IDA Ideas

A podcast by the Institute for Defense Analyses

Episode 20

Applications of Responsible AI in the DOD

Guest: David Tate

Host: Rhett Moeller

December 2024

IDA Document 3004076

Institute for Defense Analyses
730 East Glebe Road
Alexandria, VA 22305

The Institute for Defense Analyses is a nonprofit corporation that operates three Federally Funded Research and Development Centers. Its mission is to answer the most challenging U.S. security and science policy questions with objective analysis, leveraging extraordinary scientific, technical, and analytic expertise.

About This Publication

The views, opinions and findings should not be construed as representing the official positions of the Department of Defense or the U.S. Government.

For More Information

David M. Tate, Cost Analysis and Research Division
dtate@ida.org, (703) 575-1409

Copyright Notice

Rigorous Analysis | Trusted Expertise | Service to the Nation

Applications of Responsible AI in the DOD

In part two of a series about artificial intelligence (AI), IDA Ideas Host Rhett Moeller spoke with guest David Tate about the technical expertise, the training, the application and ethics of responsible AI, as applied to the Department of Defense (DOD). David is a Research Staff Member with IDA's Cost Analysis and Research Division (CARD). He has worked with IDA for almost 25 years, and has seven years of AI-related work experience. His background is in operations research, computer science and philosophy. Few people can successfully define AI, but most know it when they experience it. From digital platforms for customer service, to viral animation videos and tools for the military, artificial intelligence is ubiquitous. However, with the understanding that those who use these tools can apply them incorrectly, and that hiring talented individuals who are adept at these tools is a challenge, introducing AI enabled systems to the DOD requires a higher level of responsibility and ethics.

[Begin transcript]

Rhett Moeller: Hello listeners, I'm Rhett Moeller and I'm the host of IDA Ideas, a podcast hosted by the Institute for Defense Analyses. You can find out more about us at www. IDA.org. Welcome to another episode of IDA Ideas.

You can't look anywhere today without seeing artificial intelligence or AI. It's in everything from online tools capable of generating text, video, and images; to drive-thru kiosks and even to toothbrushes. The technology isn't without its concerns, however, and many experts have shared cautions of what might arise if it isn't used with appropriate care. In today's episode, we're going to look at the responsible use of artificial intelligence. In our time, we'll talk about what we mean by this, why it's an important consideration, and particularly how to ensure it's handled with care in the U.S. Department of Defense.

This is the second episode in an ongoing series about artificial intelligence. Earlier this year, you'll remember we spoke with IDA researcher Arun Maiya about the basics of AI. Today, I have the pleasure of speaking with David Tate, a Research Staff Member in IDA's Cost Analysis and Research Division, or CARD. David, welcome to IDA Ideas. Can you please take a moment to introduce yourself?

David Tate: Thanks Rhett. I am a longtime IDA researcher. I've been here almost 25 years now. Before I came to IDA, I did telecommunications network optimization during the dot-com boom, and before that, I taught industrial engineering for a few years. My background

is operations research, computer science, and philosophy. I didn't think the philosophy would be useful at IDA, but suddenly all those courses in ethics are starting to pay off.

Rhett: Yeah.

David: I've been working on AI-related things for probably seven or eight years now, coming at it from an odd direction. We were asked by the Air Force to think about autonomous systems and licensure. How would you license an autonomous system to perform certain missions in certain areas? Thinking like a driver's license.

Rhett: Right.

David: How would you test it? How would you become confident that that it's able to do these things?

At the time, most autonomy was implemented in sort of traditional software, and so this was a question about software assurance and debugging, and so on. Over time, our notions of how to make things autonomous have shifted more and more toward the use of AI techniques that are not like traditional software, machine learning like Arun was talking about.

And that leads to all sorts of interesting questions then about how do you become confident in their performance, how do you assure that they will be safe and secure and reliable and ethical, and all the other things that we now have policies saying our systems should be.

Rhett: That sounds very important, and we're going to be talking all about that over the next few minutes, so let's talk AI. In our previous episode, we covered a lot about the basics of AI and we covered some important terms and concepts. That was a while ago, earlier this year, so could you refresh us briefly at a high level about the distinctions between core concepts, especially with regard to artificial intelligence and machine learning.

