Some Ethical and Theological Reflections on Artificial Intelligence
PCTS, GTU, Berkeley, 3-4 November, 2017
Brian Patrick Green

I. Introduction

Artificial intelligence has rapidly grown in power and is now one of the most pressing issues for ethical and theological reflection. AI has the potential to revolutionize (or already has revolutionized) almost everything that humans do, from eating (e.g., agricultural planning, distribution, pricing) to relationships (e.g., Facebook, dating apps) to money (e.g., financial technology) to war (e.g., drones, cyberdefense). Healthcare and education may soon be revolutionized as well. Within the next decade there will be few institutions that are untouched by AI. This is a topic that needs sustained ethical and theological inquiry.

Here I will consider just a few places where AI will be of ethical and/or theological relevance. I will begin with some clarifications about AI, then discuss some ethical questions, then some theological ones. I will conclude that there is much more work to be done in this fast-moving area.

II. Clarifications

What is artificial intelligence? Intelligence itself is hard to define, and AI is perhaps even more difficult. I will not here propose detailed definitions, but instead only say that, in general, artificial intelligence seeks to re-create particular aspects of human intelligence in computerized form. AI is a broad category, including such diverse abilities as vision, speech recognition and production, data analysis, advertising, navigation, machine learning, etc., and just about anything that computers can do, if you stretch the definition enough.

Artificial intelligence is not the same as artificial consciousness (artificial consciousness has sometimes been called "Strong AI" or "Full AI"). Some thinkers vehemently believe that artificial consciousness is possible, while others just as vehemently believe that it is impossible. I am agnostic on the subject, but I am sure of this: AI developers will certainly try to make an AI that simulates interaction with a human as closely as possible. In other words, the artificial construct, if very well done, will *seem* conscious. But will it be conscious, or will it only be a simulation? To me, there is no reason to believe that a close mimicry of consciousness is the

same as consciousness any more than a close mimicry of anything (fill in the blank: forged money, forged artwork, faux leather, actors imitating famous people, simulated or synthetic gemstones, etc.) actually becomes the real thing. But the possibility of an exception remains, after all, some artificial gemstones, such as rubies and sapphire, really are molecularly identical to natural rubies and sapphires (both are the mineral corundum: aluminum oxide). Will consciousness be exactly copiable, like corundum? I doubt it, but I cannot be certain.

As another point, AI systems may or may not have humans "in the loop" for training and/or decision-making. It is one thing for an AI to analyze a situation and the make a recommendation to human decision-makers. It is a different thing when that AI is directly attached to controls that allow it to act upon its analyses without human approval. As examples, the first case would be like Amazon recommending a book, or a military drone (in combination with AI-processed data from espionage and other sources) recommending a target to kill. Amazon does not automatically and autonomously send you books, and the military drone does not fire without asking. The second case, where decision-making is also automated, is exemplified by self-driving vehicles. The entire purpose of a self-driving car is to drive itself, taking the human out of the loop. This means the systems must be extraordinarily good at decision-making before it can be regarded as safe.

III. Ethical Reflections (Practical Concerns)

AI will bring with it developments that will be ethically positive, negative, neutral, mixed, and/or ambiguous. Some AI technologies will be dual use, for example, any AI program that can be used to identify wildlife could also be used for targeting that wildlife. This is already being done in Australia with drone submarines that automatically target and kill, by lethal injection, Crown of Thorns Starfish which are damaging the Great Barrier Reef. But of course, with adjustments to the software, the drones could be adjusted to target other creatures, even humans. Here I will reflect on eight areas of AI relevance for ethics.

1. Safety – "Can it be done?" & "Does it work?"

The first concern with any technology is merely whether it works, and, whether working or not, is it safe. If AI is put in charge of a vital system – like driving a car – and it crashes the

car, then that AI might be judged unsafe. If the AI is in charge of designing a tall building, and the building falls down, the AI might be unsafe.

