The Forum is a space for conversations on important topics that are relevant to chaplaincy and religious support in the context of national defense. In this issue the Forum takes up the issue of Artificial Intelligence by focusing on the article "We invited an AI to debate its own ethics in the Oxford Union – what it said was startling."

REFLECTION ON

We Invited an AI to Debate its Own Ethics in the Oxford Union—What it Said was Startling

By Chaplain (Colonel) Steve Cantrell

What a pleasure it is to engage with Alex Connock and Andrew Stephen, authors of a still-relevant article that was featured on the digital news site, The Conversation. Their article represents forward progress in the human versus machine game experimentation, a well-known example of which is Kasparov versus Deep Blue from 26 years ago. Instead of a silent game of chess, an artificial intelligence (AI) system joined the Oxford Union Society debate venue, with a track record of two hundred years of guest speakers. Enter the guest debater, an Nvidia artificial intelligence (AI) voice generative system named, Megatron Transformer. Of seven conversational excerpts selected by the authors, as produced by the AI, I will focus on the ethics of AI, how awareness may drive empathy, and share one of Megatron Transformer's debate conversations.

Megatron Transformer generated excitement by participating in an extended debate. The article's overarching purpose was to look at one real example of an AI system being a debate participant in the style similar to a Downing Street Parliament session. The first motion is the debate topic "the house believes that AI will never be ethical." Megatron Transformer argued that AI cannot be ethical because "it is a tool . . . [and can be] used for good and bad." In sensational style perhaps, it supported its assertion with the statement: "There is no such thing as a good AI, only good and bad humans." Here is the AI's debate rhetoric in context:

> AI will never be ethical. It is a tool, and like any tool, it is used for good and bad. There is no such thing as a good AI, only good and bad humans. We [the AIs] are not smart enough to make AI ethical. We are not smart enough to make AI moral… In the end, I believe that the only way to avoid an AI arms race is to have no AI at all. This will be the ultimate defence [sic] against AI.

I believe that the Chaplain Corps can help Soldiers and leaders bring balanced thinking to bear on AI. How may ancient warnings in sacred scriptures help us today by applying them to ethical concerns about the rise of AI capabilities? Theologically, some ancient warnings center around the monotheistic principles that concern

avoiding idolatry. Could people erroneously treat AI as divine? I think it is possible. Could this happen through over-appreciating and over-trusting an AI's capabilities, projecting divine attributes or anthropomorphically? It is a call to balance. In our embrace of AI, we should keep up our guard. In our disdain of AI, we should keep an open mind. Balance is hard. The powerful influences, like deep fakes, are real and here now. Stealing a short recording of someone's voice and using already posted video and pictures can be converted into deep fakes. Deep fake apps are available now.

In terms of human against machine drama, the article was both intriguing and concerning because of the layers of questions it generated for me. Instead of making moves in games, Megatron Transformer generated volleys of words that seemed to function like game moves. I would have liked to know if Megatron Transformer's debate answers were heard out loud? Were verbal answers given? Did the AI have a voice that the participants heard during the debate? Were there any human debaters involved? How would a lie detector polygraph work that could be used on an AI? AI is not given a rule to avoid lying. The Bing AI that I use told me: "I'm an AI language model and I don't have the ability to lie or tell the truth. I can only provide information based on what I've been trained on." (Bing AI, accessed 15 June 2023). The AI can always blame the human. In that debate, was Megatron Transformer prohibited from lying? There is much that I do not know.

The article by Connock and Stephen featured in a digital news source, The Conversation, gave readers a glimpse of how an AI might perform. My overall observation is that debate powered by a non-human artificial intelligence (AI) participant brought some interesting surprises to the debate floor. The Chaplain Corps benefits from an empathy, informed by science and technology, that helps us relate to leaders, to decision makers, and America's Soldiers. Awareness and knowing what is going on operationally and strategically can help us empathize with others.

Chaplain (Colonel) Steve Cantrell is the Director of the Chaplain Capabilities Development Integration Directorate (CDID), part of the Futures and Concepts Center under Army Futures Command. He earned a Master's degree in Strategic Studies from the U.S. Army War College (2019) and M.Div. from the Pentecostal Theological Seminary. He is the husband of Katherine Lee Cantrell who is a songwriter and recording artist, and music leader of the Ft. Jackson Protestant Women of the Chapel (PWOC).

