Journal of Moral Theology, Vol.11, Special Issue 1 (2022): 1–12

An Introduction to the Ethics of Artificial Intelligence

Matthew J. Gaudet

THE ORIGINS OF THIS SPECIAL ISSUE OF the *Journal of Moral Theology* can be traced to 2018, the year I joined the faculty of the Santa Clara University (SCU) School of Engineering. Moving to SCU allowed me to reconnect with my old friend Brian Green (co-editor for this issue) who, also in 2018, was named the first Director of Technology Ethics for the Markkula Center for Applied Ethics at SCU.[1] Holding our new respective positions in technology ethics at SCU, Brian and I immediately began to discuss possible collaborations.

In 2018, I had also just wrapped up co-editing my first special issue of the *Journal of Moral Theology* (JMT),[2] was working on the second,[3] and had recently joined the JMT editorial board. As Brian and I narrowed our focus to collaborating on bringing more Christian moral theology to the subject of Artificial Intelligence, a special issue of the JMT seemed an excellent outlet and Jason King, then Editor of the journal, agreed. The JMT had released an issue on technology ethics in 2015, in which issue editors Jim Caccamo and David McCarthy noted that "we stand squarely in the midst of the digital era," but "scholarly articles [on technology ethics] from the theological disciplines [were] few and far between." [4] Today, technology has only marched onward, while responses from Catholic moral theology remain "few and far between."[5] In particular, artificial intelligence (AI)

[1] For an overview of the work accomplished at the Markkula Center, see Brian Patrick Green, David DeCosse, Kirk Hanson, Don Heider, Margaret McLean, Irina Raicu, and Ann Skeet, "A University Applied Ethics Center: The Markkula Center for Applied Ethics at Santa Clara University," *Journal of Moral Theology* 9, special issue 2 (2020): 209–28, jmt.scholasticahq.com/article/18042-a-university-applied-ethics-center-the-markkula-center-for-applied-ethics-at-santa-clara-university.

[2] Matthew J. Gaudet and James Keenan, SJ, eds., "Contingent Faculty," *Journal of Moral Theology* 8: special issue 1 (2019).

[3] Matthew J. Gaudet and James Keenan, SJ, eds., "University Ethics," *Journal of Moral Theology* 8: special issue 2 (2020).

[4] James F. Caccamo and David Matzo McCarthy, "Notes from the Issue Editors," *Journal of Moral Theology* 4, no. 1 (2015), i.

[5] A notable exception to this claim are the papal encyclicals *Laudato Si*' (2015) and *Fratelli Tutti* (2020), in which Pope Francis calls out the "technocratic paradigm" of

has entered our everyday lives like never before. To understand how quickly technology is moving, consider that in the 2015 JMT technology issue "artificial contraception" was mentioned more often (twice) than "artificial intelligence" (once).[6] In 2015, public imagination still viewed AI as a technology of the future, a novelty for tech enthusiasts, but unimportant to the broader public. That year, the major AI news in the mainstream media was that Google DeepMind's AlphaGo became the first AI system to beat a human professional at the game of Go by deploying machine learning to develop a winning strategy.[7] Today, AI is no longer limited to experimental labs and board games, but used routinely throughout our society, in ways both big and small, opaque and transparent, benign and violent. AI powers devices that determine everything from the settings of our thermostats and our driving routes to admission into elite schools, jobs, and prison sentences. The time is ripe for a sustained conversation on what Catholic moral theology can and should say to a world replete with artificial intelligence, and we are grateful to the authors of this issue and their broader interlocutors, for their contributions and leadership on addressing these issues.

This goal in mind, we bring you this special issue of the *Journal of Moral Theology* on Artificial Intelligence. It is intended, first, to reflect the ongoing conversation in AI ethics; second, to offer a set of Christian contributions to that conversation; and third, to serve as both an entry point and invitation for the AI novice to engage this topic.

---

our modern world as being particularly problematic for both the sustenance of our common world (*Laudato Si'*, nos. 101ff.) and the care for human relationships and community within that world (*Fratelli Tutti*, nos. 18–36, 164–69). However, Francis's appeals to rethink the moral and social force that technology wields upon us today are, in fact, the exceptions that prove the rule, since even the Pontiff's lead has not drawn a significant wave of moral theologians into deep reflection on these questions.

