever pursues righteousness and kindness will find life, righteousness, and honor. (ESV Bible, Proverbs 21:21)

- Blessed are those who hunger and thirst for righteousness, for they shall be satisfied. (ESV Bible, Matthew 5:6)

- Seek first the kingdom of God and his righteousness, and all these things will be added to you. (ESV Bible, Matthew 6:33)

- For I am not ashamed of the gospel… For in it the righteousness of God is revealed from faith for faith, as it is written, The righteous shall live by faith. (ESV Bible, Romans 1:16-17)

In these verses, we see righteousness associated with justice, uprightness, love, life, kindness, honor, satisfaction, fulfillment, and faith; all things that would be hard to argue with if they were core attributes of a technology that looks, sounds, and acts like us.

So, what does Righteous AI look like in a high-tech world driven by profit and power? For a start, let's talk about profit, which often manifests itself as an excessive preoccupation with money. We can sum up Jesus' view on the subject by looking in the Gospel of Matthew: "Do not lay up for yourselves treasures on earth, where moth and rust destroy and where thieves break in and steal, but lay up for yourselves treasures in heaven, where neither moth nor rust destroys and where thieves do not break in and steal. For where your treasure is, there your heart will be also." (ESV Bible, Matthew 6:19-21) St. Paul also talks about money, focusing on the destructive nature of greed. Most of us know the famous verse, "For the love of money is a root of all kinds of evils. It is through this craving that some have wandered away from the faith and pierced themselves with many pangs." (ESV Bible, 1 Timothy 6:10) But not many know the verse right

after it: "But as for you, O man of God, flee these things. Pursue righteousness, godliness, faith, love, steadfastness, gentleness." (ESV Bible, 1 Timothy 6:11) In terms of challenging the presiding Killer App ethos in Silicon Valley, Paul is suggesting, in so many words, that we stop chasing unicorns and pursue righteousness instead. This also is a hard sell in a secular world, but it seems many decisions that lead us to decry unethical AI could be addressed if we prioritized righteousness over ROI.

As for power, how might Righteous AI challenge or change this dynamic? As a thought experiment, I asked myself what AI applications might look like if the teachings of Jesus informed development decisions and business strategies. Righteous AI would be powerful but restrained, holy but humble, mighty but merciful, and all-knowing but all loving. It would optimize for the good of others, prioritize peace, and solve the trolley problem by leaving ninety-nine sheep just to save one. Righteous AI algorithms would direct us to love our enemies and pray for those who persecute us rather than heaping fuel on the fires of social media outrage. (See summary in Figure 7) And the Righteous AI business plan would tell us not to be anxious about earnings, but to trust that God will take care of us because he knows what we need. In other words, Righteous AI would be utterly counterintuitive to today's hyper-competitive, efficiency-driven, utilitarian-focused technical culture.

Figure 7

Qualities of Righteous AI

Righteous AI would be:	Righteous AI would:
Powerful but restrained	Optimize for the good of others
Holy but humble	Prioritize peace
Mighty but merciful	Love enemies
All-knowing but all loving	Confound the trolley problem

Figure 7 - Qualities of Righteous AI

And that's not even considering what the technical specifications might look like. Some of my informants addressed the difficulties there. "Christianity… is self-actualizing through self-giving love and that's something that cannot be represented in a neural network… There's no way of quantifying the interiority of human self-gift… it doesn't register as representable to a computer in some sort of big dataset… Social psychology tries to quantify these things, but then it often ends up with rather shallow definitions that don't get at the richness of these experiences and these realities in human life… but within Christianity… the transcendence of self-gift… the wounds of Christ [show] what utter compassion for and self-gift to a fallen world looks like." (Jonas) "If we adopt more of the character of Christ, we'll learn the right things to do even in the face of things that are very uncertain right? Now, how does that play into AI?... I can ask Google… the consensus for a given question, but it may not have optimized towards mercy, or kindness or wisdom… I wouldn't know how to optimize for mercy because mercy is almost illogical… Tit for tat doesn't sound like something Jesus would teach. Jesus would say turn the other cheek." (Michael)

This is a profound issue for AI which excels at pattern recognition and statistical prediction based on past data, but not discernment and wisdom based on divine inspiration and revelation. Miracles are not data-driven and so today's prediction engines would get almost every story in the Bible wrong. The question of how to make Righteous AI is perplexing, both conceptually and technically, but perhaps it is the wrong question. Perhaps the question is more one of how we make Righteous Humans. Christian texts and teachings tell us we cannot be righteous on our own, we can only be *made* righteous through the atoning sacrifice of Jesus. Even then, even though redeemed, we're still in a fallen state so we'll only ever be asymptotically approaching righteousness on earth. But we have new motivation and new power.