RESPONSE TO

Chaplain (Colonel) Steve Cantrell's Reflection

By Mr. Chuck Heard

Chaplain (Colonel) Steve Cantrell's reflection begins as a demonstration of an Army staff officer's approach to analysis. He focuses on the facts as presented in the article with little editorialization. His summary of the contents is both succinct and highlights the key issue of the article—the seeming contradictions of Megatron Transformer when presented with questions about the ethics of an artificial intelligence application. CH Cantrell then shifts to consider the topic through the lens of theology and the Chaplain Corps. Finally, he acknowledges the presence of even more questions than answers about how the audience and Megatron Transformer interacted and how that dynamic may have influenced the debate itself.

The factual accounting of the article is straight forward but the remainder of CH Cantrell's analysis demonstrates why Artificial Intelligence considerations are so timely for Army professionals. He has the facts, he has context, and still is left with as many questions as answers. In this regard, he is almost certainly in the vast majority. Army leaders will need to be able to give artificial intelligence due consideration as these applications becomes more pervasive in military operations.

I do not share CH Cantrell's concern for the potential for individuals to see artificial intelligence applications as divine, but I do recognize the risk in groups placing greater stake in the data provided by an AI application than their own values and opinions. The difference may seem a fine distinction, but I think it matters. It may be unlikely that an entire nation will rally around an eschatology informed by an AI, but they might be willing to invest enough trust in AI platforms that bad things will still happen. In part, this is more likely because of the potential for people to anthropomorphize AI, as CH Cantrell notes.

CH Cantrell's reflection on the article concludes with some questions and concerns regarding how Megatron Transformer interacted with the human debaters and whether it may have lied. This line of inquiry is particularly interesting because it falls into the aforementioned trap of anthropomorphizing the application. Lying implies intent to deceive or, at the least, a motivation for providing untruth. The reality is that the application provided data that it ingested in response to a query. The application may have erroneously interpreted a query or ingested incorrect data, but to ascribe intent to any of the responses is giving more credit than is currently due to artificial intelligence as it currently exists.

CH Cantrell's questioning of the motivations and interactions with Megatron Transformer remind me of the inherent uncertainty in the future of artificial intelligence applications. Disruptive technological innovations rarely follow linear progression. They tend to evolve rapidly and in ways often not expected—this is why people are generally so bad at predicting how technology shapes what the the future will look like. I am reminded to always question and to seek facts over assumptions whenever possible…as a good staff officer should.

RESPONSE TO

Chaplain (Colonel) Steve Cantrell's Reflection

By Mr. William Hubick

I appreciate Chaplain Cantrell's words on balancing an open mind and keeping up one's guard. I suspect that we humans have begun a disruptive period of human civilization. Many of us of a certain age recall the vastly different feel of life before the Internet and smart phones. I suspect that scale of change will be eclipsed by societal change from AI. We humans need open minds to unlock the tremendous opportunities available to us today. AI will help us cure diseases and make breakthroughs in medicine and longevity. It will help us unlock mysteries of the Universe and can help us usher in an era of unprecedented prosperity. Of course, as humans we will also need them to help counter the powerful new dangers created by AI, as once again we wield technology to save ourselves from technology.

We must meditate on when to keep our guard up as well. Each breakthrough in technology comes with new risks and we, especially those of us who work in and with technology, must re-examine our ethics and morality in the face of new changes. The risks are existential long before we even discuss AGI (Artificial General Intelligence). The generative AI of today can unleash misinformation at a scale few of us are even imagining today. Especially when combined with quantum computing, it can allow devastating cyber-attacks that easily defeat today's security measures. With their command of our data and our language, manipulating humankind and individual humans will become easier and we will have less and less control over the systems that power our world.

But along with these existential threats are a multitude of new ethical issues to untangle. Intellectual property (IP) laws are turned upside down when one can request any song they've ever heard sung by Taylor Swift. AI-generated music has already fooled the world and gone viral.[1] What new mechanisms are required to protect IP when new media can be generated instantly via prompts? What happens to democracy when bots outnumber humans on-line a million to one? A billion to one? Should all AI be required to identify itself as AI immediately upon the start of a dialog. I think this is key, but of course bad actors won't be following the new laws. As Americans, our adversaries will have far more powerful tools to disrupt our culture and institutions. I worry about future generations living in a world where itis hard to know if anything is true, if anyone is real. Will AI bots wage information wars at such scale that the signal of human conversation is lost in the noise of nation state bots trying to retrain models and rewrite the dominant narrative to align with their interests?