[6] Kara N. Slade did author an entire article on autonomous drones, a notable AI-based technology, but the fact that she did not call autonomous drones "artificial intelligence" is itself a marker of how the term "artificial intelligence" was being received a mere seven years ago. Caccamo and McCarthy, on the other hand, do make the only reference to artificial intelligence in the issue when describing Slade's paper in the introduction, indicating that the term was beginning to circulate more, though not at the level it currently does. Google nGrams data confirm this, showing a peak in usage of the term in 1987 followed by a steep decline to a trough in the first decade of the 2000s, then another exponential uptick beginning in 2011. See Kara N. Slade, "Unmanned: Autonomous Drones as a Problem of Theological Anthropology," *Journal of Moral Theology* 4, no. 1 (2015): 111–30, jmt.scholasticahq.com/article/11278-unmanned-autonomous-drones-as-a-problem-of-theological-anthropology; Caccamo and McCarthy, "Notes from the Issue Editors," ii; and Google nGrams, "Artificial Intelligence," books.google.com/ngrams/graph?content=artificial+intelligence.

[7] Cade Metz, "Google and Facebook Race to Solve the Ancient Game of Go with AI," *Wired* (December 7, 2015), www.wired.com/2015/12/google-and-facebook-race-to-solve-the-ancient-game-of-go/.

THE CENTRALITY OF CONVERSATION IN AI ETHICS

Ideas lose some of their potency when disassociated from their context. It matters that this issue is written at a time when AI is increasingly moving beyond science fiction and into current events. It matters that this issue was conceived by two technology ethicists working in the heart of Silicon Valley. It also matters that neither of us comes from computing backgrounds, in part because, we understand through layman's eyes the importance for the public at large to learn, understand, debate, and act in response to artificial intelligence.[8] Conversely, it matters that both of us *do* have training in both theological ethics and science/engineering, which means we also understand how morally impoverished the development of technology becomes when it is divorced from ethical and theological reflection.

Finally, it matters that this issue of the JMT is the third I have guest edited; those first forays taught me that tremendous value is gained when the papers of an edited volume (special issue of a journal or collection of essays in book format) are reflective of an ongoing conversation between the selected authors. The articles then engage each other in functional and constructive ways, the authors cite and build on each other's ideas, and the collection of essays is simultaneously varied and holistic.

Fortunately for us, as we were preparing the call for papers for this issue, the Pontifical Council for Culture had partnered with the Markkula Center to bring some of the leading Catholic theological voices on AI to SCU's campus for a two-day symposium in March 2020. We knew this symposium could offer the connective tissue this volume needed. Fate (or perhaps grace?) intervened and the symposium had to be shortened and moved online as the world adapted to the reality of the global COVID-19 pandemic. Rather than a single two-day symposium, the online conversation continued with regular monthly meetings for over two years![9] Those ongoing conversations appear in this

[8] I have already noted that my entry point into tech ethics was mechanical engineering, Brian was trained in biology, which is not to say that this issue does not rely on computing expertise. John Slattery, Mark Graves, and Noreen Herzfeld all hold degrees in computer science.

[9] It is necessary to acknowledge the full list of conversation partners present in these discussions, for even though only some are listed in the table of contents of this issue, and a few more make it into formal citations, it goes without saying that the ideas of all participants are present throughout this volume. To make the conversations manageable, regular monthly meetings of the working group were broken into three subgroups, with an annual plenary meeting to consolidate the ideas. The first subgroup focused on "Consciousness, Interiority, and the Soul" and comprised Brian Cutter, Marius Dorobantu, Justin Gable, Anselm Ramelow, OP, Marga Vega, and Jordan Joseph Wales. The second subgroup focused on "Relationality and AI" and comprised Levi Checketts, Marius Dorobantu, Noreen Herzfeld, Cory Labrecque, and Jordan Joseph Wales. The third subgroup focused on "Society, Ethics, and Politics" and comprised David DeCosse, Mark McKenna, Matthew J. Gaudet, Veronica Martinez, Paul

current volume in two important ways. Four of the seven peer-reviewed articles in this issue are authored by regular participants in that working group (Andrea Vicini, SJ, Noreen Herzfeld, Levi Checketts, and Jordan Joseph Wales). Second, we have included two non-peer reviewed articles purposefully reflecting the actual conversations of that working group, and broadening the range of voices to include several who did not otherwise write for this issue (Brian Cutter, Cory Labrecque, Anselm Ramelow, OP, Paul Scherz, Marga Vega, and Bishop Paul Tighe).

