ICT for peace foundation

Artificial Intelligence and Unarmed Aerial Vehicles in Humanitarian Action

Interview with Regina Surber, Senior Advisor ICT4Peace, by Camille Orend, Intern, UN OCHA
January 2020

1. CONSIDERING THE TECHNOLOGICAL ADVANCES IN AI, TO WHAT EXTENT ARE DRONES AUTONOMOUS?

To what extent drones can be considered autonomous depends on how we define 'autonomy' in the context of technological artefacts in general and drones in particular. Unfortunately, there exists no generally accepted definition of 'autonomy' for technological artefacts. What is more, my own approach to autonomy is distinct from the main-stream actors in the discussion, and will possibly not inform your answer in the way you wish.

As there does not exist any clear definition of 'autonomy' in AI and robotics, any input on your question, especially when compared to perspectives of other interviewees, must be read with great care.

Consider the following definitional controversies when talking about autonomous tech:

There exists a broader, operational, understandings of autonomy that regards *any* outcome by a machine that happens without human intervention as autonomous. In that sense, when a toaster ejects a toasted bread, that bread-ejection-operation is an autonomous one. Hence, a UAV used in humanitarian assistance (HA) that can perform a particular action (e.g. take-off) without human input, oversight, or veto-power, is 'autonomous'.

Narrower understandings, in turn, limit the term 'autonomy' to the more complex technological processes that arose through the research in the scientific field of AI. However, also here, there exist no generally accepted definitions. General or quasi-general acceptance of definitions, however, is the key value of any definition.

Main components of autonomous technologies in the narrow sense can be summarized as (1) the capacity to explore possibilities of action in a particular field (e.g. navigation, target-engagement) (2) with either limited or no human involvement, (3) in not pre-coded situations (i.e. unstructured situations / situations of uncertainty). (4) This leads to outcomes that are indeterminable, because those outcomes are per 'definition' responses to unanticipated events ('autonomy').

The components (1) – (4) can give primary information when talking about more modern and more complex software. However, they do not answer other questions, e.g. if a technology can be called 'autonomous' if only certain aspects of its functions are 'autonomous' in the above sense. Think about the following: is an armed drone autonomous if the flight control is autonomous, yet target engagement requires a human operator? Or, a question tailored to your purpose: is a humanitarian drone autonomous if flight is human-controlled, yet the topography-scanning, environmental analytics, or the identification of individuals happens autonomously?

As there exists no agreement on a definition, some experts also propose to consider a continuum of degrees of autonomy (e.g. automated, semi-autonomous, highly autonomous, fully autonomous). However, the different sub-categories (degree-range of autonomy) require definitions themselves. Hence, the problem of definition is multiplied. Also, the fact that some technologies are more 'capable' than others in a certain domain and with regards to certain technological features is common sense and as old as the first instruments in human history.

My understanding of and approach towards autonomy of technological artefacts departs from those considerations: To me, autonomy for technological artefacts does not cover clearly pre-defined characteristics, but describes the **result of a technological process for which the human cannot have control, and, hence cannot bear responsibility**. More important, the fact that it is us humans that create these technologies leads to a further observation: humans take a conscious decision to create instruments whose results they cannot control, nor bear responsibility for. Hence, I believe this outsourcing process reflects a particular psychological attitude: humans do not want to bear responsibility for those results.

Hence, the focus of discussions about autonomy for technological artefacts should shift to their human creators.

This shift, I believe, could be the real benefit of new technologies: they could shed a light on certain core human capacities that we may de-humanize in re-creating them artificially, and force us to evaluate to what extent we treasure what we are as a species.[1]

2. FULLY-AUTONOMOUS DRONES ARE MAINLY DEVELOPED BY DEFENSE CONTRACTORS FOR MILITARY PURPOSES.

And the civil sector.