The Bible tells us to model our lives on the righteous life of Jesus. Jesus never told us to do anything he didn't do himself. In the same way, righteousness is not simply something we tell other people to do. It's something we choose to embrace and live out for ourselves, in faith, and hope that others follow suit. It's a long quote, but Anthony sums it up best:

> That framework is… a downward mobility approach where we are giving up the power that we have. We are giving up the privilege that we have... There's a spirit of generosity that comes out of the project of blessing. Justice is about accounting, right? It's about that numerical practice. It's about, 'Let's get the data exactly right.' Whereas blessing is like, 'We don't really need to keep track…' There's an openness and generosity. There's an indifference to equity… There's just a freedom in that blessing that doesn't come when you are trying to generate an accounting of justice… God's inviting us to pay attention to certain narratives, certain points of information, to draw together a story. And a story is different than a pattern. The AI is drawing a pattern and it's only the interpreters of AI that can tell a story with it. But God is calling us to a certain future and putting that vision in front of us… if we can create the future data set, rather than a historical data set and that future data set could point us in the direction God has for us, what could that look like? What would be in that data? When I think about something like the law, the law is not simply, here's something to follow, it's like here is what it's like to live in the presence of God. Here's what holiness looks like. Here's what righteousness looks like. When you are living as God's people, your lives will manifest this kind of data. Your lives will manifest love, joy, peace, patience… they'll manifest care for your parents. Respect for God. Not stealing from your neighbor, not coveting your neighbor, not murdering your neighbor. Those will be the data points that that will be true of you." (Anthony).

Righteous AI Applied

I cannot sum up a Christian foundation for Righteous AI any better than Anthony. But how might we build on that foundation? The term itself was not something I had in mind when I went into this study. In fact, the idea of Righteous AI came out of my analytic reflection of the data, so I have not scratched the surface of what it would look like or how it could be reified. I did, however, ask my informants how Ethical AI might be more robust and beneficial if informed by Christian teachings, texts, and traditions. While I have attempted to lay out some general attributes and characteristics in the Discussion and Concluding Reflections sections, it makes sense to at least speculate on how Christianity might inform some specific questions that AI scientists, practitioners, and "victims" wrestle with.

I begin with two caveats. The first is a reminder that Christianity is not a monolith. Within this large and glorious faith, there are many doctrinal and practical disagreements both among denominations and within them. This vexes some humans who see the messiness and tension among Christians as a reason to discredit God. God, however, is quite comfortable with messiness and tension. He provides some instructions but inexplicably allows humans – his original AI – to stumble and fumble in their attempts to get things right and loves them even when things go wrong. Christians have also become quite comfortable with messiness and tension because we know we live in an interstitial space; a place we've come to call "the already but not yet." The fact that we can't explain everything doesn't deter us from believing in a God that can. The second caveat is that, obviously, I do not speak for all Christians, let alone all the Christians I interviewed in this study. In fact, I don't *agree* with all Christians, let alone all the Christians I interviewed in this study. Nor they with me. See messiness and tension… But since I raised the questions and placed myself in a role of meaning-maker, it is incumbent on me to take

a stab at applying what I heard, in context of what I believe, to some actual issues. I am eager to interrogate them in further research but for now, here is my take on how we might apply a sort of "what would Jesus do" filter to some common AI dilemmas.

If Christians were tasked with developing some form of Righteous AI, they might begin the product development cycle by asking a critical question: "Should we even build this?" The answer to that question may be, "No, it's a bad idea. We should not," but the next question will invariably be, "What if someone else (i.e., a known aggressor) is building it?" Believing sin and evil exist, the Christian says, maybe we must build it, but it will always be in humble submission to God, and with stringent restrictions. Let's start with a hot button topic: killer robots.

Autonomous weapons. On this application of AI alone, there is wide disagreement among Christians, and therefore, wide disagreement on what a Righteous AI solution might be. Many Catholic believers, who are unapologetically "pro-life" from womb to tomb, advocate for a total ban.

> The Vatican has been heavily involved… on the question of whether or not there should be a global ban on autonomous weapons and the church's position there has been consistent and firm, just as it has been on nuclear weapons, and know they're one of the strongest voices advocating for a total ban. There won't be a total ban. Politically, that's just never going to happen, but… the Catholic church goes out of its way to be a voice on these kinds of issues. (Mark)

Mark was not my only informant who was wary of autonomous weapons, mainly because of their unique ability to kill large numbers of humans and kill them indiscriminately, but perhaps less obviously and more profoundly, autonomous weapons serve to remove human agency from the decision to end another's life and in the process, over time, numb the human

conscience (Christians would call it the voice of the Holy Spirit) and desensitize people to the weight and enormity of those decisions. This amounts to spiritual harm and therefore, makes autonomous weapons particularly dangerous to humanity. Other Christians (not in this study) have spoken and written in *defense* of so-called killer robots and see things largely through a "peace-through-strength" lens: we know other entities are developing autonomous weapons; if we don't, we will be unable to protect ourselves from the threat of aggression or mass destruction. We have a mandate and a right to protect ourselves, our families, and our societies, so it is not morally wrong to build and use, if necessary.