A Primer on Artificial Intelligence

Before turning to the issue itself, let me first offer some brief definitions and concepts to help orient the AI novice. There is no doubt the topic of AI can be overwhelming in scope and technically daunting to those who do not already have knowledge of and interest in the technology field. Given the scope of AI's influence on our contemporary and near future society, it is absolutely vital that the general public gain fundamental understanding of the moral implications of this topic. Fortunately, there is a growing recognition that we need more ethical discussion on technology and that such a discussion cannot be restricted to a knowledgeable elite. AI is in our lives, and we must engage it morally and socially. To engage it, though, we must first make sense of it.

Our first task in making artificial intelligence accessible is to define terms. First, it is helpful to distinguish between several forms of artificial intelligence. **Artificial intelligence** is the general category including all machines or software capable of performing tasks commonly associated with intelligent beings, including learning, reasoning, problem solving, perception, and using language.[10] **Machine learning** (ML) is the subfield of artificial intelligence in which a computer "learns" how to do its task by analyzing either a set of training data or its success and failures in prior iterations of its task or both. For example, a text recognition program using machine learning might be "trained" with a set of millions of examples of text. In observing the data, the machine will learn the patterns that make certain letters so that it can recognize those letters in different fonts, handwriting, or other applications.[11] **Supervised machine learning** begins with

Scherz, Ann Skeet, Andrea Vicini, SJ, and Warren von Eschenbach. Brian P. Green, Angel Gonzales-Ferrer, and Bishop Paul Tighe were the organizers and sponsors of the working group and generally attended all three subgroups. We are indebted and grateful to each and every one of these partners for their contributions to this issue and the ongoing conversation.

[10] B. J. Copeland, "Artificial Intelligence," *Encyclopedia Britannica,* www.britannica.com/technology/artificial-intelligence.

[11] See Shan Carter and Michael Nielsen, "Using Artificial Intelligence to Augment Human Intelligence," *Distill*, December 4, 2017, distill.pub/2017/aia/.

humans defining categories and "coaching" an algorithm toward correct solutions and pattern recognition by tagging training data with correct solutions. In an example familiar to most readers, Google uses the human inputs we give to its reCAPTCHA program (e.g., those puzzles that test if you are human by asking you to "find the boxes with crosswalks or traffic lights in this picture") to train other AI systems on its network, like the tagging function in Google photos or photo sensors in its autonomous vehicle project Waymo (hence, crosswalks and traffic lights!).[12] By contrast, **unsupervised machine learning** discovers its own patterns (without human coaching or input) within a given data set and then utilizes those patterns to solve problems.

Artificial neural networks (ANN) were developed to mimic the way in which neurons work in a human or animal brain. Neural networks consist of algorithms organized to process information by feeding it through layers of "neurons" to come to a deeper understanding of an observation.[13] **Deep learning** (DL) is the subset of machine learning that deploys multi-layered neural networks in its learning process. One example of deep learning can be found in the image that adorns the cover of this issue (if you are holding the print copy) or the masthead of this introduction (if you picked up this article from the open source JMT website). This image was created using a deep learning tool called Deep Dream Generator, which applies deep learning to learn the style of a particular piece of art and then is capable of converting any other image into that "style."[14] For the cover image, we used a photo of St. Peter's Square in Vatican City and converted it to the "style" that the Deep Dream Generator saw in Van Gogh's famous painting "Starry Night." Other examples of Deep Dream generated images adorn the other articles of this issue on the *JMT* website.[15]

Many who enter the contemporary discussion of AI realize that artificial intelligence has not taken the form of humanoid robots predicted in science fiction for decades.[16] Such robots would be a form of what had been termed **general AI** or **Artificial General Intelligence**

---

[12] Rugare Maruzani, "Are You Unwittingly Helping to Train Google's AI Models?," *Towards Data* Science, January 26, 2021, towardsdatascience.com/are-you-unwittingly-helping-to-train-googles-ai-models-f318dea53aee.

[13] Later in this issue, Jordan Joseph Wales argues that the symbolic representation of the world neural networks create can be something of a spiritual lens that leads us to deeper wisdom,"Metaphysics, Meaning, and Morality," *Journal of Moral Theology* 11, special issue 1, (2022): 157-81.

