

# FTCN Replay: Speed Alone Won't Bridge the AI Trust Gap

Artificial intelligence is advancing at breakneck speed in military applications, but the real bottleneck may not be the technology itself – it's whether humans can trust and effectively use these systems. That's according to Jeff Druce, Senior Scientist, Human-Centered AI at [Charles River Analytics](#), who joined host and AOC Director of Advocacy & Outreach for a recent episode of [From the Crows' Nest](#) to discuss the challenges facing AI and machine learning in the battlespace.

“At this point, we're seeing such a rapid explosion in the advancement of many of AI tech,” Druce explained. “The slowdown may not be the tech being ready, but the humans being ready to use the tech in an effective way.”

## The Explainability Challenge

Druce, who has spent nearly a decade working on applied artificial intelligence, focuses on providing explainability and enhanced verification to AI-driven systems. His work addresses a critical gap in military AI: while these systems can generate complex courses of action virtually instantly, users have no window into the reasoning behind those decisions.

“Under the hood, you don't know what the reasoning process was for this course of action that was generated by an AI entity,” Druce said. He emphasized that even though machines can make rapid decisions, “if you want it to be trusted, if you want a human in the loop, there needs to be that mechanism to unpack it.”

The stakes for this transparency are high. Deep reinforcement

learning agents – AI systems that learn to make decisions through trial and error – use neural networks with potentially millions of computations mapping inputs to outputs. Without explainability, users have no idea what factors drove a particular recommendation, making it virtually impossible to trust AI-generated strategies when lives are on the line.

Druce pointed to the famous “alien move” made by Google DeepMind’s AlphaZero during a Go match against a human master. The AI made a decision that no expert understood, yet it proved critical to victory.

“If you can imagine if we have a confrontation and we can leverage something like this model, that can bring in a novel strategy to do something that is incredibly effective of a tactic that is not known,” he explained. But without understanding the reasoning, “it’s kind of hard to trust.”

## **RELAX: Adaptive Explainability for Military AI**

To address this challenge, Druce is working on RELAX –Reinforcement Learning with Adaptive Explainability – a program designed to crack open the black box of AI decision-making. The system provides explanations tailored to different users, recognizing that a generalist and an AI specialist need very different types of information.

Through user studies, Druce’s team tested whether explanations improve performance and trust. They created scenarios where AI agents performed well in some situations and poorly in others, then measured whether users with access to explanations could better understand when to rely on the AI.

The results were statistically significant across all three hypotheses: explanations helped users develop more accurate mental models of how the AI operates, improved their effectiveness at deploying the agent appropriately, and increased their overall trust in the system.

“We found that we asked a series of questions to basically test this where we knew the correct answers, we knew what the decision drivers were of the agent, and we asked the user their version of the mental model and we graded that for accuracy,” Druce said. The explanations included both graphical and text-based elements, allowing users to choose the format that worked best for them.

## **The Path Forward: Acceptance Over Advancement**

Looking ahead, Druce believes the next five years will be less about technological breakthroughs and more about human factors. The military’s careful approach to change – he pointed to the example of it taking 25 years to change the metal in an aircraft wing – means wholesale adoption of AI for strategic decision-making won’t happen overnight.

“I think the next five years personally is going to be understanding how to use AI effectively and what do we need to extract from the AI in order to understand and trust them,” he said. “People don’t use things they don’t trust. That’s been found with automation throughout history.”

Druce emphasized that appropriate trust is key – AI tools, like any tools, have specific use cases where they excel and others where they fall short. The goal isn’t blind faith in AI recommendations but rather giving users enough understanding to know when to rely on these systems and when human judgment should prevail.

As military operations become increasingly complex and data-rich, the ability to leverage AI’s speed while maintaining human oversight will likely determine competitive advantage on future battlefields. But that advantage, Druce suggests, depends less on building smarter AI than on building AI that humans can understand and appropriately trust.