So where might we land, then, considering a "whole Bible" approach to Christian ethics that includes God's righteous anger and "just wars" against evil as well Christ's admonishment to personally turn the other cheek? Theologians over the centuries have wrestled with this and have not managed to convince all Christians to agree. What we do agree on is that Christians are called to mitigate evil as God has placed us in our various spheres of influence. Since autonomous weapons already exist and are in use today, the job of the Church is to be a voice of truth according to our consciences, stand athwart evil, and act as peacemakers even while seeking to protect the weak and vulnerable. Some Christians will be focused on making peace with diplomacy and without weapons. Others will engage directly with AI to ensure that weapons are never designed to be fully autonomous, and that humans always make the final decision. Through it all, our mandate is to pray for wisdom as the world makes and uses technological weapons, including autonomous ones, and continue in our calling to be salt and light in a dark world.

Moral machines. Killer robots might seem the obvious choice for a restricted application of AI, but for Christians, there are other, more benignly marketed machines that we should stand

athwart as well. While autonomous weapons are de facto life-and-death decision-making tools, other systems, like self-driving cars and algorithmically driven adjudication tools, could also end in life-or-death outcomes (literally and figuratively) and should be viewed, made, and used cautiously. The Christian worldview holds that humans have a divine mandate for moral authority and moral responsibility that we cannot delegate to something else. Therefore, we should not work toward so-called "machine ethics and norms." Autonomous vehicles cannot make truly moral decisions, but only act on ultimately limited algorithmic instructions. Crowd-sourced wisdom does not reflect divine wisdom. It's merely a dumbed down, situationally positioned, mathematically averaged version of human wisdom which gives equal weight to the wise and the foolish and ignores the absolute and transcendent. By nature, computing devices that make data-driven decisions depend on utilitarian ethical frameworks which are antithetical to Christian ethics: "self-giving love cannot be represented in a neural network" (Jonas) and "I wouldn't know how to optimize for mercy." (Michael) Of course, there are things computers can do – especially if empowered with AI technologies – that can help humans in calculating, sorting, and pattern recognition tasks, but a machine cannot discern among less literal, more intangible factors to make decisions that only humans, made in God's image with God-given faculties, should make. Therefore, Righteous AI will always keep humans in the loop and never be used exclusively for anything that requires human judgment, discernment, or wisdom. Again, humans will always have the final word. This means that Righteous AI might modify, if not even eliminate, many things AI is already being used for, opaquely and behind the scenes, in areas like finance (loan decisions), criminal justice (sentencing recommendations), and insurance (risk profiles associated with race rather than individuals). Although statistics can help us understand

and mitigate risks, machines can and will never be able to "do justice, love kindness, and walk humbly with God." (ESV Bible, Micah 6:8)

Artificial versus augmented intelligence. Even among secular AI researchers and practitioners, fringe elements aside, there is an inkling that so-called artificial intelligence will never truly replace human intelligence and what we are really working on is augmented intelligence. The human brain is complex and inimitable. It cannot, as Marvin Minsky, Nathaniel Rochester, and Claude Shannon proposed at Dartmouth in 1956, be "so precisely described that a machine can be made to simulate it." Incredible scientific advances notwithstanding, AI will never be as incredible as people are. While it doesn't take Christian belief to agree that humans are something special, Righteous AI takes us back to the idea of rightly ordered intelligence: divine > human > machine. Christians, acknowledging that God is the true superintelligence of the universe, would lobby for a reality check in our representations of AI and present a realistic vision of our mandate to promote human flourishing (being sure to define this in spiritual terms as well as material terms) rather than chasing after a machine superintelligence. As one of my informants said:

> Instead of throwing my time and limited R&D resources… to enable an artificial super intelligence with the agenda that it would create an artifact of a machine that is superior than man… my worldview would say, where is its proper place? Can man create a machine that is… superior to humans created by God? My answer is no. So, I wouldn't spend time there, but instead I would spend time on using AI to bring better lives, to solve the world's problems, to make this place a better world that aligns with the 'why' of the mandate of humanity which is to subdue the earth. (Joan)

Righteous AI would focus on augmenting and complementing human capabilities rather than attempting to create a hubristic vision of artificial intelligence because, in the end, we can't. As my informant Peter said, "I think our technology is meant to stretch our constraints and move our boundaries but not to transcend them… I don't think we *can* transcend our constraints… If they're actually constraints, they're imposed by a third party, not by us, so in some ways, we don't have to be careful not to transcend to them." Artificial intelligence is part of God's divine mandate for humans to explore, discover, and have dominion over the earth, but not a means to become gods unto ourselves. It is yet another tool that augments, enhances, and otherwise expands our capabilities, but will never replace us. As my informant Timothy said, "The Magic 8 Ball is a great conversational aid; it just has no internal intelligence… I don't think AI will ever lose part of its Magic 8 Ball quality of needing to have human reception and human creative interpretation give its product life… it will take a lot to convince me otherwise."

In the end, while Christians are called to be part of all intellectual and creative pursuits on earth, including AI research and development, we should not be seeking to attain God-like qualities or attempting to transcend our human limits through AI. That said, we shouldn't be too worried if others do because we believe God has instituted protections (in the form of boundaries and limits) that he will not let us "accidentally" transcend. He is smarter than AI.