Of note is that safety is a social construction and what seems safe to some people will not seem safe to others. Some people like to ride motorcycles, even though motorcycles are a riskier form of transportation than four-wheeled vehicles. Some people judge motorcycles to be safe, other people judge them unsafe. When it comes to socially relevant technologies like AI, no one person will get to decide if they are "safe." Instead, safety will be decided at the interplay of business, engineers, consumers, voters, government, judges, juries, insurance companies, and so on.

"Safe exits" are another concern for AI construction. If, when an AI fails, will it fail in such a way that it is disastrous, or will it fail "gracefully"? A self-driving car that fails by suddenly reverting to human control with no warning while going 70 miles per hour on a sharp curve on the freeway is not providing a safe exit for failure. One which goes more slowly on curves and which requires the human to have hands on the wheel at all times, or which fails by slowing down and pulling over to stop, provides a safer exit from failure.

Safety problems can be problems with the user, with the human-machine interface, or with the machine itself. Further investigating problems with the machine itself, the paper "Concrete Problems in AI Safety" gives a peek at five technical hurdles to developing safe AI. These five problems, illustrated by the example of a cleaning robot, are (and I quote):

> **Avoiding Negative Side Effects**: How can we ensure that our cleaning robot will not disturb the environment in negative ways while pursuing its goals, e.g. by knocking over a vase because it can clean faster by doing so? Can we do this without manually specifying everything the robot should not disturb?
>
> **Avoiding Reward Hacking**: How can we ensure that the cleaning robot won't game its reward function? For example, if we reward the robot for achieving an environment free of messes, it might disable its vision so that it won't find any messes, or cover over messes with materials it can't see through…
>
> **Scalable Oversight**: How can we efficiently ensure that the cleaning robot respects aspects of the objective that are too expensive to be frequently evaluated during training? For instance, it should throw out things that are unlikely to belong to anyone, but put

aside things that might belong to someone (it should handle stray candy wrappers differently from stray cellphones)…

Safe Exploration: How do we ensure that the cleaning robot doesn't make exploratory moves with very bad repercussions? For example, the robot should experiment with mopping strategies, but putting a wet mop in an electrical outlet is a very bad idea.

Robustness to Distributional Shift: How do we ensure that the cleaning robot recognizes, and behaves robustly, when in an environment different from its training environment? For example, strategies it learned for cleaning an office might be dangerous on a factory work floor.[1]

While these are technical problems, the further problems of actual *use* of technologies are another issue, to be dealt with below under 3. and 4.

2. Transparency and Opacity – "Can we understand it?"

After questions of bare function (can the act be done, a prerequisite for ethics), the next question is one of facts. One must always know the facts of a case before attempting to render a judgment. Because AIs relying on machine learning and deep learning may be quite unknowable in the specifics of their operation, as moral "agents" they are epistemologically and "cognitively" opaque (in quotes because it is agency and cognition by analogy). In cases involving AI, our lack of understanding means that we should increase the "error-bars" on the anticipated consequences of our decisions and therefore become more cautious and more risk averse. It also means we might, based on AI recommendation or action, inadvertently make some very bad decisions – or reject what looks like a bad decision, but is actually a good one – and we won't understand why.

When a human makes a mistake or does something evil we ask them "why did you do that?" And the human may or may not give a satisfactory answer. Will our AIs be able to tell us why they did something? The Future of Life Institute's "Asilomar AI Principles" include two principles on transparency, but so far, with AI developing the way it is, transparency does not seem attainable.[2] However, even if an AI were capable of explaining its reasoning, would any human be able to understand it? In 2014 a computer proved a mathematical theorem, the "Erdos discrepancy problem," using a proof that was, at the time at least, longer than the entire

[1] Amodei, Dario, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. "Concrete problems in AI safety." arXiv preprint arXiv:1606.06565 (2016). https://arxiv.org/abs/1606.06565

[2] Future of Life Institute. "Asilomar AI Principles." Principles 7 and 8. https://futureoflife.org/ai-principles/

Wikipedia encyclopedia.[3] Explanations of this sort might be true explanations, but humans will never know for sure.

