Under Discussion

Illustration: © racken.

# "There's a danger that things operating at machine speed can spin out of control."

An Interview with Dr. Frank Sauer, Senior Researcher in Political Science at the *Bundeswehr* (Federal Armed Forces) University Munich

***Ai:** Dr. Sauer, supercomputers that take on a life of their own, robots that rise up against their creators and an earth devastated by killer machines – these scenarios have been the stuff of science fiction for years. In your research, you focus on the nexus between security and technology, such as the military-technological implications of artificial intelligence (AI). Just how far removed are we from science fiction in this area?*

**Frank Sauer:** That depends on the kind of science fiction you mean. For example, if you take a novel like Kill Decision by Daniel Suarez, some of the ideas in it no longer seem so far-fetched. On the other hand, apocalyptic scenarios such as those in the Terminator films are still a long way off, or will never actually happen – or so we hope. I myself enjoy reading science fiction, but I'm not too worried about robot uprisings, terminators and artificial super intelligence. I'm concerned with more mundane things in the here and now.

***Ai:** Such as what?*

**Frank Sauer:** I'm currently looking at the risks of the short-sighted application of technology in security contexts – technology that is currently available and, relatively speaking, quite "dumb". Especially, as you rightly point out, when it comes to the military and the use of these applications in weapons systems.

Despite all reservations: The military should not forego technology. Source: © Charles Platiau, Reuters.

***Ai:** Can you give us an example of the short-sighted application of comparatively "dumb" technology in weapons systems?*

**Frank Sauer:** Take automatic image recognition systems. They are at the forefront of current breakthroughs in the area of artificial intelligence. If you save your photos in Google Cloud, you can ask Google to sort them for you – say, all the photos of your last beach holiday, your new car or Grandma Erna. All well and good. Or take autonomous driving. Tesla is so convinced of the capabilities and potential of automatic image recognition that it is committed to using it to produce a self-driving car at some point. Tesla simply does away with other components that most other car manufacturers consider indispensable, such as lasers for measuring distance. And it's true that automatic image recognition is amazing. But it has nothing to do with intelligence. Unfortunately, the terms "artificial intelligence" and "machine learning" are very misleading for the majority of people. The neural networks trained for image recognition based on machine learning are developed for a single, extremely limited purpose. They are competent, but not intelligent. They can recognise cats in photos – in some cases more reliably than a human. But that's it. And they can only do this under certain conditions; if they are faced with inputs that they haven't been optimised for, they fail spectacularly. Therefore, we are not dealing either with intelligence or learning, at least not in the way that we humans have previously understood these terms, and how they can make sense for a species such as ours, which is so much more capable and adaptable. This is what I mean when I say that a modern image recognition system is, relatively speaking, "dumb", even if it performs extremely well in certain applications.

***Ai:** What does this mean for the use of such technologies in weapons systems?*

**Frank Sauer:** It would be dangerous to rush into recklessly using automatic image recognition technology in weapons systems. This doesn't require much imagination, as we have had examples of this for some time now. Last summer, for example, Kalashnikov came out with an autonomous gun turret that combined an image recognition system with a weapon – this is all current technology, not science fiction. But, of course, the Kalashnikov image recognition system cannot understand a battlefield like a human can. This turret would probably have difficulty distinguishing soldiers from civilians. And with a probability bordering on certainty, it wouldn't be able to recognise and understand whether a soldier is trying to surrender, or is perhaps injured and therefore no longer constitutes a legitimate target. Which brings us to the risks. If the weapon were fired automatically, because the algorithm can only identify basic patterns and not interpret what is actually happening in a given situation, then this would be a violation of international law governing the conduct of war. And on top of that, it would be difficult to determine who should be held criminally responsible for such a violation. But this doesn't mean that the military should forego technology. What it means is that we first have to think very carefully about when and how much decision-making power can be delegated from human to machine. This will vary according to context and – also dependent on this context – the military will need a few new rules. This is no big deal, after all the military are very good at drawing up and obeying rules. But unfortunately, the many misunderstandings surrounding the terms "artificial intelligence" and "machine learning" and all the hype about "AI in the armed forces" in general is

currently making it difficult to implement this kind of common-sense approach. Drawing up new rules for dealing with autonomy in the weapons systems of today is a lot of work, and not as sexy as the continued dreaming of tomorrow.

***Ai:*** *But one could argue that, particularly in the military sector, it's not so much dreaming of tomorrow as focussing on very real security issues. Or would you say that there's no justification for worrying that we might end up lagging behind China, for example, the longer we continue to dwell on the risks of new technology? It's painful enough to lag behind in economic terms, but when it comes to the military this can quickly take on an existential dimension.*

**Frank Sauer:** It's interesting that you specifically mention China. China is well aware of the security risks associated with an unregulated, offensive use of weapons systems that "autonomously" select and engage targets, i.e. without effective or meaningful human

Looming threat? When humans are totally removed from the decision cycle, the humanitarian risks rise significantly. Source: © Ognen Teofilovsk, Reuters.

control. One of the main effects of having a completely automated decision cycle would be the enormous acceleration of operations. The Chinese have coined the eerily beautiful term "battle-field singularity" to describe the point at which human cognition can no longer keep pace with developments on the battlefield. Everyone – and above all the countries at the forefront of technology – is well aware that this entails considerable risks of escalation.