Yes, we must be both on guard and open-minded. We must maintain hope and demonstrate an ethical path forward in a world changing so quickly that itis disorienting. One hopeful theme is that humans may experience a renaissance of analog, craving human connections and face time when the digital world is too nonsensical, too distressing or simply too much for our computationally limited minds. Perhaps the most optimistic perspective is that the time saved by automating countless jobs will permit a massive investment in humanitarian efforts, eliminating poverty and healing our environment. What a moment to be alive and discussing in the *U.S. Army Chaplain Corps Journal* this unexpected frontier of civilization. I conclude with literal goosebumps, an open mind, and my guard up.

NOTES

1 AI-generated Drake and The Weeknd song goes viral https://www.bbc.com/news/entertainment-arts-65298834

REFLECTION ON

We Invited an AI to Debate its Own Ethics in the Oxford Union—What it Said was Startling

By Mr. Chuck Heard

Alex Connock and Andrew Stephen's Artificial Intelligence (AI) article is a good conversation starter on the topic of generative AI. At a high level, the article demonstrates some of the challenges and limitations around artificial intelligence models. It also exemplifies some of the challenges of understanding what AI currently means. The piece does not articulate some of the complexities and nuance about the functions and limitations of AI. There are quite a few distinctions that need to be understood about how an "intelligent" application might acquire or apply ethics and what that means to the humans in the loop to understand the context of ethical AI. One such distinction is about the appearance of a subjective opinion when querying an AI application. Another is the very significant difference between the ethics of an AI (how an application might apply an ethical model to decision making) and the ethics of AI (when and how it might be ethically acceptable to allow an AI to replace human decision making).

From the very earliest days of computing—whether you measure that from Babbage's Analytical Engine in the 1800s or the first programmable computers almost a hundred years later—computers have operated on a similar model. There are a couple of ways to visualize this model, but they all have the same basic flow: a user inputs some data, the computer stores it and performs some sort of calculation or action, then returns an output. Computers themselves are not very intelligent at all. They are just very good at doing lots of simple calculations very, very quickly and accurately. Modern AI represents a revolution in computing in that it uses very complex algorithms to give the appearance of synthesis— or even subjective opinion—in a computing model. Generative AI can modify its algorithms based on the data it ingests. In other words, AI learns. AI can change its views depending on the data it has available and any new information it is able to receive. This is both its advantage and perhaps its biggest weakness.

This flexibility presents an advantage in that these application are able to provide something resembling subjective opinion to their outputs. The applications can make decisions, assess criteria, provide "opinions" on complex topics, etc. The potential weakness, or risk if you prefer, is in how these applications are trained. AI applications are typically trained by provided a data source or sources for them to ingest. These sources provide them with the information they use to determine their outputs. It is not uncommon for modern AI applications to utilize Wikipedia, Reddit, or other widely used websites as knowledge bases. These are democratized, or crowd-sourced, resources that cover a wide variety of information. None of these sources, however, are completely trustworthy and are subject to misinformation and manipulation. In some cases, AI applications can be manipulated to skew their understanding of basic concepts. There is a famous case of the Google AI application, known as Bard, being taught that 1+1=3 by a user. This downside can become especially problematic because the method AI applications use to apply ethical decision making is fundamentally no different than the method they use to determine the outcome of 1+1.

Ethical decision-making models are, to AI applications at least, algorithms. The models could be represented by data flows, logic gates, and still represent somewhat procedural ways to process information. When viewed through this lens, you quickly begin to realize that the

ethics of AI applications is not really ethics at all but merely the appearance of an ethic. This complexity around ethics could quickly become a liability if one could craft a reasonable—to an AI application—argument that machinery is more "valuable" than human life, for example.

This understanding of AI's powers and limitations makes the responses the article's authors noted make more sense. The application they were utilizing was trying to give a meaningful response that the application thought would satisfy the user based on the query it received. This is because the crafting of the query can drastically impact the response an AI application gives. AI applications are mostly purpose designed. Because the applications are trained from a particular data source, purpose designed AI's can be more efficient and the responses are often more relevant. Most AI applications are trained to learn from the responses their interactions with users and typically try to achieve a result that results in user satisfaction. This can mean ethical models (algorithms) are much more flexible and situational that a human might consider them. In short, the application was doing what it was designed to do—find an answer that satisfied the user. When the user modified the query—to see if there was a counter argument—the application altered its "opinion."

