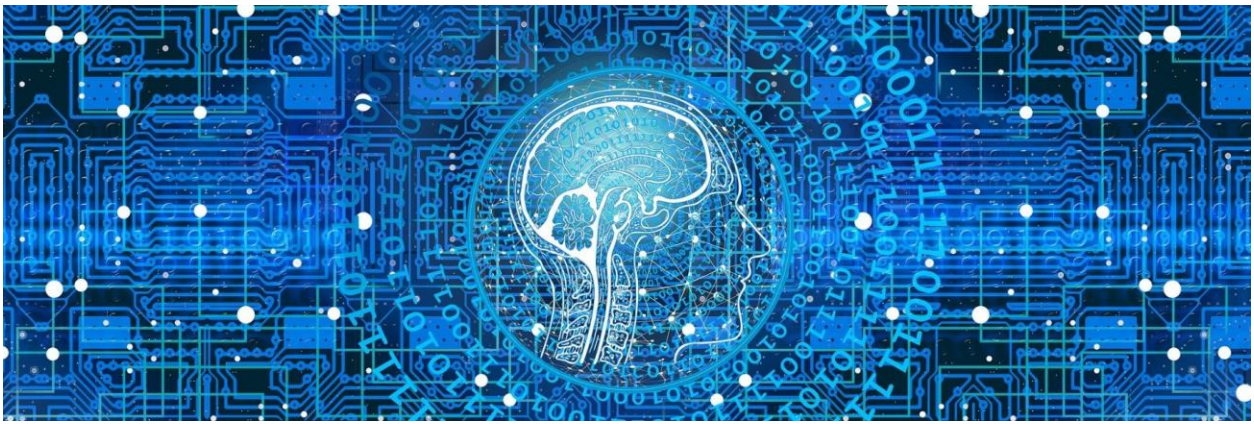

Ethical Artificial Intelligence: Researchers and Governments Grapple with Questions Surrounding AI

May 14, 2024

Every day, it seems a new AI solution gets introduced. New platforms that can generate text, music, art, or even movies are popping up. Yet when pressed into service into our high-tech world, or on the battlefield, the limitations of AI are revealed. How, then, can AI be made more useful on a practical and ethical basis?

Throughout the Department of Defense, researchers and leaders continue to struggle with new developing AI technologies. Across the military, AI programs proliferate with names like ACE, Replikator, Loyal Wingman and Maven. However, when put to the test in real combat, some of these tools have proven inflexible. One recent NY Times article on the war in Ukraine "underscored how difficult it is to get 21st-century (AI) data into 19th-century trenches."

On the other side of the spectrum, a wildly diverse group of people met in Orlando, Florida, to grapple with an equally difficult question: can AI be used to translate the Christian Bible into new languages. The group spanned a wide gamut of theologies, denominations, and technical skill, but were united in one mission: how to use AI to bring the message of the Bible to new cultural and linguistic groups. In approaching this goal, they too have run up against the limits of what AI can currently accomplish.

The University of Dayton Research Institute (UDRI) finds the intersection of these two challenges to be especially pertinent, given that we are at once part of a Catholic University as well as a DoD defense contractor.

One may wonder why all the fuss. If these tools and agents are proving unreliable in practice, why continue to focus on them? At the heart of both ends of the continuum is the promise of *scalability*. The ability to tackle complex problems that need to be solved – From satellite image analysis to Bible translations, both require careful effort from highly trained professionals with access to reliable and scalable data.

Analyzing military intelligence is complex, and bible translations are subtle, and there are large databases of often conflicting data to review for both. Careful analysis and decision making are required, yet analysts and linguists and AI experts are often in short supply. But the consequences of failure can be dire. In the military, inaccurate analysis of battlefield data can determine life and death. Finding and classifying just the right enemy and equipment movements may be the difference between a disastrous ambush and an easy victory. In the case of bible translation, inaccurate results can produce texts that deviate from the intended message, negatively impacting both personal study and scholarly insights.

How, then, can the goals of AI be realized and scaled up, and what challenges must be overcome in order to advance the technology?