[14] See Deep Dream Generator, deepdreamgenerator.com/.

[15] See jmt.scholasticahq.com/issue/.

[16] E.g., Rosie the Maid in *The Jetsons*, C3-PO and R2-D2 in *Star Wars*, Data from Star Trek, KITT the talking car in *Knight Rider*, HAL in *2001: A Space Odyssey.* Artificial General Intelligence is also sometimes termed "Strong AI" but we will avoid that terminology in this volume because it implies that more narrow applications of AI are "weak," an inaccurate and problematic labeling.

(AGI), a computer capable of adapting to any task given, just like a human. Such humanoid robotic form is highly unlikely without radical technological advancements. The reality is that AGI will more likely take the form of vast datacenters or be distributed across networks of computers.[17] Such AGI is still theoretical, but at least seventy-two different organizations are working to make it a reality and several, such as DeepMind and OpenAI, have deep pockets.[18] If AGI ever does come to be it will require significant theological discussion about AI personhood, robot rights, human-AI relationships, and so on.[19] These questions only get stickier if the capacities of AGI reach **superintelligence**, the point where artificial intelligence *surpasses* the capacities of human intelligence.

While AGI remains largely theoretical, today applications of what is known as **narrow AI** are increasing exponentially. These applications are "narrow" in that computational intelligence is used for a very specific task or set of tasks.[20] Familiar examples of narrow AI include the algorithms that power Google's web search or Facebook's ad targeting. Digital assistants, such as Amazon's Alexa, Apple's Siri, or Google Assistant—though seemingly capable of near general AI—are actually just integrating several different forms of narrow AI, including voice recognition, textual autocomplete, geolocation mapping (e.g., when Apple Maps "learns" the commute you take regularly), and biometric tracking (e.g., when your watch identifies that you have been sitting for too long). Theologically and ethically, narrow AI does not force us to wrestle with notions of agency or personhood in the same way AGI might, but this does not mean narrow AI is not posing ethical questions. Emerging narrow AI applications include facial or other biometric data recognition raising significant privacy concerns, prison sentencing algorithms raising questions about the necessity for

[17] "Smart" hardware like home assistants do very little computing within the device. They are actually just data conduits, sending requests to data centers where the actual computation is done before solutions are returned to the device.

[18] McKenna Fitzgerald, Aaron Boddy, and Seth D. Baum, 2020 Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy, *Global Catastrophic Risk Institute Technical Report* 20-1, gcrinstitute.org/papers/055_agi-2020.pdf.

[19] Roberto Dell'Oro begins to take up the question of personhood in "Can A Robot Be A Person? De-*Facing* Personhood and Finding It again with Levinas," *Journal of Moral Theology* 11, special issue 1 (2022): 132–56.

[20] Science fiction has given us models for embodied AI in humanoid form (such as Rosie, Data, and C3-PO), embodied AI in non-humanoid form (R2-D2 and KITT) and seemingly non-embodied forms (HAL) but even in cases deprived of a human body, other humanoid traits, abilities, and characteristics remain a staple of the genre (e.g., HAL or KITT's voice, R2-D2's emotionally charged language. Green and I actually disagree as to whether HAL is embodied—he sees the removal of chips to disable HAL as akin to a lobotomy. Such are the more entertaining, but less consequential debates in the field of technology ethics). Narrow AI, on the other hand, rarely wastes computing power on trying to appear human (with the notable exception of the voice in digital assistants like Amazon's Alexa or Apple's Siri, designed to pass as AGI).

compassion in our systems; and autonomous vehicles placing the (sometimes life and death) driving decisions into the hands of sensors and algorithms, and many more.

Having laid out these basic definitions, our second task in this primer is to briefly summarize some of the major ethical and theological issues AI raises. Since narrow AI is capable of completing tasks faster and with fewer errors than humans, many of the moral problems related to AI are simply exacerbations of moral issues already present in our society. Among the most prevalent of these is the problem of bias.