David: So, I have strongly resisted defining AI in all of my papers. I usually start by saying I am not going to define it. [I just mean] things that do things that we would normally have thought people would do; or animals … mimicking nature a little bit. I don't, I don't really care whether something is AI or not.

Machine learning, though, that's a very technical, specific thing. Machine learning is a set of techniques for creating algorithms or systems that learn patterns from data or from repeated behavior and use that learning to then react to future inputs, right?

And so, it's sort of like instead of writing an algorithm down that says, okay, here are the steps … do this … divide that by three and then add seven. It says, here's a bunch of examples of what the input looks like and what the output looks like. You figure out how to get from one to the other consistently.

It's extremely powerful for processing very large complicated inputs like images or video, but it's also … a black box. You don't know exactly why the output you're getting is the

output you're getting unless you do a lot of careful … peeking under the hood and probing at it to figure out what's going on.

Rhett: Right. And … in my own reading and thinking about it, even if you get the right answer, you can't necessarily have confidence that the way it got to that right answer is appropriate.

David: Right, it's always a challenge in testing. It's not enough that the system is doing the right thing. You want it to be doing the right thing for the right reasons. And with traditional physical systems, that's a lot easier to verify than it is for some of these machine learning models. And especially recently these generative AI models where we're using the AI model for purposes that are not really the thing it was trained to do. When you think about the classic example … ChatGPT: it's basically trained to predict what the next word will be in a body of text. Now, the fact that you can then use that to write code or to provide … help desk functions, or … write a sonnet in the style of Shakespeare — that's crazy.

And how do you test then whether this new use we want to use it for that nobody thought of before is dependable? That's … a challenge.

Rhett: There's obviously a lot going on in this field, so it is helpful to revisit these concepts from time to time just to make sure we're thinking about things clearly. As I said at the outset of this episode, we're here to talk about the responsible use of AI and specifically how important it is to test AI systems and some of the challenges involved with that. We would like to know more about why this testing is important and what makes it different from, say, traditional software.

David: So traditional software, you think of debugging, right? We, … can run the software and we see if it's doing what we want it to do, and if it's not, you know, trace through it and you find the place where, oh, we're just dividing by zero here, or, well, how did that get to be zero? We can … figure out the why fairly easily.

With the machine learning models, as you are training … [them], you can see that the performance is improving and you can uh adjust to get better … outputs. But once you have the trained model, it kind of does what it does, and there have been some disturbing theorems proved about how there's always the possibility that a slight change in the input will lead to a sudden change in the output.

And they're brittle, I guess is the technical term … it's not a smooth transition as you smoothly change the inputs, you don't get smooth changes in the outputs and you get surprises everywhere. And so, the errors that you're liable to see from a machine learning model … don't behave the way the errors from a normal software system or much less a hardware system would behave.

If you think of an artillery piece, right, you shoot the shells, and you miss the target by a certain amount. There's a nice statistical distribution … of miss distances around the target.

Machine learning models aren't like that. You'll get a bunch of points right near the target, and then you'll get some weird ones that are … how did that happen? And so, there's an extra layer of assurance that needs to be applied either during the development process or at runtime … when you're using it to trap those weird outputs and make sure that they don't lead to unacceptable use of the system in practice.

Rhett: It's interesting that you've mentioned slightly different inputs. Even in my admittedly limited use of say ChatGPT, even using the same prompts, sometimes I get different results, and I don't know if that's a common thing or if that's to be expected.

David: Well, for extra difficulty, ChatGPT is randomized. It is choosing … rolling dice and choosing from the most likely predicted next words … so that it doesn't always pick the same one. And that makes it sound much more natural in its ability to produce English language, but also for testing makes it much harder because the tests aren't reproducible.

Rhett: Right.

David: You can't get the same output always from the same input.

Rhett: So … anybody who's been online for the last year plus has seen flashy videos of ultra-realistic creations, whether it be video images … [or] interesting text that's been generated. And so, we see a lot of commercial focus on the technology and its development. But obviously your focus, IDA's focus, is more on Department of Defense. And so, I'd like to see if you can walk us through some of the differences there and what AI means to both of these industries.