***Ai:** Comparisons are often drawn between a hand-wringing West, which allows itself to be held back by ethical and regulatory issues, and China, which forges ahead without hesitation. Do you think that's fair?*

**Frank Sauer:** Don't get me wrong, there's definitely an element of truth in that. We only have to look at the latest developments in China with regard to human germline engineering, which clearly breaches existing taboos. And I still have my doubts about the willingness expressed by China at the United Nations in Geneva to sign up to an international treaty banning the use of fully autonomous weapons systems. China loves to create this kind of diplomatic smokescreen. The point I was trying to make was that, despite this, there is a general awareness of the risks on all sides. Not only in China, but also in the US. The former US Deputy Secretary of Defence Bob Work, for instance, who was responsible under Obama for promoting the issue of AI and robotics in the US armed forces, made it abundantly clear that the US was not willing to be the first to cross the Rubicon, but that it had to be prepared to be the second across in an emergency. So risk awareness is one thing, but internationally binding political agreements are another. This brings us back to the dilemma addressed by your question – the classic security dilemma in the international system, including all the associated incentives offered by unregulated arms. To put it in a nutshell: "Since I can't be sure my opponent won't build killer robots, I'd better build them myself." But in addition to this individual risk, there are collective risks, which are now well understood. Just think of the implications for international security and stability. When humans are totally removed from the decision cycle, there's a danger that things operating at machine speed could spin furiously out of control and escalate unintentionally. There are also significant humanitarian risks, such as civilian suffering, not to mention the key ethical question of whether we want future wars to involve this kind of "automated" killing, thereby uncoupling it from our judgments, decisions and consciences. The German government uses this risk of crossing an ethical red line to justify its negative attitude towards delegating kill decisions to machines in wartime – an attitude that former Defence Minister Ursula von der Leyen and the German Ambassador to the United Nations in Geneva, Peter Beerwerth, recently publicly reiterated. Recognising these risks should not be dismissed as simply hand-wringing on the part of the West. On the contrary – who else is supposed to stand up for the values and standards affected by these developments on the international stage? It's not likely to be China.

***Ai:** In the end, then, it comes down to a classic risk assessment: how highly do I rate the risk of the unregulated use of autonomous weapons systems as compared to the risk of lagging behind on military technology, perhaps because I misjudged the intentions of my counterpart? Is that right?*

**Frank Sauer:** Yes, that's right.

***Ai:** Given this kind of risk assessment, do you believe it's realistic to expect the stakeholders involved to come to some kind of agreement on effective arms control in this area?*

**Frank Sauer:** In principle, it's possible. That's how we ended up with agreements between the superpowers on things like nuclear arms control. If the collective risks are understood and taken seriously, then it should be possible to steer particular developments on arms control in other fields too, and in this way limit a potential arms race. That's obvious, as otherwise we wouldn't have any form of arms control at all, neither for nuclear, chemical or biological weapons, nor for anti-personnel mines, cluster munitions or blinding lasers. But we do have these controls, so I think it's too early to throw in the towel in this case. As a community of states, we can still insure ourselves against these collective risks, which are far greater than the risks posed to the individual state.

This would above all benefit the countries that are at the forefront of technology, as the kind of technology that is used for autonomy in weapons systems has largely been borrowed from the civilian sector, and so it diffuses much more quickly than the sophisticated military technology of the past. This means there will not be a monopoly on autonomy in weapons systems, such as that enjoyed for a while by the US with its stealth technology. Nevertheless, we are currently in a political phase in which enthusiasm for international arms control is on the decline rather than on an upswing. Existing treaties and agreements are being eroded, and urgently needed new ones are not being negotiated. At the UN in Geneva, talks on autonomy in weapons systems are progressing slowly, to put it mildly. This is why – although I believe arms control is both possible and necessary – I think we can't realistically expect to see any great progress in the near future. We will probably have to put the "arms control winter" of the Trump-Putin-Xi-era behind us first.

***Ai:** To what extent is it possible to control these new technologies? You say these technologies are spreading much faster than in the past, so what are the possibilities for effectively preventing this spread, or for identifying potential violations and then imposing sanctions where necessary?*

**Frank Sauer:** It cannot and should not be about controlling technology. Especially as most of the progress being made in technology is in the civilian sector, where we hope to take every imaginable advantage of the developments being made. We shouldn't try to stop progress and anyway we probably can't. But we need rules for dealing with this kind of technology. Our best chance of developing such rules is to stop talking about technology and instead to take a differentiated look at humans and their potential future role in warfare. How should we design meaningful human control over weapons systems, and when should it be used? Do we need to intercept projectiles approaching at lightning speed? If so, then humans can confidently be taken out of the decision cycle and the task delegated to a defensive machine. If, on the other hand, it's a matter of planning and deliberately carrying out an attack that may cost human lives, then humans should continue to decide on the selection and engagement of targets, take legal responsibility for the decision and bear it on their consciences. So we are basically talking about the regulation of military practices and the context-specific adjustment of the man-machine relationship in the military.

***Ai:** This sounds like an enormous challenge in itself – not to mention the question of how to effectively verify compliance with the rules once they have finally been agreed.*

**Frank Sauer:** Of course this is no easy task; and of course we know that rules are broken, including in the area of arms control. Not constantly and everywhere, but now and then, in specific cases. But that's not a reason to have no rules at all. It's only on this basis that sanctions can be legitimately applied. It is indeed difficult to verify the retention of meaningful human control over weapon systems as a general rule, with the exception of defending against incoming munitions. This is a much greater challenge than monitoring compliance with arms control treaties in other areas, such as nuclear weapons, where we can for example count warheads and delivery systems. Yet when it comes to new technologies and domains – including cyberspace and space – there are no comparable, quantitative, monitoring procedures. And research into new instruments for qualitative arms control is still in its infancy. As things stand, I simply don't know whether, or how, we can ensure verification – i.e. the monitoring of rule-compliant behaviour in future arms control. This has not yet been seriously or adequately researched and attempted, so it is too early for a final verdict.

*The interview was conducted by Sebastian Enskat.*

*–translated from German–*