A machine learning algorithm can only be as good and reliable as the data set it is trained on. If the algorithm is set up to learn from interactions with our real, sinful society, it will naturally come to reflect the inherent biases of that society. When Microsoft connected Tay, an ML driven chatbot, to Twitter and used its exchanges on the social media platform to "learn" how and what to tweet, within hours Tay was spewing racist and misogynist tweets.[21]

In theory, with the correct training an ML algorithm should be *more* apt than a human actor at avoiding bias, since it has no subconscious informing its results. Biases are sometimes so baked into our society that even in cases where ML is trained on a selectively screened data set and restricted from using certain categories—like race or gender—to make its determinations, machine learning often finds proxies that bias the final results anyway, as was the case when a prison sentencing algorithm used zip code instead of race to predict recidivism,[22] or an Amazon application screening algorithm used certain keywords (like the names of all-women's colleges or participation in certain clubs or sports) rather than gender as a means to maintain the glass ceiling.[23] The problem of bias is often compounded in systems using deep learning because the connections that neural networks make through the deep layers are often opaque to human observation, precluding easy verification that the logic leading to a solution is biased. A cautionary tale often told in ML literature tells of a defense contractor tasked with building a targeting algorithm for autonomous weapons to recognize enemy tanks, discovering that the photos of

---

[21] James Vincent, "Twitter Taught Microsoft's AI Chatbot to Be a Racist Asshole in Less Than a Day," *The Verge*, March 24, 2016, www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.

[22] See Ellora Thadaney Israni, "When an Algorithm Helps Send You to Prison," *New York Times*, October 26, 2017, www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html.

[23] See Jeffery Dastin, "Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women," *Reuters*, October 10, 2018, www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G. In this issue, Andrea Vicini, SJ, examines several applications of AI in which societal bias finds its way into the algorithm ("Artificial Intelligence and Social Control," *Journal of Moral Theology* 11, special issue 1 [2022]: 41–69).

tanks used to train the neural network had been shot on sunny days, while the photos without tanks in cloudy conditions, leading the deep learning network to use brightness as criterion to determine the presence of a tank. In opaque systems it can be very difficult to find such errors.[24]

Beyond bias, we also must be cautious about how AI removes from more traditional systems some of the friction inadvertently rendering the system more moral. For example, as deadly as war can be, the amount of destruction is reduced simply because some people refuse to act.[25] When AI is deployed in autonomous weapons systems, it removes any hesitancy soldiers might have in killing another human being,[26] thereby eliminating the friction and making warfare more efficient. Is greater efficiency or ruthlessness at killing actually the more moral course? Could there be goodness in the friction?

Similarly, when AI speeds up processes otherwise impossible for individual humans to complete, it can remove barriers inherent to the system itself. For example, the central premise of an insurance system is that it allows individual risk to be spread across a large pool of contributors. When we pay our life insurance premiums, most of us actually get paid out less in death benefits than we paid in over the course of our lives. The beneficiaries of the rare person who dies an early death get the benefit without paying decades of premiums. In this way, the risk of an early death is shared and spread over the entire pool. Medical insurance works similarly: we (or our employers) pay medical premiums at usually much greater cost than the medical expenses we incur in a year. The surplus is used to pay for the small set of people who end up with serious medical conditions and high medical costs. In this way, the entire pool of contributors shares the risk, even though only a few "profit" in the sense that they get more out than they put in. Traditionally, life insurance or medical insurance premiums could vary on the basis of general factors such as age or smoking, but beyond these generalities, one cannot predict who will die young or suffer from the expensive medical condition, only that *someone* in the large pool will die early or require significant medical care. The use of AI

---

[24] James Bridle, "Known Unknowns," *Harpers Magazine*, July 2018, harpers.org/archive/2018/07/known-unknowns/. Some have argued that the story is apocryphal; whether the story derives from actual events is unimportant if the story is intended to illustrate problems with opaque neural networks (see Gwern Branwen, "The Neural Net Tank Urban Legend," www.gwern.net/Tanks).

[25] One study showed that as many as 80% of American infantry soldiers in World War II never fired their weapons in combat. See Fredric Smoler, "The Secret of the Soldiers Who Didn't Shoot," *American Heritage* 40, no. 2 (1989), www.americanheritage.com/secret-soldiers-who-didnt-shoot.