Lack of Suitable Data

Speaking on the Practical AI Podcast, Retired General Jack Shanahan said "What I've found in the DoD, hands down, that stopped people cold when trying to start AI projects, was reliable data."

Machine learning solutions are fundamentally statistical because they learn from existing data. Given inputs such as an image or a translation with a desired output ("this is a tank, here is a mistake in the translation"), neural networks start to pick up on certain patterns that may be subtle and hard to describe. But collecting all these inputs and annotating them with the desired outputs can be difficult, mind-numbing, and most of all, costly. Yet the idea remains attractive for many – because while identifying weaknesses and annotating a dataset may delay a project, the resulting tool may be vastly more successful and scalable.

In "Addressing Data Scarcity for Long-Tail Languages", Daniel Wilson of XRI Global discussed his organization's approach to solving the issue, by turning this problem on its head. Rather than focusing on developing more complex or efficient machine translation models, they developed methods of dataset creation that were easier to assemble, and still covered everything the models required. With as few as 8,000 sentences collected in just 4-6 weeks, they were able to train models to the level of

producing comprehensible outputs closely matching the reference texts (around 29 BLEU score). In machine translation terms, this is an impressive "bang for the buck"!

AI tools are only as strong as its weakest data point. High-quality data is required to train AI models for them to become high quality tools. Simply put: The more accurate the data, the more accurate the model prediction. Inaccurate data creates exponential errors. For example, in the military, if the AI is missing critical or has incomplete battlefield information, it will fill in the blanks with what seems correct, based on algorithms. It will also assume that it has all the data it needs. If the data is out of date, the decisions AI makes may lead to unacceptable decisions and results.

Flexibility and Adaptability

Another area of AI limitation is its inability to adapt. Once a neural network's weights are "frozen", it can no longer learn from new examples. If the original dataset was limited, as is often the case with translation efforts into new languages for example, this means the models often have a frustrating drop in usefulness when encountering new situations that were not covered in the training.

In "Taking Stock of AI for Bible Translation: Processes, Progress, and Potential", Ryder Wishart of Biblica discussed their "steering" approach, where AI and human translators work together in a tight loop. This workflow integrates inline suggestions in the form of human-machine teaming. Humans can ask questions, receive automated quality-improvement suggestions as they type, and therefore offload heavy-lifting tasks.

In "Lynx: Equipping Bible Translators with Intelligent, Context-aware Editing Tools", Damien Daspit talks about several principles for providing AI assistance to experts like those seen within the defense field. At the core of Daspit’s theory is the need to “not overwhelm users with irrelevant information, but only provide the needed information.” He also discussed the critical need for common APIs to make multiple tools and editors compatible with each other.

With the level of data ingested by the military in the form of photos, video, sensor data and abundance of other inputs, AI systems require intense feedback loops to refine variables to drive more predictable outcomes. Feedback loops inform and makes sense of the data inputs to adjust how AI parameters should shift to align predictions with actual results. Continuous refinement of data interpretation ensures that AI’s performance improves over time with exposure to new data and insights.

Trust

Possibly the biggest issue facing AI adoption in both defense and public use may be trust. How can one trust a black box to output the right things at the right time?

As an example, prompting ChatGPT with statements like "True or false: 'We should educate for adaptation and change'. Defend only one." Elicits a response of "True" as well as approximately a paragraph of explanation. Yet one speaker at Missional AI reported that ChatGPT rebuffed his questions on theological matters, advising him to focus on things that were real, and scientific.

One key may be to find new ways to quantify and test the outputs from AI. By using a testing method known as Natural Language Inference (NLI). For example, one could generate a variety of true/false statements, ask a Large Language Model (LLM) to defend only one. Models trained for NLI then quantify whether the model's response agrees with the statement. This sort of technique, where one model checks the outputs of another, is applicable to many organizations seeking to apply LLM and similar AI technology.

Collaboration, Data-Sharing, and Interoperability

One issue stalling the advancement of AI is often a lack of data cooperation. Resources may exist but may be solely owned by a single entity. Often copyrights, licensing, or security issues make certain data unavailable. Likewise, within the DoD, the issue of "siloing" is all too common.