David: I think for defense, people don't realize that all of the big commercial successes using AI have come in areas where the cost of mistakes is very low, right? If Google recommends the wrong link to you with the search tool, nobody cares. If Amazon recommends the wrong product to you to buy, nobody's harmed. If cancer screening software incorrectly tells people that they probably have cancer and causes a lot of expensive testing and stress, that's bad. And we haven't seen yet major commercial successes in AI in these high consequence areas, right?

DOD has lots of high consequence things that they do. Not just with weapons, but also in personnel systems and in lots … of back-office applications. And we know from bad experiences in the public sector, in criminal justice and so forth that there's a real possibility of bias or of discriminatory treatment of different groups because we're training on historical data that captures the historical biases that we have.

And so, the things that [the] DOD most wants to use AI for are inherently higher risk, … higher cost of error than the things that the commercial world knows how to do well. And

we can't really depend on the commercial world solving those problems. If you've been watching the self-driving car industry … the commercial world's been trying that for a long time now with very limited success.

Rhett: You've mentioned some pressing issues and obviously that needs good minds to think about it. What challenges does the DOD face in maybe siphoning some of that talent from commercial and bringing them over to work on DOD projects?

David: Oh, that's a good question. If you want actual uniformed military personnel to have AI skills, we need a new pipeline for that. There are some efforts out there, like the new Defense Civilian Training Corps, which is … a Reserve Officers' Training Corps like program for civilians, especially in tech areas that are trying to do that.

There's a limit to what we're allowed to pay defense contractors in terms of salaries. And so, for the top tech companies, the salaries of the people who are doing their most impressive work are well beyond the limits of what [the] DOD … is allowed to do. And that's a barrier. We tend to need our contractors to be cleared US citizens. And so, the security clearances and the US citizen requirement are … major barriers … in the AI world to getting top talent….

And, in general, the fact that the AI industry is booming means that we're competing against a booming industry for attracting talent, and that's always a problem that's not specific to AI.

Rhett: So, speaking of [the] Department of Defense, obviously this is a large organization and as you've already mentioned, there's a lot of different ways that artificial intelligence and machine learning could be used to augment operations, administration, that sort of thing, but we're looking at a huge scale, not only in the in the number of tools available or potentially available, but also in the size of DOD alone. You add that complication of size and that's got to be incredibly challenging to manage.

David: Well, there's certainly a workforce issue. The DOD doesn't feel like they have the skilled personnel that they need to take advantage of AI and all of the application areas they would like to, and there's a backlog of developing training materials and guidance materials for the DOD workforce. IDA has actually been active in helping to produce guidebooks and policy documents. … And we're actually writing courseware for use at Defense Acquisition University (DAU) to help train the test and evaluation workforce in how to deal with AI and how to think about the particular challenges.

Rhett: I see.

David: The acquisition workforce has been retooling their certification processes for the last few years. And I think this is partly driven by changes of COVID with more … remote work and … less in-person education. But also, I think there was for a long time, a feeling that certifications in the acquisition workforce were something that you got in your first

few years in the field, and then 10 years later, somebody wanted you to use those skills and, and you'd forgotten everything you knew.

Rhett: Right.

David: And so, they're trying to shift to a more, let's say, agile system where people get focused credentials in specific topic areas that expire after three or five years and have to be refreshed. … And so, the plan is that in the future you will put people on projects based on which certifications they have, which credentials they have. And that people will sort of pick and choose what they want their career to look like, by which credentials they get. But there's kind of a chicken and egg problem there. You can't really do that until the credentials exist for people to get, and you can't start requiring them for specific jobs until people have the credentials.

And so, there's a big backlog of creating the training materials for these credentials and the testing materials for these credentials, to be able to support this new concept of … workforce certification. And when we saw that DAU was fully booked in trying to develop courses and that they were falling farther and farther behind, we said, hey, you know, we have some expertise in some of these subject matters, and we know how to develop courseware. How about Department of Defense pays IDA to develop some of these credentials, … and we can take some of the load off of DAU. And I'm the guinea pig.

We are trying to do that now for the first time. We're developing an eight-course credential and test and evaluation of AI. The first course should go live in a week or two.

Rhett: Great.

David: And I hope to have them all out there by sometime in the spring.

Rhett: You mentioned, David, a very thorny problem of needing qualified people in positions, but then needing the training to make sure that people are qualified. … I can see how challenging that is to get such a program started.