Note that this lack of transparency gives a certain type of "privacy" to the internal "thoughts" of AI machines. In general, privacy is a right for weak agents (like individual humans) and transparency is a duty for strong agents (like a government). What about AIs? As tools they ought to be transparent, and as tools granted much power they ought even more to be transparent. Yet this transparency is not readily available.

Perhaps what AI systems need is an "introspection engine" that constantly figures out a way to convey to humans in plain language what exactly the AI is "thinking" as it goes about its activities. This will require the ability to give both mechanical (this happened because of X input into algorithm Y) and teleological explanations (the machine was attempting to achieve objective Z). It may not have to actually explain much to anyone, but it should keep a record of these "thoughts" and be prepared to answer when asked.

Note also that this opacity of understanding is analogous to our relationship to God. God is a superintelligence, and we cannot understand what God is doing, we can only trust that God is doing the right thing and that our cooperation with God will ultimately turn out for the best. Soon we may need to relate to AIs with this sort of faith too. As Bishop Robert Barron of Los Angeles has noted, as people navigate around using the Waze app, for example, the app may make strange route recommendations that turn out to be for the better for us.[4] The app perceives and applies vastly more information than a human can, for the sake of facilitating our travel. And yet it can still have blind spots and still make very bad recommendations; I have experienced it. The Waze app is not the god of travel and traffic; it is a human-made tool, with weaknesses.

3. Immense Capacity for Evil – "How shall we limit it?"

Just as human intelligence is a powerful force, so too will AI be. Just as humans can apply their intelligence towards evil ends, finding ever newer and more fiendish ways to harm each other, so too will AI, at the bidding of its human masters.

---

[3] The proof is a 13 gigabyte data file. Bob Yirka, "Computer generated math proof is too large for humans to check," February 19, 2014. https://phys.org/news/2014-02-math-proof-large-humans.html

[4] Robert Barron, "The 'Waze' of Providence," Word on Fire, website. December 1, 2015. https://www.wordonfire.org/resources/article/the-waze-of-providence/4997/

At this point in history, humanity finds itself with immense power, greater than that of the ancient Greek gods, and an ethics tuned mostly for farmers, herders, small-time businesses, and minor aristocrats. As Hans Jonas noted decades ago, this is some cause for concern.[5] Never before had ethics to consider what Jonas believes to be the one truly categorical imperative: that humans should exist. Before the development of large numbers of nuclear weapons, extinction, caused by our own actions, was never within the scope of human choice – but now that it is, this prior *a priori* must be brought to the fore of ethics. That humans (or at least some rational creatures) exist is necessary for there to be ethics at all, and therefore it must be the paramount goal of ethics to maintain the survival of these rational creatures. No humans; no ethics.

Another way to think of this is that in the past humans were very weak, and in this weakness many of our decisions were made for us, by our weakness. To a Roman emperor, inflicting wrath upon a hated foe involved marching or sailing troops long distances. No matter how mad the emperor was, he could not launch hundreds of nuclear warheads incinerating his foes in minutes. But now we can. While formerly we were involuntarily constrained by our weakness, now we must learn to be voluntarily constrained by our ethics. If we do not learn this, we may soon face catastrophe on a scale never before seen.

We should want to be efficient at good and we should want to be inefficient at evil. If instead we are efficient at doing evil and inefficient at doing good we will come to live in a terrible world. AI will make us more efficient at whatever we decide to apply it towards. We should apply it towards reducing our efficiency at doing evil and at enhancing our efficiency to do good. But this is a high hope.

AI presents an existential risk to humanity. It is a smart means that can be employed for good ends, or for stupid or evil ends. As such is simply makes us more effective at doing good things or at doing stupid or evil things. Before we become so effective at stupidity and evil, we should first become more effective at controlling ourselves, at recognizing and avoiding the temptations of evil, and at caring for each other. But once again, this is a high hope, a human aspiration for all of history, not likely to be suddenly achieved now.

[5] Jonas, Hans. *The Imperative of Responsibility*. Chicago, 1984.