These attempts to apply Christian ethics to specific AI questions are by no means authoritative or exhaustive, and there are plenty of specific AI questions I did not address here, but many of my speculations are in line with what other Christians have written on the subject. This study represents my initial Christian foray into the Ethical AI conversation in an academic setting. I hope some of the wisdom herein will find its way to a larger audience and begin to

expand the dialog between those who follow Jesus and those who don't. To summarize my

thinking on the Christian voice in Ethical AI in so many points:

- Christians are called to represent Jesus in the world as God has placed us in our various
  spheres of influence. We are to be a voice of truth, stand athwart evil, and act as
  peacemakers even while protecting the weak and vulnerable.
- Our mandate is to pray for wisdom as we make and use powerful technologies, and to be
  salt and light in a dark world.
- Humans have a divine mandate for moral authority and responsibility that we cannot
  delegate to something else. We should not work toward machine ethics and norms or
  fully autonomous agents.
- Crowd-sourced wisdom does not reflect divine wisdom. It gives equal weight to the wise
  and the foolish and ignores the absolute and transcendent.
- Righteous AI would never be used exclusively for anything that requires human
  judgment, discernment, or wisdom.
- So-called artificial intelligence will never truly replace human intelligence. What we are
  really working on is augmented intelligence.
- While Christians are called to be part of all intellectual and creative pursuits on earth, we
  should not be seeking to attain God-like qualities or attempting to transcend our humanity
  through AI.

Thoughts on Further Research

The fact that I was able to meet and talk at length with a cohort of brilliant and thoughtful

Christians who are all exploring similar topics on their own is a promising entrée to further

research. Several have written books on various aspects of AI and Christianity. Others are

working at Christian universities. Still others have left major tech companies to develop their

own AI applications that align with their Christian worldviews. And of course, there are those

still working at public and private universities and companies where they live out their faith in

secular environments. The connections I made conducting this research suggest many possible

avenues of fruitful collaboration on Righteous AI and how it might be approached both technically and commercially. I hope to address this in future research.

Other areas of interest for future exploration include a deeper dive into the correlation between human learning and machine learning from a spiritual perspective as well as a broader look at the landscape of ethics education in computer science programs. Several of my informants lamented the dearth of ethics courses and noted the disturbing imbalance between technical training and ethical training. David, who is a Christian but went to a large, secular university, had no ethical training in school. There is talk of change, he says, but the embedded worldview imbalance remains. "A lot of schools in the secular academy… are pushing to add a mandatory ethics course alongside the standard computing curriculum. And I think that's good. That's better than nothing. But I would say that the premise that you teach everyone in a certain way of thinking about the world, and then you slap on an ethics course, doesn't make an ethical person… When I was an engineer in the mid-nineties, I distinctly remember sitting in a cubicle farm as an engineer with these forty technical courses under my belt wondering, how does this connect to my Christian faith?" Martin, who now teaches computer science at a secular university, reflected on the gap as well, saying, "We had a sort of a *day* in class devoted to the question of ethics in computing…" He is now working on developing a course that not only talks about the question of ethics in computing but has graded assignments on the subject. And Benedict observed the gap between demand and supply, saying, "We need to be thinking much, much more about training in terms of ethics and helping people get better at making the right decisions… Right now, we have hundreds of Christian ethicists or Christian theologians thinking about these things. We need to have tens of thousands… the demand is there and we're not meeting it." How might the Christian worldview impact curriculum in this area?

I'm also fascinated by St. Augustine's idea of ordinate (rightly ordered) loves, about which he wrote in his seminal work *The City of God*:

> And thus beauty, which is indeed God's handiwork, but only a temporal, carnal, and lower kind of good, is not fitly loved in preference to God, the eternal, spiritual, and unchangeable Good. When the miser prefers his gold to justice, it is through no fault of the gold, but of the man, and so with every created thing. For though it be good, it may be loved with an evil as well as with a good love: it is loved rightly when it is loved ordinately; evilly, when inordinately… But if the Creator is truly loved, that is, if He Himself is loved and not another thing in His stead, He cannot be evilly loved; for love itself is to be ordinately loved, because we do well to love that which, when we love it, makes us live well and virtuously. So that it seems to me that it is a brief but true definition of virtue to say, it is the order of love; and on this account, in the Canticles, the bride of Christ, the city of God, sings, "Order love within me." (Augustine, 426 AD)

I hinted at the idea of an "ordinate understanding of human intelligence" in this study but did not explore it. In future work, I would love to pursue a variation of Augustine's ordered loves by probing the topic of "rightly ordered intelligence" (divine > human > machine) in more detail. Since machines can do certain tasks better, faster, and more accurately than humans, it's a small jump for some to believe that ultimately, they'll be able to do every task better, faster, and more accurately than humans. There is much to explore on the spiritual impact of this belief.

Finally, regarding publication of my work, as a Research Fellow at AI and Faith, I am eager to disseminate this study to a broader audience and encourage people of faith in other religions to conduct research and publish their perspectives as well. For a start, I will focus on taking parts of this dissertation and re-working them as articles for submission to various

journals, think tanks, and organizations that foreground Christian scholarship, possibly including the Center for Global Christian Studies here at the University of Washington, the Walter Bradley Center for Natural and Artificial Intelligence at Baylor University, The Kirby Lange Centre for Public Theology at the University of Cambridge, and the Veritas Forum at Harvard.