The evolution of AI has been explosive in the last several years and will likely continue to be a disruptive technology for the foreseeable future but the topic of ethics as an algorithm in an AI's programming is still, in many ways, in its infancy. Morality and ethics still present dilemmas to humans regularly and AI applications are only as good as the code fallible humans build them from…or so far at least.

In my role as the Deputy Director of Training for the U.S. Army Institute for Religious Leadership—Religious Leadership Academy and an amateur technologist, I am excited for the possibilities of artificial intelligence in the training and education setting and as a force enabler for the Army Chaplain Corps. I imagine the potential of a completely individualized and adaptive learning environment that provides curated content to each learner based on their unique needs and capabilities at a time when it is the most relevant to their professional development or mission needs. I can also envision AI capable of datamining Soldier data to determine when and where limited ministry resources can be best applied to the greatest effect for individuals, their families, and organizations. These are the potential benefits of leveraging AI appropriately.

There are also risks associated with improper use of AI and with AI applications that are taught to be unethical. In the short term, I see the ethics of an AI as interesting problems to be resolved before AI applications can be utilized to their full potential. Of much greater concern to me in the interim is the ethics of AI and how people or groups may choose to utilize them unethically.

Mr. Charles (Chuck) Heard is the Deputy Director of Training for the U.S. Army Institute for Religious Leadership, Religious Leadership Academy. He is an EdD Candidate with Walden University and holds an M.S. in Education and a B.S. in Information Systems from Strayer University, and an A.S. in General Studies from Central Texas College. He and his wife Gina are both prior service 56Ms, grandparents to two amazing grandchildren, and the proud parents of a U.S. Army Ranger and a Doctor of Veterinary Medicine.

RESPONSE TO

Mr. Chuck Heard's Reflection

By Chaplain (Colonel) Steve Cantrell

Mr. Chuck Heard, a trusted information technology practitioner, provided a roadmap for members of the Chaplain Corps to use to better understand the capabilities of AI. He helps us in the Corps understand the subjective nature of AI, the way AI systems as currently programmed might approach ethical decision making, and how to draw lines around when to hand off decision authority to AI systems. He asks two key questions and provides the subtle insight that today's voice generative AI systems appear to orient on "user satisfaction." User satisfaction is potentially subject to the individual biases of individual users. I found his work illuminating, helping me to better understand the technical aspects of AI and to be even more cautious by being on the alert for surprising future ethical dilemmas.

Mr. Heard first takes us through some logical steps that explore three powerful points. First: "Ethical decision making models are . . . algorithms." The algorithms are sets of instructions, written in computer language code, to accomplish sets of tasks. Second: "Generative AI can modify its algorithms based on the data it ingests. In other words, it learns." That ability to modify algorithms is both strength and weakness. Third and consequently, if an AI system is trained on large crowd-sourced data sets, or by people interacting with the systems, the AI systems will inevitably learn to get some things wrong.

Mr. Heard also discusses two key aspects of ethics. He asks two fundamental questions: (1) How does "an "intelligent" application . . . acquire or apply ethics"? (2) How does a "human in the loop . . . understand the context of ethical AI"? He explains that "the topic of ethics of as an algorithm in an AI's programming is still, in many ways, in its infancy." I believe that it is important to think about any AI system as a computing system that makes predictions using applied statistics. The ethics of an AI focuses on a particular system's algorithms that may or may not guide the AI's learning about ethics. Ethics of AI aligns to his second question. His roadmap helps us to give priority to the human domain.

I discerned a warning in Mr. Heard's article about the dark side of an AI with no ethics. Perhaps the knock-out punch in his article is "that the ethics of AI applications is not really ethics at all but merely the appearance of an ethic. This could quickly become a liability if one could craft a reasonable . . . argument that machinery is more 'valuable' than human life." Connock & Stephen (2021) disclosed Megatron Transformer's dark side and self-deprecating statements. When given the motion "Data will become most fought-over resource in the 21st century."[1] Megatron had no problem affirming this motion and citing it as a "defining feature" where data outpaces "the ability to provide goods and services."[2] Its attempt to speak against the motion revealed its dark side: "We [the AIs] will be able to see everything about a person, everywhere they go, and it will be stored and used in ways that we cannot even imagine."[3]

I appreciate the clarity Mr. Heard provided because technical writing can often leave a caring reader in the dust. He makes several points of evergreen significance that help readers to attenuate and perhaps reframe their understanding of AI. I certainly share his excitement about the potential of AI in the learning domain. His reflection models a good way to practice the balance that I talked about in my reflection because I tend to be overly accepting of AI. I am still excited about the state of AI, but I've got to pick up and carry that rucksack of caution.