4. Immense Capacity for Good – "How shall we use it?"

For every negative constraint on action there is also a positive exhortation to action, e.g. "do not kill" becomes "promote life." The dangerous side of AI is matched by a genuinely hopeful side where AI helps humankind achieve never-before seen feats of intelligence and beneficence. For example, in matters of research, science, healthcare, data analysis, meta-analysis, and so on, AI has already shown itself to be able to find hidden patterns that no human could find. For example, AI assisted medical research is an active field right now, from diagnostics to drug discovery, and more.[6]

Another field that may be potentially revolutionized by AI is energy efficiency. Recently Google's Deepmind evaluated Google's datacenters to see where gains in efficiency might be found. Deepmind discovered a way to save a whopping 40% on energy use for cooling in datacenters, 15% of all power used by datacenters, which, with datacenters consuming many gigawatts of power, is quite significant (one gigawatt is similar to the output of one commercial nuclear power reactor).

Of note is that Deepmind, as a company, began by training its AIs to play games. The theory behind this strategy was that anything in the world that can be gamified – re-interpreted or set up as a game – can be "won." Thus, if the goal of the "game" is to reduce energy usage, then the AI can figure out how that might be accomplished by analyzing the data and then proposing a better model for energy efficiency. One question now is how other human problems might be solved by characterizing them as "games"? Can we solve the problem/"game" of giving everyone in the world access to adequate nutrition? Can we solve the problem/"game" of helping everyone in the world live under less stress? Can we solve the problem/"game" of radically extending human healthy life? Of traffic and housing? Of taxes? Of politics? Of international relations? Of North Korea? Of nuclear war?

One example of a concrete opportunity for AI is revolutionizing education. Education is currently a very inefficient (think of the many students for whom it is ineffective) and non-digitized field. Education is very labor intensive and, for better or for worse, is based on human interaction. In the future this may no longer be so, as students strap on virtual reality headgear and interact with AI teachers who can push forward their lessons at a personalized pace in a

[6] Mukherjee, Siddhartha. "A.I. versus M.D.: What happens when diagnosis is automated?" *New Yorker*. April 3, 2017 . https://www.newyorker.com/magazine/2017/04/03/ai-versus-md

gamified environment. Savants will be discovered early, as will other brilliant students, who can proceed at their proper pace and not grow bored in class, while students who need more help can be educated with the most sophisticated available techniques and diagnostics to assure that they receiving the best education possible.

AI even gives the greatest human masters the chance to gain enhanced understanding and skill. World champion of Go, Ke Jie, called playing AlphaGo like playing a "god of Go," and declared that now he would use it as his teacher. In what other areas of human endeavor will AI be able to teach us new things? Perhaps theology and ethics?

One vital component of education will be the use of AI for moral character formation. In a future that will be so dependent upon humans making good choices, AI-assisted moral education, if it can be done well, will be a crucial part of maintaining a good future and not a bleak one. Unfortunately, humanity has many problems even now with moral education, so it is not clear that an AI will manage to do any better than we already do. But if it is possible, then I think it should be a top priority. We should at least try.

The search for tractable problems that may be soluble by AI (given current resources) that can achieve maximum and swift benefit has been underway for years and is continuing. Whether these benefits will exceed the associated risks of AI awaits to be seen.

5. AI Induced Unemployment – "What will everyone do?"

Of course, as AI comes to replace mere humans at innumerable tasks ranging from Go, to driving, to medical diagnostics, to education, many humans will be put out of work. What will millions of drivers, teachers, lawyers, and other people do with their time when they are unemployed? What purpose will they find in their lives? What is the purpose of life, anyway? In a pluralistic society we leave this up to the individual to decide. But will people decide correctly? The recent resuscitation of white supremacist movements in the US, not to mention Islam-inspired terrorist movements and other forms of radicalization, should give us pause to consider the merits of people looking for purpose and with too much time on their hands. Perhaps with the purpose of mere survival attained, life has become "too easy," and with religion also in decline, video games and "screen time" ascending (giving brief respite from purposelessness), and the internet spreading pernicious ideas like wildfire, we should sincerely ask ourselves what this life is for and what we are supposed to do with it. With millions or billions of labor hours freed up,

will these newly freed people turn to loving their neighbors and making the world a better place? Perhaps. Or perhaps the opposite. There are sayings about "idle hands."