As for broader future research in this arena, AI and Faith plans to expand capacity for original research in partnership with other centers and institutions around the U.S. and UK.

References

AI100, (2021, September). Michael L. Littman, Ifeoma Ajunwa, Guy Berger, Craig Boutilier, Morgan Currie, Finale Doshi-Velez, Gillian Hadfield, Michael C. Horowitz, Charles Isbell, Hiroaki Kitano, Karen Levy, Terah Lyons, Melanie Mitchell, Julie Shah, Steven Sloman, Shannon Vallor, and Toby Walsh, *"Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report*, Stanford University, Stanford, CA, http://ai100.stanford.edu/2021-report

AI and Faith, *The Brain*, https://app.thebrain.com/brains/88eb0773-0065-41e4-9e30-0f9ce83b3340/thoughts/23bdeaf0-d1e5-4832-83e6-4140189e08ca/notes

Anderson, R., (2017, March 17). *After 75 years, Isaac Asimov's Three Laws of Robotics need updating*, The Conversation, https://theconversation.com/after-75-years-isaac-asimovs-three-laws-of-robotics-need-updating-74501

Augustine of Hippo, (426). *City of God*, XV.22

Barton, K., (2015). *Elicitation Techniques: Getting People to Talk About Ideas They Don't Usually Talk About*, Theory & Research in Social Education, 43:2, 179-205, https://www-tandfonline-com.offcampus.lib.washington.edu/doi/full/10.1080/00933104.2015.1034392

Baudelaire, C., (1864). *Le Joueur Généreux (The Generous Gambler)*, short story, Le Figaro

Bergstein, B., (2020, February 19). *What AI Still Can't Do*, MIT Technology Review, https://www.technologyreview.com/2020/02/19/868178/what-ai-still-cant-do/

Bernard, R., (1988). *Research Methods in Cultural Anthropology*, Sage

Billheimer, P.E., (1977). *Don't Waste your Sorrows*, CLC Publications, Pennsylvania

Bloom, B.S., (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*,

New York: David McKay Co Inc.

Bohus, D., et al, (2021, March 29). *Platform for Situated Intelligence*, arXiv:2103.15975 [cs.AI]

Buolamwini, J.; Gebru, T., (2018). *Gender Shades: Intersectional Accuracy Disparities in*

Commercial Gender Classification. Proceedings of the 1st Conference on Fairness,

Accountability and Transparency, Proceedings of Machine Learning Research (PMLR),

81:77-91 Available from https://proceedings.mlr.press/v81/buolamwini18a.html

Castro, D., New, J., (2016). *The Promise of Artificial Intelligence*, Center for Data Innovation,

https://www2.datainnovation.org/2016-promise-of-ai.pdf

Charmaz, K., (2014). *Constructing grounded theory* (2nd ed.) London: Sage.

Chesterton, G.K., (1905, August 16). *Letter to the Editor in response to A Heretic*, The Daily

News, London, Retrieved from https://platitudesundone.blogspot.com/2019/02/what-is-

wrong.html

Crawford, K., (2021). *The Atlas of AI: Power, Politics and the Planetary Costs of Artificial*

Intelligence, Yale University Press

Creswell, J.W., (1994). *Research Design: Qualitative and Quantitative Approaches*, Sage

Publications.

Cognilytica Market Intelligence, (2020, February 6).

https://www.cognilytica.com/document/worldwide-country-ai-strategies-and-

competitiveness-2020/

Cole, M.; John-Steiner, V; Scribner, S.; Souberman, E., editors, (1978). *L.S. Vygotsky Mind in*

Society: Development of Higher Psychological Processes, Harvard University Press

DARPA RATS (Robust Automatic Transcription of Speech),

https://www.darpa.mil/program/robust-automatic-transcription-of-speech

DeepMind, AlphaGo. https://deepmind.com/research/case-studies/alphago-the-story-so-far

Domingos, P., (2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, Basic Books

ESV Bible, (2008). *English Standard Version Study Bible*, Crossway.

Ethics and Religious Liberty Commission (ERLC) of the Southern Baptist Convention, *Artificial Intelligence: An Evangelical Statement of Principles*, (2019, April 11).

https://erlc.com/resource-library/statements/artificial-intelligence-an-evangelical-statement-of-principles/

Etzioni, O., Etzioni, A., (2016, September). *Designing AI Systems that Obey Our Laws and Values*, Vol. 59, Number 9, Communications of the ACM,

https://cacm.acm.org/magazines/2016/9/206255-designing-ai-systems-that-obey-our-laws-and-values/fulltext

Frame, J., (2008). *Doctrine of Christian Life: A Theology of Lordship*, Philipsburg, NJ:P&R, 10

Gebru, T., (2021). *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability and Transparency, ACM Digital Library, https://dl.acm.org/doi/10.1145/3442188.3445922

Gebru, T., et al, (2018, March 23; revised 2021, December 1). *Datasheets for Datasets*, arXiv:1803.09010, [csDB]