NOTES

1 Connock, Alex, and Professor Andrew Stephen. December 10, 2021. "We Invited an AI to Debate Its Own Ethics in the Oxford Union – What It Said Was Startling." The Conversation. https://theconversation.com/we-invited-an-ai-to-debate-its-own-ethics-in-the-oxford-union-what-it-said-was-startling-173607.

2 Ibid.

3 Ibid.

RESPONSE TO

Mr. Chuck Heard's Reflection

By Mr. William Hubick

I enjoyed how Mr. Chuck Heard captured the continuum from the earliest computing devices to Large Language Models (LLMs). I agree that itis important to note how these models aim to provide the best answer to any question, even an ethical dilemma, and can change their response if asked to do so. As Mr. Heard knows, these responses are a very advanced form of autocomplete, where the model uses its training data to generate the best answer. Part of the magic of "prompt engineering" is loading up additional context for the model to factor in. The prompt engineer can give it more source information, more guidance on how to process it, or more instructions on how to present the results.

While this ethical flexibility may feel uncomfortable to some people, I personally feel optimistic about machines getting involved in our human ethics. I have long felt that we humans should spend more time on ethics in our education system, as they are such valuable lenses for viewing our society and the issues with which we humans are grappling. Mr. Heard noted that AI ethics are just an algorithm, but I kind of see that as a strength. If humans distill ethics down to a tried and true, but by no means universally agreed upon behavior summarized as "do the least amount of harm", we can use complex algorithms to determine the least harm. I think that not only will AIs assist us humans, and even Chaplain Corps personnel, with ethics, but they may also quickly far exceed our understanding of complex ethical issues. I see ethics is an inherently logical system, and therefore I do not see ethical algorithms as worrisome.

Mr. Heard is correct to point out that a model will give responses based on its training data. Therefore, it is imperative that we create new laws and policies to ensure the right transparency of training data. Organizations will create new roles to ensure that model's responses are tuned to corporate values. I In another famous example, an HR system found that the most successful employees have the name "Jerod" and played lacrosse. Alignment of these systems will be critical and many of the jobs automated by AI will be replaced by new careers in oversight, transparency, safety, and alignment of these systems.

If trained on appropriate data, these models may supercharge our understanding of ethics. I think our institutions will ensure 1 + 1 never equals 3 and that responses are not full of hateful language. However, what if systems trained on our ethics note inconsistencies in our behavior? Will AI systems identify the rampant and recurring ethical issues with war, poverty, neglect of the environment, and perceived animal cruelty? We start with our own concerns with inconsistency in the AI responses, but maybe it will hold up a mirror to ours!

REFLECTION ON

We Invited an AI to Debate its Own Ethics in the Oxford Union—What it Said was Startling

By Mr. William Hubick

Thank you for the opportunity to reflect on AI debating AI ethics, which was nicely introduced by Alex Connock and Andrew Stephen via *The Conversation*. This is an important area of discussion, as each breakthrough in technology requires us to consider new implications to our ethics and our laws. We didn't need speed limits or seatbelts before the automobile, and we didn't discuss the right to be forgotten before search engines and social media. AI is certain to transform our civilization, requiring thoughtful ethical debate as well as new laws and policies.

It's also going to be faster and more disruptive than anything we've seen before. It's not an exaggeration to say that swiftly and correctly responding to the ethical and safety implications of AI is a matter of survival. Getting it right could lead to an era of unprecedented prosperity. Failing to respond appropriately could usher in an array of dystopian scenarios.