3. HOW CAN WE ENVISION THE USE OF AUTONOMOUS UAVS IN HUMANITARIAN ACTION? WHAT ARE THE OPPORTUNITIES FOR HUMANITARIAN ACTION?

A key benefit and, hence, reason for use, is access to sites that humanitarian aid personnel cannot physically access, due to destruction through natural catastrophes or physical risks due to conflict situations.

Autonomous UAVs can be used for both **collection** and **delivery** of 'goods'. Collected 'goods' are mainly **immaterial** (data). Delivered 'goods' are mainly **material**, but can also be immaterial (information).

Collection:

- Topography mapping: video recording, processing of aerial images, creation of 3D imagery, environmental analytics.
- Identification of persons of interest: individuals in emergency situations.

Delivery:

- Selective material aid supplies: medication, food products
- Selective immaterial supplies: information (esp. when satellite networks are sabotaged or break down)

1 For a more detailed analysis of the term 'autonomy', 'AI', and definitional intricacies, see Surber, R., 2018, AI: AT, LAWS, and Peace-time threats, Geneva: ICT4Peace Foundation.

4. WHAT ARE THE CHALLENGES FOR HUMANITARIAN ACTORS RELATED TO THIS TECHNOLOGY?

Challenges of autonomous UAVs are **mainly ethical**, not operational.

One general ethical challenge **(1)** is that we do not know with certainty what risks emerging technologies involve, neither today nor in the future. Financial and cognitive resource investments in risk analysis of new technologies fall short compared to investments in the development of those technologies.

From an ethical point of view, the duty to analyze and minimize the risks of new technologies is especially strong when those technologies are used for purposes involving and affecting central human aspects, such as human life and human physical integrity and health. Humanitarian assistance is a prime example in this regard. What is more, HA's primary purpose is to address not only the human, but the *suffering* human, who is vulnerable and in special need of care. If emerging technologies are used in the process of addressing those needs, risks must be laid open, investigated in, minimized, and the humanitarian use of those technologies may need to be rejected in case it is ethically wrong, also despite minimal risks **(2)**.

Further intricate ethical problems arise when looking closer at what the humanitarian principles require, and to what extent the use of emerging technologies and autonomous UAVs could undermine those principles.

HA's principle of Impartiality requires that aid is delivered based exclusively on the criterion of need, meaning that priority is given to the most urgent cases. Analyses of who is in need, to what extent, in what way, and how in comparison to someone else, will pose a tremendous algorithmic challenge, not only in terms of bias, but of technological feasibility in general **(3a)**.

HA's principle of Independence requires that HA must itself be autonomous (!), meaning that assistance must be independent of economic, military, or other objectives. When using autonomous UAVs that gather, analyze, categorize, and create data that needs to be stored, Independence requires that this data will never be used for informing anything else than the humanitarian situation, i.e. no national intelligence, no businesses focused on behavioral analytics, no environmental agencies, etc. Whether or not this can be ensured (e.g. where to safely store the data?), and, more importantly, whether or not we *want* this to be ensured (e.g. data is money, also for HA organizations; also consider: HA data can be useful for developmental purposes, which would require either a stretching of the term 'humanitarian' or a violation of the principle of independence), are key questions that need to be addressed by high-level institutions **(3b)**.

A further, more intricate, ethical challenge arises through the potential requirement of value-tradeoffs **(4)**. See Question 5.

5. SHOULD THE ECONOMIC CONCEPT OF 'EFFICIENCY' BE TAKEN AS A REFERENCE POINT TO EVALUATE HUMANITARIAN RESPONSE TECHNOLOGY?

I believe that efficiency could be taken as a reference-point to evaluate humanitarian response if it is de-coupled from its classical economic objective. Consider the following:

The HA principle of Humanity requires that (a) human suffering must be addressed whenever it is found, that (b) life and health must be protected, and that (c) human dignity is respected. Those are three different goals.