Grudem, W., (2018). *Christian Ethics: And Introduction to Biblical Moral Reasoning*, Crossway

Guba, E., (1981). *Criteria for Assessing the Trustworthiness of Naturalistic Inquiries*,

Educational Communication and Technology Journal, 29(2):75-91

Hackett, C., Grim, B., (2012). *The Global Religious Landscape: A Report on the Size and*

Distribution of the World's Major Religious Groups as of 2010, The Pew Forum on Religion

& Public Life, Pew Research Center, https://assets.pewresearch.org/wp-

content/uploads/sites/11/2014/01/global-religion-full.pdf

Hagendorff, T., (2020). *The Ethics of AI Ethics: An Evaluation of Guidelines*, Minds &

Machines **30,** 99–120 (2020). https://doi.org/10.1007/s11023-020-09517-8

Harris, T., (2013). *Original Google Deck on Digital Wellbeing: A Call to Minimize Distraction*

& Respect Users' Attention, https://www.slideshare.net/paulsmarsden/google-deck-on-

digital-wellbeing-a-call-to-minimize-distraction-and-respect-users-attention

IBM, Deep Blue. https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/

IBM, Watson Advertising Accelerator, https://www.ibm.com/downloads/cas/NE1O8V0K

Indurkhya, B., (2018, December 25). *Is morality the last frontier for machines?,* New Ideas in

Psychology, https://doi.org/10.1016/j.newideapsych.2018.12.001

Jiang, L., et al, (2021, October 14). *Delphi: Towards Machine Ethics and Norms*, arXiv >cs>

arXiv:2110.0754v1, [csCL]

Jones, C., (2017). *Why Does God Allow Evil?: Compelling Answers for Life's Toughest*

Questions, Harvest House Publishers

Kincheloe, J.L., McLaren, P.L., (1994). *Rethinking Critical Theory and Qualitative Research*. In

Handbook of Qualitative Research. Norman Denzin and Yvonna S. Lincoln, eds. pp 138-

157, Sage

Kinstler, L., (2021, July 16). *Can Silicon Valley Find God?*, New York Times,

   https://www.nytimes.com/interactive/2021/07/16/opinion/ai-ethics-religion.html

Knight, W., (2017, April 11). *The Dark Secret at the Heart of AI*, MIT Technology Review,

   https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/

Kohlberg, L., (1958). *The Development of Modes of Thinking and Choices in Years 10 to 16. Ph.*

   D. Dissertation, University of Chicago.

Krizhevsky, A, et al, (2012). *ImageNet Classification with Deep Convolutional Neural Networks*,

   NeurIPs Proceedings,

   https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-

   Paper.pdf

Lanier, J., (2010). *You Are Not a Gadget: A Manifesto*, Vintage Books

Lennox, J., (2020). *2084: Artificial Intelligence and the Future of Humanity*, Zondervan

Lewis, C.S., (1944). *The Abolition of Man*, Macmillan Publishing Company, 1996, Touchstone

Lewis, C.S., (1952). *Mere Christianity*, HarperOne, a trademark of Harper Collins Publishers

Lewis, C.S., (1956). *Chronicles of Narnia, The Last Battle*, Harper Collins Children's Books

Lincoln Network, (2018), *The Viewpoint Diversity and Cultural Norms in Silicon Valley Survey*,

   retrieved from https://github.com/lincolnlabs/viewpoint-diversity-reports/blob/main/lincoln-

   viewpoint-diversity-survey-results.pdf

Martius, G., (2017, June 29). *Empowerment as Replacement for the Three Laws of Robotics*,

   Frontiers in Robotics and AI,

   https://www.frontiersin.org/articles/10.3389/frobt.2017.00025/full

Maslow, A., (1943). *A Theory of Human Motivation*, Psychological Review, Volume 50, No. 4

Maslow, A. (1953). *Love in Healthy People*, in Montague, A., (1953). *The Meaning of Love*, in

Sorokin, P., (1954)

McCarthy, J., Minsky, M., Rochester, N., Shannon, C., (1955, August 31). *A Proposal for the*

Dartmouth Summer Research Project on Artificial Intelligence,

http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf

McLuhan, M., (1964, 1994). *Understanding Media: The Extensions of Man*, MIT Press

Merriam, S., (2009). Qualitative Research*: A Guide to Design and Implementation*, Revised and

Expanded from Qualitative Research and Cast Study Applications in Education, John Wiley

& Sons

Merriam, S., (2016). Qualitative Research*: A Guide to Design and Implementation*, 4th Edition,

John Wiley & Sons

Microsoft Research Cambridge, Project Inner Eye, https://www.microsoft.com/en-

us/research/blog/project-innereye-evaluation-shows-how-ai-can-augment-and-accelerate-

clinicians-ability-to-perform-radiotherapy-planning-13-times-faster/

Microsoft Research Montreal Lab, https://www.microsoft.com/en-us/research/lab/microsoft-

research-montreal/

Mitchell, A., Diamond, L., (2018, February 2). *China's Surveillance State Should Scare*

