

Reflections on AI

Q&A with
David Barnes

“We need to ask the question: should we be doing this?”

The [TUM IEAI](#) had the pleasure of speaking with BG (Ret.) David Barnes, PhD. David Barnes advises senior government and business leaders on mitigating AI regulatory and ethical risks affecting their business or investment decisions. Most recently he is serving as Professor, United States Military Academy (PUSMA) Department of English and Philosophy, [West Point](#), NY and as Chief AI Ethics Officer for the US Army's Artificial Intelligence (AI) Integration Center (AI2C).

1. What is the biggest misconception about AI?

One of the things that are problematic in this space, is that there is too much over-promise of what AI enable technologies can do today and for this foreseeable future. To paraphrase Carl Sagan (the astronaut?): “There is a lot of bologna out there”. Part of the problem is this notion of cognitive anthropomorphism; we see AI doing very powerful and unique tasks. We believe it exceeds or meets human level intelligence or ability, but then we end up subconsciously shifting the goal post from it. For example: the game ‘Go’. It is amazing what Alpha Go and Alpha Zero did with that game, but is that really a matter of the AI becoming the I part- the Intelligence piece- or is it really that we’ve discovered that the game ‘Go,’ as complex as it is, is very heavily reliant on computation. Moreover, that the AI is better at those types of tasks. So, I believe that this ties back into one of

my main themes: how do we educate the ‘every-person’ about AI? In terms of the government, military or the business world: you are going to invest billions of dollars in this technology and we want to ensure we are doing it in the right sort of way, but we need to understand what its limitations are.

2. What is the most important question in AI ethics right now?

One question that really stands out in my mind- typically in the tech industry- we ask: can we do this? However, I think what we need to think about in terms of AI: should we be doing this? In other words, is AI the right set of tools to solve a particular problem space? It is really about aligning values. We see a problem area and we want to leverage this exciting technology. We need to make sure this aligns with our values. Bryan Christian calls this the “alignment problem”. So even starting from that initial question, when we are looking at researching the problem to address and potential solution, [accordingly planning needs to be done] very early in the design process. However, we need to ask ourselves should we be using this technology. Moreover, we need to ensure, in the broad sort of sense, we include those legal and ethical considerations. Also, from a practical standpoint: is this the right investment moving forward? Will it actually solve the problem we are attempting to solve? Therefore, I believe it is really about aligning our values.

3. What are some of the major ways militaries around the world using AI?

As I mentioned in the main body of the talk, the military is looking at AI and how it can best increase efficiencies throughout all of the enterprise processes, not just on the battlefield. Some of these [are focused] on the battlefield, which militaries around the world are seeking to explore, including intelligence collection and the fusion of this vast body of data. Additionally, there is cyber defense: because of the concerns over cyber [attacks] from the last years (we have seen examples of different sort of cybercrimes and cyber strikes against different companies as well as national institutions). [The question is:] Can we leverage AI to help protect those critical and vulnerable systems? [Or with] predictive maintenance, we can save potentially hundreds of millions of dollars by more accurately predicting when particular systems will fail in their life cycle and be more aligned, therefore spending our resources in a better and timely manner, ensuring that they are kept at a high level of maintenance. I also discussed a bit about fires and the concept on the battlefield of the sensor to shooter clearance of fires- work is being done in different armies to identify a way that we can reduce that chain. So we can get accurate and timely [control] fires, that are discriminate on the right target by the right system. Of course, there is a lot of work being done on crude systems. Whether it is UAVs, air systems, ground systems - even work in the oceans or underwater. Much of that work is focused on fully autonomous systems, yet AI can help improve those autonomous processes. So, those are the main areas of the battlefield. Of course, we have to remember we are looking at how we can better manage talent. How can we

think of running a bureaucracy- is there a way to leverage AI to help streamline and generate efficiencies there as well. Therefore, all areas are currently being explored.