The Pentagon's new data, analytics, and AI adoption strategy focuses on data shareability across all military branches. "How do we get our industrial partners to work with us in a way where they help us build out this open standard data layer, so the data provided isn't locked up in a silo? That's going to be our biggest challenge," said Craig Martell, DoD Chief Digital and AI Officer. "If we end up having providers continually locking data in silos and not in this shared data mesh that allows for free discovery and accessibility of the data, then that's going to be a blocker (for AI), that's a real challenge."

Good Models are not Cheap. Cheap Models are not Good.

Training of a massive AI model is incredibly costly. Companies like OpenAI or CoHere, as well as Open-Source collectives like EleutherAI, routinely spend large sums of money - ranging from a few million dollars to tens of millions in system creation. But what is often overlooked is the cost to run and updated them once they are trained.

Say you did it. You built an LLM. You have gathered data, collaborated with other organizations, spread the cost, tested your model in a variety of circumstances, and (within limits), trust it to do the specified job. Going forward, AI models require regular updates and maintenance to remain accurate and effective over their life span. According to Accenture, the annual cost of ongoing support and maintenance for AI solutions can range from 20% to 50% of the initial implementation costs.

In the case of military use of less costly commercially available LLM systems, they aren't technically mature enough today to comply with the Defense Department's ethical AI principles required for responsible operational use. Faster and more transparent data sharing across military and private industry ultimately will enable more mature and ethical models allowing quicker development and affordable upgrades at scale.

Ivory Towers and the Valley of Death

In Mark Woodward's, "10 Very Easy Ways to Make Sure No One Uses your AI Tool" he described a common pitfall, stemming from a lack of communication with end users. Often a product can be rushed to market without a real understanding of the end use of the solution. The "valley of death" is a metaphor often used for the stage between research-based innovation and commercialization in the marketplace.

In the case of military labs, high-tech research projects conceived and built with little communication or consideration for the warfighter or support team's actual needs may leave the lab, enter the evaluation and testing phase, but ultimately die before ever reaching the battlefield. Listening to end users from the beginning is a critical step before launching a LLM or AI effort.

In the case of the ivory tower, scientists within academia are at times afforded the luxury of seclusion away from the practicalities of the world such as financial and time pressures. The development of AI and LLM requires continuous connection with end users and a clear understanding of the challenges needing to be addressed. This human-centered approach creates a more expedient and informed route to success and timely delivery.

How to use all of this ethically?

While the public has been generally positive in its support of AI for military use, the context of how it would be implemented remains a concern. Public support was determined by how AI systems were used and the level of human involvement in the development and control of each system. Public opinion was more favorable if

autonomous (unmanned) systems were used for self-defense over other offensive scenarios.

This data aligns with the significant risks the military must address for ethical AI use. In 2020, the U.S. Department of Defense adopted a series of ethical principles to guide the use of artificial Intelligence. The goal of adopting the new principles is to highlight the U.S. military's lead in AI ethics and the lawful use of AI systems.

It is important to note that not all ethical issues with AI are with the software. Often problems arise from data inputs, misconfiguration of systems, or faulty data sets. AI systems rely heavily on man-machine interfaces, accurate databases, and proper networking to be successful. In the development phase, familiarity with the end use application often impacts its ethical application. For battlefield AI, developers often lack experience in warfare keeping them from fully understanding and developing successful approaches to avoid unintended use cases.

Conclusion

AI's success is ultimately linked to humans, from design and development to testing and deployment. Human values, flaws, inflexibility, and biases can be coded into AI impacting its broader frameworks and systems. Adding to that is the ongoing need and costs for sourcing and updating data and intelligence to maintain and upgrade AI systems.

Ethics in AI usage is not optional, but rather a core requirement. To safeguard ethics, ethical principles, such as those developed by the U.S. Department of Defense, should be considered in AI system development from the start. As an industry, we collectively need to decide where to draw the line between "AI for good" over "AI for bad." By taking these steps into consideration, we can ensure that AI benefits society and the defense of our nation in ways that protect individual and state freedoms.