

18 ETHICS AND AI

PJD's opening remarks

In our presentations and questions about AI the past weeks, we've come across a number of ethical issues. They appear as dilemmas for which there are no ready answers. They cannot be resolved by the AI machines. Human designers must grapple with them. I've been keeping notes and already have compiled a list of nine dilemmas that all require ethical judgment.

One example is fragility. Remember, this means that many neural networks are very sensitive to small changes in their input. Modify a stop-sign image slightly by changing a few pixels and the AI says it is a speed-limit sign. It seems that a lot of neural networks are sensitive in this way. If you are asked to put one of these AIs to work in a critical operation, such as driving a vehicle or firing a weapon, would you trust it? How do you know when you can trust the AI enough to deploy it?

Another example is control. Even if we can build an AI weapon system that decides on its own when to fire, do we want to operate that way? Do we want a human operator to make the decision? How do we deal with our fear that the enemy's system is completely automatic and will fire on us before our human operator can decide anything?

Today B J Strawser will talk about these kinds of issues and offer some ways of thinking about them. He is Associate Professor in the Defense Analysis Department. He has written and spoken extensively on ethical issues of all kinds relating to warfighting.