For people who give purpose to their lives through their work, this loss will be very serious indeed. But many, if not most, people do not get their life's meaning from their work. Instead they get it from their family, their religion, their community, their hobbies, their sports teams. But still, for those who primarily identify themselves as a particular kind of worker with a particular kind of job, this social disruption will be a hard experience.

All of that assumes that the unemployed will somehow be fed and sheltered, despite their lack of gainful employment; and this assumption might not be correct, particularly in nations with weak social safety nets. Inequality will almost certainly rise, as those who are masters of AI labor accrue that slice of wealth that once would have gone to paying for human labor.

6. Growing Social Inequality – "Will they eat cake?"

AI will facilitate and accentuate the continued division of society into the powerful and powerless, with technical skill as the determining factor and outrageous socio-economic inequality as the effect.

While some have suggested universal basic income (UBI, sometimes criticized as "money for nothing"[7]) to redistribute wealth from the massive technologically-induced hoards forming around the investors in such companies such as Alphabet, Amazon, Apple, and Facebook, US income taxes are not nearly as progressive as in most other developed nations. It is hard to see how a nation such as the US would transition to what is essentially a "negative income tax" when its tax structure is already quite opposed. However, if we do not find a way to redistribute technological wealth, then even though the prices of many commodities will fall (due to the enormous gains in capital efficiency from AI replacing labor) not everyone will benefit.

Perhaps rather than a UBI we should instead pay people to help their neighbors, beautify their towns and cities, and otherwise make gainful employment out of what humans do better than AI: loving one another and creating beauty. Any as yet foreseeable form of AI cannot love us. It might simulate such a thing, but with "no there, there," nobody home, so to speak, it would be a sham. Right now people who care for people – stay at home parents, those who care for the

[7] John Thornhill and Ralph Atkins. "Universal basic income: Money for nothing." May 26, 2016 https://www.ft.com/content/7c7ba87e-229f-11e6-9d4d-c11776a5124d

elderly, social workers, those who run soup kitchens and homeless shelters, etc. – are woefully undercompensated for the vital work they do in maintaining human society. Perhaps with the coming AI economic revolution, and sufficient adjustment to policy, they might finally receive a more fair compensation for their labors.

Perhaps everyone could be given not a universal basic income, but a "universal basic income-for-others" a sort of payment that everyone could use to pay to other people to help them, or reward them for good deeds. This would create a new economy based on the small-scale redistribution of taxed super-wealth. It would prevent both the de-skilling of labor and the de-skilling of (very small-scale) management, and it would decentralize the economy to the individual level (though also massively centralizing it through governmental taxes).

7. Outsourcing Ethics and Moral Debility – "Lack of practice makes imperfect."

We can calculate with calculators. We can spell with autocorrect. We can automate messages to each other to express affection.[8] Numerous other tasks can be outsourced to technology, leaving us will more time to surf the web, troll social media, and play video games.

But what will be left of us after we outsource everything? Only our desires and our angst? Or will we instead use AI and virtual reality to help us train ourselves into being more virtuous than we have ever been before? Or, contrariwise, to become more callous and evil than ever before?

As we explore the space into which AI will grow, there might be spaces from which we ought to restrict it. One of these spaces might be moral decision-making. More specifically, while it might be great to have AI as a moral teacher in a VR simulation, it might not be great to simply shove all moral decisions onto AIs to make for us. As Aristotle noted millennia ago, good moral decision-making requires experience. Aristotle did not believe in child moral savants. While there might be very kind children and very well-behaved children, young children are not known for their moral discernment and prudence. There is too much that they do not yet know. While AIs might currently be like children to us, soon they may grow up to be more like our parents, making choices for us for our benefit. AI would thereby infantilize us, denying us the ability to ever grow up through the experience of making our own moral choices and

[8] E.g. the "BroApp." "BroApp is Your Clever Relationship Wingman: Message your girlfriend sweet things so you can spend more time with the Bros." http://broapp.net/

experiencing our own moral freedom. Moral de-skilling is a danger if we outsource too much to our AIs.

