A Technical Forum for Confidence-Building in the Autonomous Weapons Realm

By Malte Göttsche

Prof. Dr. Malte Göttsche, Leader of the Nuclear Verification and Disarmament Group at the Aachen Institute for Advanced Study in Computational Engineering Science (AICES) Graduate School of RWTH Aachen University

Today, we find ourselves in a world of diminished trust among global actors, one characterized by power competition and a qualitative nuclear arms race. Against this background, research and development efforts that contribute to enabling autonomous weapon systems (AWS) are particularly worrisome, as such efforts may aid the initiation of yet another technological arms race. Preventing this requires confidence-building, to which not only the policy, but the scientific community should also contribute.

Autonomous Weapon Systems

No consensus exists about the definition of AWS. A key characteristic is that these systems could autonomously select and engage targets. They “will be able to operate without human control or supervision in dynamic, unstructured, open environments […].”[1] However, it is hard to define a threshold, as the degree of autonomy is a spectrum.

The use of AWS may not be far in the future. Prototypes are being tested in several countries, and several precursors already exist. The current main competitors in this field are the United States, Russia, and China. AWS may offer advantages to the military: fewer soldiers would need to directly engage in combat. In the absence of the human need for rest, endurance during warfare would be enhanced. Reaction times would be reduced if systems did not require a remote soldier to make decisions. Individual communication links that can jam would no longer be required. Weapon swarms would become possible.

However, the risks ultimately outweigh the benefits. AWS would reduce predictability and control on the battlefield. Given the impossibility of training the control program for all possible circumstances in combat, potentially grave mistakes could occur. Other limitations include that artificial intelligence will in the foreseeable future not be able to reliably distinguish between combatants and non-combatants. This inability, along with quick response times, could cause conflicts to almost instantaneously escalate.

Regulating AWS

States may be tempted to invest vast resources to develop AWS, either to be the leaders of the development, or to avoid falling behind. While there could be temporary military advantages, there seem to be no long-term benefits of such an arms race. Instead, it would increase the probability of war, including by erroneous decisions of AWS.

[1] Altmann, Sauer, Survival 59, 2017

These risks should be an incentive to regulate AWS. There are ongoing discussions in the Group of Governmental Experts in the context of the Convention on Certain Conventional Weapons in Geneva. At least thirty states propose to ban their development, deployment, or use. However, those states investing in relevant research object to banning AWS. If consensus is required, as is the case under the current format, a ban is unlikely. Germany seeks a middle ground by proposing to formally declare that all weapon systems must be undergirded by meaningful human control.

Overall, the discussions are highly controversial and, at best, slowly evolving. So what are additional options to seek progress?

Scientific Contributions to the Debate

The history of nuclear and chemical arms control shows the importance of integrating scientists into the discussions. In particular, the Nuclear Non-Proliferation Treaty, most bilateral United States-Russia nuclear arms control agreements, and the Chemical Weapons Convention have strong verification regimes whose development depended crucially on scientific expertise. In the case of the Comprehensive Test Ban Treaty, it was the Group of Scientific Experts that helped pave the way by developing the verification approach of the treaty many years before it was finally negotiated.

Technical work also acts as a confidence-building measure, as the currently active International Partnership for Nuclear Disarmament Verification demonstrates. At a time when nuclear disarmament is a highly divisive issue, this group nevertheless successfully discusses how it could be verified and conducts exercises. The participating countries, who hold diverse views on nuclear disarmament, do so by not spelling out how it could be achieved politically. Indeed, a success factor of the Partnership is that it is more a technical than a political forum. Besides diplomats, technical experts also participate, including academics. This enables them to achieve concrete scientific results.

How could a technical forum of interested parties be an avenue for progress in the AWS context? Here, the debate is much less framed than in nuclear disarmament discussions, which have a legal basis in the Non-Proliferation Treaty. A verification regime as part of an arms control treaty will likely not be the first step in preventing or limiting AWS.

A technical forum could build confidence by preparing the ground for future voluntary transparency initiatives. For example, it could develop technical approach-enabling exercises in which states could demonstrate that during certain tests of weapon systems, no autonomous modes were explored.

How to Assess the Non-Use of AWS?

Since it is unlikely that direct participation in such exercises would be possible for reasons of sensitivity, there is no simple way to establish the non-use of AWS. Also, there are no clear characteristics that could be identified upon observing the actions of a weapon system, for instance via video, to prove that it is acting or has acted autonomously. Even with the ability to fully examine software and hardware – a highly unlikely scenario – it would be hard or even impossible to reach a conclusion: the same hardware could be used with or without autonomous mode, and the authentication of complex software is extremely challenging, sometimes impossible.

Even though it was developed for the arms control verification context, a cryptographic method could provide a way forward through voluntary demonstrations that prove the actions of a weapon system are the result of orders given by humans.[2] According to this concept, human-machine interactions would be recorded in an encrypted database. When asked for proof that a specific weapon system did not act autonomously during a specific event observed on video, the state could make available the particular records.

This is only a preliminary idea; much more work will be required to develop it, and perhaps also different and approaches can be thought of. The focus of the proposed forum should be on technical dialogue. Only when a certain level of trust has been built through this process can actual exercises be discussed.

In conclusion, as consensus in Geneva is far from emerging, other avenues should be explored, and new actors should be engaged in the debate. The scientific community has an important role to play, as it can contribute to building confidence and generate new and innovative ideas.

[2] Gubrud, Altmann, Compliance Measures for an Autonomous Weapons Convention, ICRAC, 2013