[26] For a detailed examination of the moral status of Lethal Autonomous Weapons Systems, see Noreen Herzfeld, "Can Lethal Autonomous Weapons Be Just?," *Journal of Moral Theology* 11, special issue 1 (2022): 70–86.

to process medical and other factors increases the specificity with which an insurance company can predict those who will die early or have high medical costs, and consequently charge them with higher premiums or exclude them from getting insurance altogether. AI powered insurance systems are more efficient than traditional systems, but by being so, they lose the randomness that made these systems more morally acceptable. With AI powered systems, however, insurance systems no longer pool the risk of death or severe medical condition; instead, they actually *remove* the risk by excluding the most needy individuals from the system. Morally, this raises serious questions both about the common good and the Catholic principle of the preferential option for the vulnerable.[27]

Finally, AI powered systems have heightened questions regarding personal autonomy and privacy. Today, nearly every purchase you make, every term you search, every location you map, and every link you click is tracked to help companies build a profile that can better target the advertisements you see. These mountains of data would be overwhelming if they had to be organized by hand, but through the deployment of AI to sift and sort the data, the reality today is that Google, Amazon, and Facebook often know us better than ourselves and use this profile not for our good, but their profit. The problem only gets worse as we connect more and more "smart" devices: home assistants collecting our voices, connected refrigerators and toothbrushes monitoring our daily patterns, robot vacuums mapping our homes[28] (this set of devices is collectively known as the **internet of things** or **IoT**). These are the data we *voluntarily* provide to these companies through our searches and clicks on own smart devices. As smart doorbells, traffic cameras, and other surveillance systems become ubiquitous throughout our cities and suburbs, we need also be concerned about the troves of data gathered by these cameras. We should remember that AI systems make surveillance capable of integrating and processing data drawn from across an entire city. AI also expands the markers by which an individual person can be identified; in addition to the traditional means of face, voice, or handwriting recognition, AI is now capable of uniquely identifying you by your speaking or writing style, heart rate, or even your gait. The set of decisions or actions we can make without being surveilled is ever shrinking. Serious public discussion must be raised about how this data is used.

---

[27] For a more extensive analysis of Catholic social thought as it applies to AI today, see Levi Checketts, "Artificial Intelligence and the Marginalization of the Poor," *Journal of Moral Theology* 11, special issue 1 (2022): 87–111.

[28] See Maggie Astor, "Your Roomba May Be Mapping Your Room, Collecting Data That Could Be Shared," *New York Times*, July 25, 2017, www.nytimes.com/2017/07/25/technology/roomba-irobot-data-privacy.html.

The moral issues I have identified thus far only relate to what narrow AI is already capable of. But hardware companies like IBM are working on developing **quantum computers** which process data using quantum bits (or qubits) instead of the standard binary bits used on digital computers. Qubits are superimposed on one another allowing computers to process data millions of times faster than digital processors. The possibility of developing general AI or superintelligence will raise deep theological and philosophical questions about the nature of creation and the place of humans, AI, and God in that creation. Will AGI be worthy of some or all of the rights and protections we ascribe to humans? Will it require us to develop new and different rights or moral principles? Will the creators of AI be like gods to their intelligent computers or, if we reach superintelligence, will AI become a god to us? The advance of technology has, throughout history, challenged our understanding of the divine; these advancements may shatter our current comprehension of the relationship between the Divine creator God and creation itself and strain our theology in novel ways.

Moreover, the above remarks presume AI and humanity will remain distinct, which is unlikely. Even today, there is a growing discussion about **transhumanism**, the movement to integrate technology and the human body to enhance human capacities. As these applications increase it will raise justice questions about who has access to such augmentations and what happens to those who are not "lifted" in such a way. Some even claim that such technologies pave the way to extending human life, even indefinitely, raising further moral and theological questions about the nature of death and the afterlife. If a human could live forever, what would this imply for the Thomistic presumptions of *exitus-reditus* (we come from and return to God) or the Augustinian notion that those who follow Christ in this world are akin to travelers in a foreign land, working our way home? Is transhumanism the path to spiritual immigration away from the City of God? These are deep theological questions we must begin to contemplate if Catholic theology is to be prepared for what is to come.