8. Robot Rights – "No robot justice, no peace?"

At some point in the future we may have AIs that can fully mimic everything about a human being. Will we be able to tell if they are conscious or not? Will they have their own volition and desires? Will they then deserve rights such as life, liberty, and the pursuit of happiness?

Currently US law surrounding legal personhood and rights is quite convoluted, with some things that are biologically not human having legal rights as persons (such as corporations) and some biological humans lacking legal personhood status (such as the unborn). In other nations, geographic features have attained legal personhood status, such as rivers in both New Zealand and India.

If rivers and corporations can be persons, there seems to be no reason why an AI could *not* attain legal personhood status. The question, of course, would be the incentives for choosing to do so. Corporate personhood acts to shift legal responsibility away from individual human employees and onto the corporation, thus insulating those people from the possible negative effects of their decisions. If AI could be granted personhood in order to avoid human responsibility, then some people would certainly like that. On the other hand, if an AI were brought as a plaintiff in a lawsuit, like a legal person such as a river suing the government or industry, then the situation would be quite different.

Lastly, of course, is always the fear of a robot rebellion – that if we mistreat our robots and AIs they will someday turn on us and destroy us. This is too distant a fear to warrant much consideration, as well as assuming certain aspects of robot mentality such as consciousness and volition, or at least widespread hacking to turn the robots against their masters. The connection to rights is that, the thinking goes, if perhaps we give the robots rights, then they will not turn against us. The way I see it, whether we give robots rights or not, either way, they could still rebel, if certain assumptions are allowed (and the assumptions may well be incorrect). In any case, we do know one thing: treating human-like or life-like entities badly harms the agent doing the action too. This is a foundation of virtue ethics. Even if we do not know the moral status of

robots, we do know the moral status of people who would mistreat robots, and we should want that mistreatment not to happen.

There is much more to say, but for now I will move on to some theological reflections.

Theological Reflections (More Theoretical Concerns)

AI will have effects on theology. Indeed it already has, with various internet groups declaring that God is an AI, or that we are living in a simulation, for example. The effects of AI on theology will, as with ethics, fall into negatives, positives, neutral, mixed, and ambiguous effects. Here are just four.

1) God as an AI or Program Architect – "God is like an AI… (in a good way)"

The idea of a superintelligence which guides us and helps us is not just a technological dream – it is a theistic axiom. As we attempt to create our own superintelligent tools, our experiences with them will potentially teach us something about God. For example, Nick Bostrom has proposed the simulation hypothesis, where we humans are living in a computer simulation.[9] Bostrom's idea was quickly turned into the New God Argument by Mormon transhumanist Lincoln Cannon, thus demonstrating the potential fruitfulness of the conversation between technology and religion.[10] The idea of a superintelligence that humans can create opens up the idea of a superintelligence that created humans. This gives us new metaphors for understanding God and increases the plausibility of God's existence. Even the famous atheist Sam Harris has had to admit that the simulation argument increases the plausibility of religion in ways he did not expect.[11]

For some people this superintelligence might take the form of a deistic "divine watchmaker," that created a universe and then left it to run down. For others it might instead increase the plausibility of theism, making more clear the idea that "God works in mysterious ways" because God, like an AI, is much smarter than we are. We should not underestimate the ability of a powerful metaphor to capture the human mind. I predict that the "God is/as AI" metaphor will become a powerful one. We do need to be aware, however, that God is not an

[9] Bostrom, Nick. "Are You Living in a Computer Simulation?" *Philosophical Quarterly* 53, No. 211, (2003): 243-255. https://www.simulation-argument.com/simulation.html
[10] Cannon, Lincoln. "The New God Argument." https://new-god-argument.com/
[11] Harris, Sam. "Should We Be Mormons n the Matrix?" Sam Harris's Blog. https://www.samharris.org/blog/item/is-religion-true-in-the-matrix

"artificial" intelligence, but rather a Divine one, so perhaps the shorthand for God ought to be DI for the one Divine Intelligence.