4. What do militaries need to do to maximize the benefits of AI-enabled systems while mitigating the risks they entail?

It partly goes back to the earlier question: should we use AI to solve this particular problem. Currently, the AI we are working with is limited (or the layered narrow AI system, they are brittle, there are problems of over-fitting), so we may not fully understand the long time unintended consequences. We haven't developed a full ontology of failures. How the system may fail in ways that right now are currently unanticipated. Part of it is, thinking about the law of armed conflict and international humanitarian law. Are we developing systems that are keeping up with the laws that are currently out there? [This applies to] the rule of law that we feel is so important. As well as the AI principles, e.g. the US Dept. of Defense has developed and published an article on AI principles. Can we ensure that the systems we are developing applications, whether it is on the battlefield or back office, are aligned with those principles? It really ties back into this notion of risk management, again risk management is only part of the solution- maximizing the benefits- but if you think about that process of identifying risks- looking at steps to mitigate risks, looking at residual risks and then in a sense having the ability to say: once this risk is mitigated is it acceptable or not. This is going to be very important moving forward.

5. Given the high risk associated with the use of AI in the military, what about the question of accountability?

One of the things the military has (that you may not see in the industry and other institutions) is the notion of command responsibility. That the commander is ultimately responsible for all the actions that the unit does on the battlefield. This is a very important piece, because when we look at employing a system on the battlefield, the buck stops at the commander. However, we also need to be careful when we are talking about the accountability of it. We are not just talking about whom we should blame, should there be an error or an unintended consequence, no matter how dire. In the entire design development of the system, there are many steps along the way, where error may have been introduced. So, in a sense, the commander is ultimately responsible, but the question will still remain to what degree. That is going to be determined by the law, regulations, the particular society and even the international potential condemnation of certain acts. The key to this, of course, goes back to education for every person. In this case to every soldier. What does each individual, from the youngest private to the highest level, need to know about the benefits, risk and limitations of the AI-enabled systems. The better educated they are, the better educated [decisions] they will be able to make in terms of whether they believe they are willing to accept certain risk. In conclusion, that engenders trust in the system. Again, soldiers are not going to use these systems unless they trust them. By that, I mean, not just that it is reliable, but that they are confident they know how it's going to react in the different contexts they are using it in. This means they have been trained in using it, they've

tested it in this case - and if not - they may choose not to use it.

6. Who should be in charge of, or involved in, developing ethical frameworks and standards for AI?

That is really a great question and it really lines up with this concept of many efforts/many hands. One of my personal lines of operations (one that we are doing with the US Army), is how do we increase productive collaboration across industry, government and academia. We know that regulations are coming; there should be a desire to be a part of that. To help shape it. Because part of the concern is that, we want to ensure that we have the right kind of regulation to ensure that it is keeping up with our values in the rule of law. However, we also want to make sure it does not stifle innovation, because, again (although there are certainly areas of high concern and high risk), we want to be able to leverage this technology in areas that can really be productive to society at large. If you remember back in 2020, the Berkman Klein Center identified 36 steps of principles. So a lot of work is being done nationally and internationally in identifying different principles or guidelines for designing, developing and using AI-enabled technology. The question here is: how do we put those principles into practice? In addition, part of that is going to be the development of clause and regulation policies, which are keeping up with the values of the institution of the country they represent, as well the international community.

Disclaimer: *These are my own personal views and do not necessarily reflect those of the US Army, DoD or US government.*
(BG (Ret.) David Barnes, PhD.)

Meet the Expert



David Barnes, PhD.

As Chief AI Ethics Officer, David Barnes advised the Army on incorporating ethics, law, and policy into Army AI design, development, testing, and employment, directly contributing to the development of the DoD AI Ethical Principles and the 2020 US Army AI Strategy. In this capacity, he provided assistance to the Joint AI Center (JAIC), Defense Innovation Board (DIB), National Security Commission on AI, OSD Autonomy Community of Interest (CoI), OSD Biotechnology CoI's Ethical, Legal, and Social Implications (ELSI) Subcommittee, and others. He also served as the US Army's Senior Service Representative for the USSOCOM Commander's 2019 Comprehensive Review of Culture and Ethics.

His research interests include normative and applied ethics, especially the ethics of war and the ethics of emerging technology. He is a DARPA Senior AI Ethics Advisor, a Senior Advisor for the JAIC Responsible AI Subcommittee, and a Non-resident Fellow at the Stockholm Centre for the Ethics of War and Peace.