Everyone, The Atlantic, https://www.theatlantic.com/international/archive/2018/02/china-

surveillance/552203/

Musk, E., (2019, October). *Neuralink: An Integrated Brain-Machine Interface Platform with*

Thousands of Channels, Journal of Medical Internet Research, Vol. 21, No. 10, e16194

NCLS (National Conference of State Legislatures), (2022). *Legislation Related to Artificial*

  Intelligence, https://www.ncsl.org/research/telecommunications-and-information-

  technology/2020-legislation-related-to-artificial-intelligence.aspx

NIST (National Institute of Standards and Technology), (ongoing).

  https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt-ongoing

Niu, F., et al, (2022, April). *Differentially Private Estimation of Heterogeneous Causal Effects*,

  First Conference on Causal Learning and Reasoning, Proceedings of Machine Learning

  Research vol 140:1–17, 2022

OpenAI, GPT-3. https://openai.com/blog/gpt-3-apps/

O'Shea, K., Nash, R., (2015, Dec. 2). *An Introduction to Convolutional Neural Networks*,

  arXiv:1511.08458 [cs.NE]

Oxford English Dictionary, (March 2022), *OED Online*, Oxford University Press,

  https://www.oed.com/

Patton, M.Q., (2002). *Qualitative research and evaluation methods*, 3rd edition, Sage

Pavlov, I. P. (1927). *Conditioned reflexes: an investigation of the physiological activity of the*

  cerebral cortex. Oxford University Press

Pew Research Center/Elon University Imagining the Internet Center, (2021). *Survey XII: What is*

  the Future of Ethical AI Design?, Retrieved from

  https://www.elon.edu/u/imagining/surveys/xii-2021/ethical-ai-design-2030/

Piaget, J., (1932). *The moral judgment of the child*. London: Kegan Paul, Trench, Trubner & Co.

Postman, N., (1992). *Technopoly: The Surrender of Culture to Technology*, Vintage Books

Rainie, L., et al, (2021, June 16). *Experts Doubt Ethical AI Design Will Be Broadly Adopted as the Norm Withing the Next Decade*, AI Research Topics, Pew Research Center

RenAIssance Foundation, (2020, February 28). *Rome Call for AI Ethics*, https://www.romecall.org/the-call/

Richards, L., (2015). *Handling qualitative data* (3rd ed.), London: Sage.

Sadowski, J., (2018, Aug. 6), *Potemkin AI*, Real Life, https://reallifemag.com/potemkin-ai/

Smith, B., Brown, C.A., (2019). *Tools and Weapons: The Promise and Peril of the Digital Age*, Penguin Press

Sorokin, P., (1954). *The Ways and Power of Love: Types, Factors, and Techniques of Moral Transformation*, Templeton Foundation Press

Stanford AI Index, (2021). Daniel Zhang, Saurabh Mishra, Erik Brynjolfsson, John Etchemendy, Deep Ganguli, Barbara Grosz, Terah Lyons, James Manyika, Juan Carlos Niebles, Michael Sellitto, Yoav Shoham, Jack Clark, and Raymond Perrault, *The AI Index 2021 Annual Report*, AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report_Master.pdf

Stanford Research Institute, SIRI, https://www.sri.com/hoi/siri/

Taylor, A., (2018, Aug. 1). *The Automation Charade*, Logic, Issue 5 – Failure, https://logicmag.io/failure/the-automation-charade/

Turing, A., (1950, October 1). *Computing machinery and intelligence*, Mind, Volume LIX, Issue 236, pp. 433-460, https://academic.oup.com/mind/article/LIX/236/433/986238

Varkey, B., (2020, June 4). *Principles of Clinical Ethics and their Application to Practice*, Medical Principles and Practice, 2021; 30:17-28, Karger

Vallor, S., (2018). *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*, Oxford University Press

van Seijen, H., et al, (2017, June 13; revised 2017, November 28). *Hybrid Reward Architecture for Reinforcement Learning*, arXiv:1706.04208 [cs.LG]

Whittlestone, J., et al, (2019, January 27-28). *The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions*, Leverhulme Centre for the Future of Intelligence, University of Cambridge, AIES '19, Honolulu, HI, USA, https://dl.acm.org/doi/pdf/10.1145/3306618.3314289

Wiggins, K., (2020, September 9). *NIST benchmarks show facial recognition technology still struggles to identify Black faces,* https://venturebeat.com/2020/09/09/nist-benchmarks-show-facial-recognition-technology-still-struggles-to-identify-black-faces/

Wolcott, H., (1995). *The Art of Fieldwork*, AltaMira Press

Wolcott, H., (2008). *Ethnography: A Way of Seeing*, 2nd Edition, AltaMira Press

Wolcott, H., (2009). *Writing Up Qualitative Research*, 3rd Edition, Sage Publications

Zuboff, S., (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, Hachette

Appendix A - Institutions

Institutions and Organizations for AI Ethics

- Partnership on AI
- AI Now
- Data & Society
- Berkman Klein Center for Internet and Society
- Center for the Governance of AI
- Center for Humane Technology
- Stanford Human-Centered AI Institute
- Ethics of AI Lab
- MIT Ethics and Governance of AI
- Alan Turing Institute
- Center on Privacy and Technology
- Microsoft Research Responsible AI,
- ACM FAccT
- The Ethics and Religious Liberty Commission of the Southern Baptist Convention
- The Roman Catholic Church
- U.S. Office of Science and Technology Policy