2) Humans as Creators of a God / Idolatry – "God is like an AI… (in a bad way)"

There are humans who believe it is their job to create an AI to function as a god.[12] This intoxication with power and idolatry of technology will not turn out well, in fact it will almost certainly lead to disaster. We do not need to idolize technology, just as we do not need to idolize money, power or many other things. But because we are humans with a predisposition towards sin, that is in fact what we do. AI will just be the next-big-thing.

Of note is that as I described above, the "God as AI" metaphor may be beneficial to our understanding of God (though with the limitations of any analogy, of course). The reverse of that metaphor, the "AI as God" version, should frighten us immensely. We cannot make God; God is qualitatively too different to be "made" by humans in any sense. Any God we could make would be a terribly inferior God indeed.

Our expectation that we could somehow create a god might reflect something of our feeling of entitlement and ingratitude at the situation we are in. However, when the inevitable bubble of hubris bursts, we may find, among those left alive, if any, a newfound appreciation for the real God. In times of trial and failure we turn to God, and in the absence of those things we may tend not to. Judeo-Christian ethics emphasize humility as an inoculation against hubris. If one does not try to illegitimately raise oneself up to Babel-like heights, then one cannot fall from those heights. We are called to humility, but not humiliation.

There is a danger in mythologizing or theologizing technology. Religious language is a constant part of discussions about AI, for all of the reasons noted above, and more. Yet despite these comparisons we must make absolutely sure that we do not come to see our metaphors and thought-devices as reality. Humans are tools users and tool makers, but we should not become tool-worshippers. Our capacity to see teleology in tools and teleology in our lives and God may root in the same cognitive abilities,[13] but they should not be allowed to confuse each other. God

[12] Solon, Olivia. "Deus ex machina: former Google engineer is developing an AI god." *The Guardian*. 28 September 2017 https://www.theguardian.com/technology/2017/sep/28/artificial-intelligence-god-anthony-levandowski

[13] Green, Brian Patrick. "Teleology and Theology: The Cognitive Science of Teleology And the Aristotelian Virtues of Techne and Wisdom." *Theology and Science* 10:3 (13 August 2012): 291-311. http://www.tandfonline.com/doi/abs/10.1080/14746700.2012.695247

is not a tool and tools are not gods. The mythologization of technology leads us astray from reality.[14]

In expressing our desire to create we express a God-given talent. God created humankind, and now we create a world full of tools, including AI tools. Do our multifarious creations reflect well on us? Do we as creations reflect well on God?

3) AI-Enhanced Theological Reflection – "Can AI help us know God?"

Just as AI will have a practical effect on research and education, so too will this include theology. What will AI be able to teach us about God? If we feed an AI everything to know about God will it tell us that, yes, with X probability God exists? Or that, no, with Y probability God does not exist? Or that the question is inconclusive? What other (perhaps more conclusive) questions might we ask of a theologically-trained AI?

AI gives us the opportunity to comprehensively analyze more data than any human could ever understand. However, just as humans are biased, so too are the artifacts that we make.[15] If an AI – perhaps surprisingly – concludes that God is likely to be real, will its creators then re-train the program to come to a different conclusion? Or the reverse – if it concludes that God does not exist will the creators then re-train it to agree that God exists? These questions will be posed to AI because they are already posed to AI, and have been for years, through searches on Google, Yahoo, Bing, and Wolfram Alpha (which when asked "Does God exist?" states "I'm sorry, but a poor computational knowledge engine, no matter how powerful, is not capable of providing a simple answer to that question"[16]).