If technology can help to achieve those goals more efficiently, i.e. through not over-using resources to achieve a humanitarian objective, then considering to what extent new technologies could lead to a more efficient way of achieving these goals, and to apply those technologies, can, prima facie, be regarded as a humanitarian duty.

However, using 'autonomous' technology for humanitarian goals can lead to a trade-off between those goals: e.g. topography-mapping and information-collecting drones may help create more efficient situational analyses and, hence, faster and situationally tailored disaster relief strategies (a). Yet, this could violate human dignity (c). How?

Autonomous UAVs in HA could violate human dignity in roughly four ways. First, (1), whether or not the collection of human (emotional, behavioral, biological, genetical, etc.) data violates human dignity is an open question. Human data collection in HA makes this question more pressing, as the data that is collected is data from vulnerable, suffering humans. Second (2), human dignity entails that humans are related to as humans; again, especially if they are in a situation of special vulnerability (HA). Autonomous UAVs analyze data points and, arguably, cannot relate to the human as dignity may require, as the human is a particular data point. Third, (3), UAVs in general (not only autonomous ones) used in HA could violate human dignity as they can instill a particular fear and undermine trust: sadly enough, military signature strikes by UAVs are not rare, and civilians know that, classically, the artificial bird brings death. Using UAVs in conflict struck areas for HA purposes may instill immediate fear, and undermine long-term trust in the humanitarian use of UAVs.

In other words, new technologies may be efficient ways to achieve a humanitarian goal (a) while at the same time undermining another humanitarian goal (c).

Which of the principles will outweigh may require either a general or a situational analysis. A general analysis that puts more importance to human dignity may lead to a general proscription of the humanitarian use of drones. A situational analysis may allow the use of drones in some humanitarian situations, but not in others. This may require a constant knowledgeable HA-ethics committee and constant supplies of on situational information.

6. THE RISKS OF USING AI-ENABLED / HIGHLY AUTONOMOUS TECHNOLOGIES ARE HIGH IN TERMS OF SAFETY AND ETHICS (ALGORITHM BIAS).

Safety is an ethical consideration as well.

SHOULD THE HUMANITARIAN SECTOR GIVE UP ON AI POWERED TECHNOLOGY?

This depends on value-judgments, value-weighing, and situational and contextual analysis. However, the overall ethical challenge (Questions 4 answer (1)), i.e. that we ignore possible risks of emerging technologies as they are underanalyzed, is a strong argument against a general acceptance of autonomous UAVs in HA with potential exceptions in ethically problematic situations.

In my view, the argument of Q4 A1 should force us to take the contrary position: we should err on the side of caution. In other words, we should adopt strong presumptions and decisions against using autonomous UAVs in HA until (a) risks are analyzed and (b) thorough ethical analyses have taken place, and in the meantime only use autonomous UAVs in situations where we are certain to run no or only minimal risks of harm.

IF NOT, WHAT SHOULD HUMANITARIANS DO IN ORDER TO ENSURE THAT THE PRINCIPLE OF 'DO NO HARM' IS GUARANTEED?

The answer depends on the understanding of 'harm', what 'do' means (whether it includes only intentional or also 'foreseeable' acts, and whether 'allowing' counts as 'doing' as well), and 'who' is doing / acting. Hence, a detailed answer may need more time and analysis.

The 'Do no harm' principle requires that every humanitarian situation is analyzed in both situational detail and context. Reaching an in-depth understanding of the disaster shall ensure that assistance is not misused or distorted.

Autonomous UAVs run on software and software can be hacked. The possibility that HA UAVs are hacked and misused exists. This risk must be weighed against the benefit (efficiency) of using them.

What is more, again, emerging technologies involve data gathering, analysis, categorization, creation, and storage. Data is untouchable, not to be grasped physically, easily manipulated, and easily stolen. In other words, autonomous UAVs create property (either of private individuals or semi-public HA organizations) to whose security from theft there arguably still exist no sufficient mechanism. In that sense, it may create a new platform for doing harm.