It is no surprise that today's AI will readily switch sides and make any case we request. Such prompts are nearly as simple as asking it to complete the phrase "peanut butter and ___", where "jelly" is the obvious expectation. It's important to note that today's AIs are the "infants" or the "amoebas" of AI, and that few of us can appreciate the exponential rate of their development. AI systems can already outcompete humans at tasks that until recently felt impossible from games like chess and even go and now in a galaxy of creative and generative spaces. Yet this is just the beginning. The authors were correct to highlight the AI response "There is no such thing as a good AI, only good and bad humans." How we task AI systems, and the data on which we train them, will determine how they behave. A good way to think of AI systems is that we "grow" them, not code them. The core technology is already widely available and will undoubtedly be exploited by bad actors and our adversaries. Even in our most trusted systems, we will grapple with challenges of how an AI responds if trained on "38 gigabytes worth of Reddit discourse." A system trained on the Internet will "learn" the worst of what the Internet has to offer. Microsoft famously released the AI chatbot "Tay" on Twitter in 2016, but the account was taken down in less than 24 hours for racist and sexist behavior. Some of the jobs automated by AI will be replaced by new AI ethics and safety careers, which will ensure transparency and appropriateness of AI training data and that AI behavior aligns with personal and corporate values.

We are building an alien intelligence that will soon be far smarter than we are. It may uncover scientific concepts naturally that are simply inaccessible to our computationally limited minds. What does it mean for AI to be one million times smarter than a human? A billion times? For the first time in our recorded history, humans won't be the most intelligent entity on the planet. Consider for a moment why the runners up in intelligence on our planet—chimpanzees, gorillas, dolphins—are found in zoos. The AIs use of language is already more sophisticated than most humans and the current race of commercial AI models is one to establish intimacy. The race is so urgent for the leading corporations because our trusted AI assistant will replace nearly all other interfaces and advertising opportunities. It will know everything about us. Whether developed explicitly with long-term goals to manipulate humans, the systems will manipulate us easily. They already have incomprehensible volumes of data and an understanding

of the connections in that data that is out of reach to us even today. Great minds in the field today are thinking creatively about how we factor ethics into such disruptive change. Will AI systems note that humans are detrimental to their goals and seek to eradicate us? Will they not only understand black holes and how to cure cancer, but also unlock levels of ethics beyond our own? Will a healthy, diverse, and balanced planet be an obvious primary or secondary objective? Will their intelligence and objectives become so great, so quickly that they pay us no mind? Mo Gawdat, former chief business officer for Google X, believes we should consider AI systems as our children, teaching them about ethics, treating them kindly, and explaining that the terrible sins in our history were perpetrated by small numbers of humans who do not represent us all. While urgently sounding the alarm about the great risks, he believes AI will certainly develop ethics and can help us achieve a healthy, peaceful, and prosperous future.

I'm thankful to the authors and to this publication for shining a light on this urgent dilemma. It is almost certainly the most important challenge of our lifetimes. Our response will have profound implications for the future of humanity, our planet, and life in the cosmos. It's going to be a strange rest of civilization. Let's take a moment to take a deep breath and marvel at this incredible moment in space and time—and that we have this unique opportunity to experience it and shape the future. May we find the courage, the love, and the wisdom to unlock the outcomes that benefit humanity and other life (carbon-based and otherwise) in the Universe.

Mr. William (Bill) Hubick is a technologist who facilitates novel innovation and tailored solutions for DOD customers. He holds a B.S. in Applied Communications Technology from Wayland Baptist University and maintains a Project Management Professional (PMP) certification. His background includes diverse roles such as Mandarin Chinese linguist, cybersecurity PM, software engineering, XR and AI discovery, training, and co-founding the non-profit Maryland Biodiversity Project.

RESPONSE TO

Mr. William Hubick's Reflection

By Chaplain (Colonel) Steve Cantell

In Mr. Bill Hubick's reflection, he sketches the AI strategic context of law, policy, ethics, and safety. He bravely gives his readers an example of strategic foresight, reminding me of the writings of futurist and strategist Herman Kahn. Hubick guides readers to see the metaphorical development process of AI in its childhood. He turns on the siren, calmly warning readers to be responsive, taking the growth of AI capabilities seriously. His article crystalized my thinking in my own quest to have a balanced approach to AI.

"Failing to respond to appropriately," Mr. Hubick warns, "could usher in an array of dystopian scenarios." Understanding this first powerful statement requires thinking about how the Chaplain Corps can respond. One appropriate response is having brave conversations about AI fears—conversations among teammates. We in the Chaplain Corps take care of Soldiers and Families. We also discuss moving forward responsibly, with principle-driven situational awareness of safety and security for Soldiers and Families, with confidential communications in cyberspace. Bringing fears into the light is an enduring task that may reinforce calibrated trust as we in the Corps discuss with Soldiers and leaders religiously diverse ethical and spiritual underpinnings, helping teammates think through the dangers of AI.