David: Traditionally, it's much harder to design and build things than it is to test them. And all of the hard … engineering challenges were in making it work. We're starting to think that for machine learning in particular, it's going to be harder to test them and be confident that they're dependable, than it was to build them in the first place.

Rhett: Right.

David: And so, things will be showing up as we want to use this faster than we can become confident that we should use them or that it's not overly dangerous to use them. And that's a new situation, and DOD has always hated testing because it's expensive and slows down your program. This will make that even worse. … There's a real risk that people will either not get to use AI for the things that it could be doing, or that we will deploy things before they've really been thoroughly tested and verified and something bad will happen.

Rhett: Obviously there's a lot to consider in this topic, and I can tell it's exciting work with long term ramifications. Is there anything that you want to revisit or elaborate on?

David: I think the whole concept of responsible AI has changed the way the policy world thinks about test and evaluation. I don't think anyone would argue that we only want our AI enabled systems to be responsible. We want all of our systems to be responsible. We want them to be employed responsibly. That's always been true. But AI, and machine learning in particular, bring these issues to a certain head … and introduce certain new risks that cause us to need a whole framework for thinking about how are we going to systematically make sure that our systems and their employment … are responsible and ethical.

And I think that's going to be good in the long run for test and evaluation. I think we're moving away from the historical stovepiped system, where the safety engineers are over in that corner worrying about safety and the cybersecurity guys are over in a different corner worrying about cybersecurity, and the operational testers are worrying about effectiveness and suitability, and none of that is really coordinated in a systematic way to say, what are all the things we want to be true about this system? What is the most efficient way for us to collect the information and make the arguments to show with reasonable confidence that all of those things are true? So that we can confidently employ these systems to do the things we need to do.

And so, I know that the Chief …Digital and Artificial Intelligence Officer, …CDAO, is promoting a holistic assurance approach for AI enabled systems where you think about all of these issues simultaneously. We are trying to help them develop test and evaluation policy and guidance and training, as I said before, to get people thinking about all of these things at once and how they fit into the development life cycle. And what kind of tools they can use to then assure as we go from the beginning. Rather than trying to come in at the end with test and evaluation and say, Okay … did you do it right?

…Let's instead build it in a way that produces assurance over the course of the development and also produces the information that decision makers need in order to make informed trade-offs between risk and capability.

Rhett: David, thank you very much for taking the time to discuss this timely topic with us and for sharing your expertise. It's really been illuminating.

David: It's been my pleasure.

Rhett: As always, if you want more information on IDA and its ongoing work, please check us out at IDA.org. We also have a presence on X at IDA_org, and we have a channel on YouTube. …IDA Ideas is hosted by the Institute for Defense Analyses, a nonprofit organization based in the Washington DC area. Once more, you can find out more about

us and the work we do at IDA.org. Thank you for tuning in, and we hope you'll join us again next time as we discuss another big idea here at IDA Ideas.

Show Notes

Learn more about the topics discussed in this episode via the links below.

Haga, Rachel A., John W. Dennis. "WEAI 2023 – Assurance of Responsible AI in Personnel Management." IDA Document 1038161. June 2023. ida.org/research-and-publications/publications/all/w/we/weai-2023-assurance-of-responsible-ai-in-personnel-management

Haga, Rachel A., John W. Dennis, Erin P. Eifert, David M. Tate, Connor P. Trask. "DATAWorks 2024: Data Verification Validation and Accreditation for AI Enabled Capabilities." IDA Document 3001831. March 2024. ida.org/research-and-publications/publications/all/d/da/dataworks-2024-data-verification-validation-and-accreditation-for-ai-enabled-capabilities

Maiya, Arun S., and Rhett A. Moeller. "IDA Ideas (Podcast Transcript) – Exploring AI and Machine Learning Capabilities." IDA Document 3002620. June 2024. ida.org/research-and-publications/publications/all/i/id/ida-ideas-exploring-ai-and-machine-learning-capabilities

Shapiro, Daniel G., Joshua Alspector. "A Grounded Introduction to Large Language Model and Generative AI Technology." IDA Document 3002626. September 2024. ida.org/research-and-publications/publications/all/a/ag/a-grounded-introduction-to-large-language-model-and-generative-ai-technology