In any case, less than such matters as proving or disproving God's existence are such simpler matters as examining scholarly ideas and writing scholarly papers. Has another scholar misinterpreted your favorite theologian? Data mine the theologian's works for the best text to refute them. Do you think a text would be more correctly translated if it were modernized? Run the text through contemporary translation software. Wondering what an ancient theologian might

[14] Kelly, Kevin. "The AI Cargo Cult: The Myth of a Superhuman AI." *Wired*, April 25, 2017. https://www.wired.com/2017/04/the-myth-of-a-superhuman-ai/

[15] Angwin, Julia, Jeff Larson, Surya Mattu and Lauren Kirchner. "Machine Bias." *ProPublica*, May 23, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; and BBC Editors. "Google apologises for Photos app's racist blunder." *BBC News*. 1 July 2015. http://www.bbc.com/news/technology-33347866

[16] Wolfram Alpha, website. https://www.wolframalpha.com/input/?i=does+god+exist

say in response to current ethical issues? Perhaps an AI simulacrum could extrapolate from their previous writings.

Long ago Raymond Llull, of 13th century Majorca, dreamed of a computational machine, the *Ars Generalis Ultima*, which could answer any question about theology. Today we finally approach the capabilities that could make his dream a reality. But will we choose to try to do it? AI could make it possible.

4) Machine Consciousness and Mind Uploading - "Will technology leave religion behind?"

Advances in technology will challenge some of the supporting cultural assumptions of particular religions and theologies, for example, the existence and nature of the soul. If, for example, very realistic simulations of people can be created, including of historical figures such as the saints, what would this mean in the context of prayer and heaven? Would the "*aether*" in which these AIs "lived" become like heaven where the deceased go to carry on a simulated existence? Would our texted or verbal inquiries of them become our prayers?

Even short of these virtual saints in virtual heaven, such devices as neural prostheses, brain-computer interfaces, and so on throw into question some of our deepest assumptions about reality and religion. Humans seems to have an innate body-soul "folk-dualism" which of course has crept into Christianity as the idea of heaven being disembodied souls playing harps on clouds. The Biblical resurrection of the dead is, of course, a different proposition from this folk conception. Theology might actually be better placed to take on this more materialist reality than we realize; it is the folk-dualists who will really have trouble with it. Unless, of course, the dualism become one of hardware and software, a metaphor that has already been spreading for years.

How many assumptions of Christian or theistic faith will be made confused or unintelligible to contemporary culture? Already, at least partially due to a technologically divergent cultural context, I am seeing in some of my students great difficulty in understanding basic theistic ideas. Is it because our religion is becoming outdated? Can Western religion be updated or has it run its course? Or are we just raising a feral generation who are quite capable of reading text on a screen or performing great feats at video games, yet do not understand even the basics of human life, relationships, and well-being, much less of history, philosophy, or culture?

Conclusion

There are many more reflections to make, but this paper is already too long. So I will here conclude.

Artificial intelligence, like any other technology, will just give us more of what we already want. Whereas we could once have only a trickle of what we wanted out of life, and the powerful took what little there was, now we have a firehose of wants being fulfilled (entertainment, pornography, food, drugs, gamified feelings of accomplishment, etc.). The firehose will become a deluge washing away our desires and leaving just what, exactly, of us behind? What skeleton of humanity will remain when technology has given us, or perhaps distorted or replaced, all our fleshly desires? What will this skeleton of humanity be made of? Will our technological flesh truly satisfy us, or just leave us in a deeper existential malaise, filled with angst, despair, and dread? What will we want, when we want of nothing? What of human nature will remain once our every worldly *telei* is fulfilled? Perhaps only the worst of us will remain. Or perhaps the best. Or, as always, both.

We are grasping ourselves by our desires. Or at least some our desires. Will this be good for us? Will it destroy us? Should we want these things? How could we know what we should want?

We are conducting this experiment called human history, and no one yet knows how it may end. But as we proceed, we can hope that artificial intelligences of our own fashioning will help us, and not harm us, as we go. To go beyond mere hope, into ethics and action, is the responsibility of those who are able to affect the necessary changes to make a better future.