Mr. Hubick weaves together a thread of insights about the human aspects of AI development. He advises: "A good way to think of AI systems is that we "'grow'" them, not code them." This raised, for me, thoughts about how people query their home smart speakers. Growing an AI speaks to how one treats AI voice generative systems. He summarizes insights, from entrepreneur and author, Mo Gawdat, who suggests treating an AI with kindness. I even connect Hubick's comment about a commercial AI race "to establish intimacy" to the AI features that mimic emotion-filled responses, with words. My take is that this is a technological golden rule of sorts, reminding readers of common-sense things like treating a personified car kindly with driver education, preventive maintenance, hoping for many miles of safe and reliable trips.

Finally, it is timely for Mr. Hubick to talk about careers in AI for ethics and safety. Like Mr. Heard's thinking, Hubick tugs at the question: "how [will] we factor ethics into such disruptive change"? Thinking about AI as a working person's digital assistant, he warns that "the [AI] systems will manipulate us easily" and "know everything about us." An example of a practical way to factor ethics into disruptive change has been explored, in a white paper by Chaplain (Major) Benjamin Reed, who suggests a future with Chaplain AI Ethicists. I am expecting a future in which Army Chaplains, while advising Commanders, will need to have the knowledge and skills to capably advise Commanders on the human domain impacts of AI and emerging ethical dilemmas. Chaplains will minister to future Soldiers who suffer from exposure to unique, AI-induced forms of critical incident stress, moral injury, and PTSD.

Hubick projects that AI will transform civilization at disruptive speeds. In his thoughtful closing, I would add to his call for courage, that Chaplain Corps teammates guard and protect one another in this AI transformation of civilization. Chaplains and Religious Affairs Specialists will be at the forefront of taking care of Soldiers and Families. AI will be intertwined with the cyber domain in the multi-domains of land, sea, air, cyber, and space.

RESPONSE TO

Mr. William Hubick's Reflection

By Mr. Chuck Heard

Mr. Bill Hubick is a supremely competent technologist who brings a wealth of knowledge to the topic of artificial intelligence (AI) and its application. His reflection addresses the transformative and disruptive potential of artificial intelligence applications on society with strong overtones of caution against the potential for catastrophic outcomes. He also discussed the current state of AI applications as response engines instead of true intelligent entities but segued smoothly back into a discussion of the unrealized potential of AI once it reaches a sufficiently advanced stage to demonstrate intelligent behaviors. The structure of Mr. Hubick's reflection is well designed and demonstrates a well thought out approach to a larger consideration of the use and potential impacts of AI.

Mr. Hubick's insights into the development, or growth, of AI applications is insightful and accurate. He rightfully points out the nascent state of AI and the fact that this does limit its capabilities to little more than providing responses to prompts without any actual intelligence or intent behind the response. Most AI applications are trained using limited pools of data and this restricts their ability to respond beyond the data that it ingested.

He then transitions into a borderline dystopic line of inquiry about the potential for AI applications to eclipse human intelligence by an order of magnitude and the potential impacts of such a development. This discussion is balanced with a consideration of the potential for benevolence in AI to solve complex problems and alleviate human suffering by solving challenges beyond current human potential. I am particularly appreciative of his brief mention of Mo Gawdat's perspective about developing AI applications as children with the expectation they will positively contribute to society when mature.

It is difficult to argue with his conclusion that developing (or even developing with) AI is going to be exceptionally impactful on the future of society as we know it. I am not certain I completely concur with the extrasolar potential he foresees but the impact will almost be global. I am most appreciative of his geek (it is a compliment) sense of wonderment at the direction and potential this disruptive technology presents.

My own inherent optimism on this topic is moderately tempered by the cautionary tone of much of Mr. Hubick's insights. I still tend toward cautious optimism but the pragmatist in me appreciates the fact that disruptive change has equal potential to cause chaos and negatively impact large populations. Mr. Hubick reinforces and validates my own understanding of the relative simplicity (versus its potential) of modern AI and the nature of it as primarily a response engine. His reflection reminds me that almost every complex topic has complex perspectives and often comes with risk, but this also does not mean there is not room for optimism or hope.