The Structure of this Issue

Fortunately, some of us have already begun to ask these types of questions. The present issue of the *Journal of Moral Theology* gathers some of these reflections. The topic of AI is vast. The theological dimensions of and the moral challenges wrought by AI are extensive. There is simply no way to capture that vastness in a single volume. The metaphor of an hourglass conveniently describes the structure of the present issue. Our first task must be to funnel the reader toward a narrow neck of information, without losing essential elements of the conversation and debates we, as a society, need to have. This introduction serves as the first part of that funnel both by offering a brief primer on terminology and concepts and orienting the reader to the

structure of the issue itself. Next, we offer the first of two non-peer-reviewed articles in the issue. In this article, we (quite literally) attempted to capture the salient aspects of the conversations the PCC working groups have had. In order to respectfully reflect some of the tensions and debates inherent to the ongoing conversation, we invited nine members of that body to engage in an online written conversation reflecting the work we have been doing over the past several years. Brian and I have moderated the conversation by offering initial questions, collecting answers, editing responses for saliency and overall flow, and then inviting the participants to engage the conversation again. Several iterations of this process were completed to allow the participants to fully engage and debate each other. In the end, our hope is that this conversation provides an introduction to some of the important questions AI poses for us and the variety of responses available.

Following the conversation paper are seven peer reviewed articles. The first four address one or more current applications of or moral issues related to artificial intelligence through the lens of an established tradition of Catholic ethics. Andrea Vicini, SJ, uses Pope Francis's theology (and the "Rome Call for AI Ethics") to analyze the ethics of facial recognition systems, the use of AI in judicial sentencing, and the use of AI in job hiring. Noreen Herzfeld applies the just war tradition to the recent emergence of AI-driven lethal autonomous weapons systems on the battlefield. Levi Checketts asks what Catholic social thought has to say about the effects of increasing usage of AI on the poor and marginalized. Finally, John Slattery takes aim at the persistent moral problem of gender and racial bias in AI systems with a theological critique drawn from M. Shawn Copeland's womanist theology. Taken together, these four articles offer an excellent sampling of the moral issues being debated under the current state of (narrow) AI development as well as a demonstration that established Catholic moral thought already has much to contribute to such debates.

In the next three articles, the conversation begins to move from asking how theology might inform the ethical use of AI to how theological questions about AI might inform our ethics. Roberto Dell'Oro uses the theological anthropology of Emmanuel Levinas to take up the classic question of whether a machine can achieve the moral status of personhood. Next, Jordan Joseph Wales challenges those who suggest that AI is merely a tool, incapable of anything more than expressing the will of its programmer(s). Employing an Augustinian theology of the natural world, Wales argues that complex computational processes (especially those black box deep neural networks leaving humans unable to understand how a solution was reached) do constitute a significant interpretive layer that "stands between" us and the world we seek to understand. In the final peer reviewed article of the issue, Mark

Graves attempts to articulate a "pragmatic theological anthropology" specifically adapted to thinking about artificial intelligence.

These seven articles merely scratch the surface of the conversation Catholic moral theology needs to be having with broader society about the continuing development of AI and the expanding integrations of AI into our personal and social lives. While the first two pieces in this issue aim at bringing a vast topic down to a narrow neck, the final two articles aim to widen the scope once again, connecting the wisdom present in this volume to the wider world and the questions and conversations we could not include here. First, we have an interview with Bishop Paul Tighe, the Secretary of the Pontifical Council for Culture and one of the leading Vatican voices on the moral and theological questions related to technology and AI specifically. He also convened the working group from which many of these papers emerged. In the interview, conducted by Brian Green, he provides a clear account of current Vatican thinking on the ethics and theology of AI. Following this interview, my co-editor Brian Green offers an epilogue to the entire issue. Just as this introduction has attempted to guide the reader from a dauntingly broad and deep topic down to those aspects most salient and ripe for discussion, the epilogue's function is to return the reader back to world of AI beyond these pages, and especially the problems we see lurking on the horizon.

In summary, the time has come to recognize, first, the capacities AI already has brought to the world and the moral challenges these capacities raise, and second, the exponentially greater potential capacities that will put our foundational theology to the test. We hope this issue serves as both a challenge and a resource to Catholic theologians, ethicists, technologists, and the Catholic faithful, as well as to all people of good will, as we begin to address this difficult topic. M

Matthew Gaudet is a Lecturer of Ethics in the School of Engineering at Santa Clara University and a Fellow at the Grefenstette Center for Ethics in Science, Technology, and the Law. In addition to his work on tech ethics, Gaudet also works on issues of university ethics, disability ethics, and the ethics of war and peace. He has written for and edited several previous issues of the *Journal of Moral Theology* among other outlets.