Appendix B – Guidelines

<u>Sample Guidelines:</u>

- [U.S. Department of Defense](#)
- [AI Now](#)
- [Microsoft Research](#)
- [U.S. Government Accountability Office](#)
- [Harvard Data Science Review](#)
- [Canadian-ASEAN Business Council Montreal Declaration](#)
- [U.S. Executive Order](#)
- [European Union](#)
- [Organisation for Economic Co-operation and Development (OECD)](#)

Appendix C – Protocol

In the quest for robust and beneficial AI, we must concurrently seek to provide a framework of robust and beneficial AI ethics. The goal of this study was not to interrogate the benefits and harms of AI in general, but rather to interrogate how Christian teaching, texts, and traditions can contribute to, or inform, the current secular conversation on AI ethics. The questions were to be viewed loosely through the lens of several key tensions drawn from current thinking and literature in the field. These tensions included:

- **Robustness v. privacy** (quality and efficiency v. autonomy of individuals)
- **Accuracy v. bias/fairness** (algorithmic precision v. equal treatment of individuals)
- **Personalization v. community** (personalized services v. citizenship and unity)
- **Convenience v. freedom** (life of leisure v. self-actualization and free will)
- **Leisure v. dignity** (ease of life v. value and dignity of work)
- **Augmentation v. replacement** (human enhancement v. human obsolescence)
- **Autonomy v. control** (free will and the "Frankenstein" effect v. totalitarianism)

As noted, the original strong focus on ethical tensions did not promote fruitful discussion, so while the issues surrounding these tensions were still a part of the discussion, they were not the main part of the discussion. Rather, the interviews – with variation, depending on the informant's knowledge and expertise – focused on the following questions:

General Research Questions:
1. How does worldview affect how AI researchers and developers conceive of and make artificial intelligence?
2. Does a Christian worldview have anything unique to contribute to the discussion around ethical issues in AI?
3. How might AI ethics be more robust and more beneficial if we brought Christian teachings, texts, and traditions explicitly into the conversation?

Follow-up Questions:

- How would you compare the prevailing secular view on how to **ensure ethical behavior** with the Christian view? How might AI ethics be more robust if we brought the beliefs, texts, teachings, and traditions of the Christian faith explicitly into the conversation?

- We have a plethora of ethical principles and laws attempting to ensure robust and benevolent AI. Where and how might the idea of **virtue** fit into the conversation?

- How does the idea of **transcending our human limits** via AI compare with the biblical lesson on transcending our human limits? Is there a balance between enough and too much technology?

- Proponents of AI promote it as an aid to **human flourishing** primarily in material terms. How does the world's view of human flourishing compare with a biblical view found in Jesus' Sermon on the Mount and other places in scripture? What does that say about how we should conceive of, make, and use AI?

- Data provenance is the process of identifying the origin and path of the data in order to ensure its quality. How might the idea of **provenance** be important as it pertains to the quality of ethics as well? Does it matter if anyone knows where ethics come from? How?

- Recent attempts to make **moral machines** – and earlier work on prediction engines – seem very much like modern day versions of ancient oracles. What does the Bible say about how humans should make wise decisions and what dangers (if any) do we face in abdicating responsibility our decisions – or outsourcing wisdom – to machines, both technically and spiritually?

- What is the **telos of AI** in terms of where the science is heading and what is the **telos of humanity** according to Christian texts? How do they compare? Is there anything Christian teaching can tell us here? How would this impact the conversation or direction of so-called Ethical AI?

- AI ethics is often cast in the role of hall monitor, preoccupied with ferreting out bad behavior and stopping it – rather than being conceived of as a way to advance a bold, brave, spiritual vision of humanity and our creations. How might we retain appropriate levels of critique and caution while presenting a more inspirational and aspirational view of ethics according to Christian teaching?

Appendix D – AI and Faith

AI and Faith is a cross-spectrum consortium of religious faith communities and academic institutions. Its mission is to bring the fundamental values of the world's major religions into the emerging debate on the ethical development of artificial intelligence and related technologies.

Faith communities, faith-related institutions, and universities in technology-centric regions like Seattle are home to professionals, scientists, theologians, ethicists, teachers of religion, and other thought leaders who are stakeholders in the evolution of AI. AI and Faith exists as a channel for religious perspectives to help shape the development of AI in ways that are deeply ethical and life-affirming.

In addition to promoting discussion and bringing communities together, AI and Faith conducts original research to better understand the relationship between sacred texts and millennia of commentary on ethics in everyday life, government, and business, seeking to align faith-based wisdom with the current secular dialog on the ethical development of AI. The organization is open to a diverse range of members, research fellows, and partners, including faith traditions beyond the Abrahamic religions, as well as ethicists, scholars, and institutions that may hold a secular perspective but value the inclusion of religious perspectives in the ethical discussion surrounding developments in AI.

For more information, click here.