AIR SPACE
ASPC's crowd-sourced blog
https://airpower.airforce.gov.au

Cook, D. (2023) Air/Space 2 bp33316508
Date published: 27/11/2023

Conversations in human-machine teams – are we on the record?

Daniel Cook
Australian Defence Force

Most of us have blithely accepted the terms and conditions of software products we use in our private and business life. This generally hasn't been that important as we assume a mild erosion of our privacy in exchange for access to services we want or need. However, when translating the issue to human-machine teams for earning our livelihood we will need to pay much more attention to what we are agreeing to. When the future pilot signs for their aircraft and agree to the terms and conditions for using a futuristic AI-based aircraft, they had better understand what they are really signing for – their job may depend on it.

Rather than humans controlling machines, it will become increasingly possible for future Air Force capabilities to consist of systems requiring collaboration between human and intelligent machines. Human-machine teaming is 'a relationship – one made up of at least three equally important elements: the human, the machine, and the interactions and interdependencies between them (Konaev & Chahal, 2021).' The interaction between the human and 'intelligent' machine will likely have some physical element such as a pilot that currently controls an aircraft with a throttle and side stick. This will potentially be augmented with spoken language as natural language processing, such as ChatGPT[1], is added as an interaction method for human-machine teams.

Spoken commands could be an easy method for the human to direct the machine's actions and for the machine to communicate information/intent back to the human. For example, a human pilot provides directions to their aircraft to begin readying for take-off to start their mission using a language that is understandable to both human and machine. The aircraft then talks with the human pilot to work through checklists and communicates with air traffic control to arrange taxiing for take-off. This could work well in building the trust within the human-machine team as each element understands the other and their actions correlate with the intent.

The ease of communication between human and machine could raise team performance by relieving work load on the human or limit the number of humans required, such as replacing the co-pilot with a smart aircraft.

So what happens when the work is done? Long distance transits or lengthy time on combat air patrol leaves space for humans to be bored. Humans being humans are going to want to break the silence. If the machine can talk, then there is likely to be a conversation that may not necessarily be about work. With a human like response, the human could start to anthropomorphise their machine. While this may seem good or necessary for building trust within the human-machine team, this could begin to blur the line between what is private and what is work.

If it were with just another human, then the interaction would generally be considered an

[1] Chat Generative Pre-trained Transformer (ChatGPT) is a web interface that mimics human conversation developed by OpenAI Inc., which is an American company conducting research on artificial intelligence.

‘off-the-record’ private conversation. Maybe a problem is raised, an admission of guilt, regret over a poor action, or anger over a decision/action a superior made. While a colleague would probably understand the context of the conversation and provide some advice or light hearted banter, how would the machine respond?

This is where the terms and conditions agreed regarding interactions with the machine matter. Is the venting of a complaint against a superior within listening distance of the machine now going to result in a charge of insubordination (actively supported by the machines digital recording)? Will a human talking disrespectfully about a colleague obligate the machine to report this as an incident of unacceptable behaviour? Will a human lamenting their poor financial situation obligate the machine to report issues of security concern to vetting agencies? The complaints automatically injected into command by the machine could have serious consequences such as administrative or disciplinary action that could lead to termination of employment for the human in this team.

There is also the potential for the machine to initiate actions that break the trust in the team. Humans don’t start with the skills to do everything and have to learn how to employ the machines. Not everyone is the best at their job. Narrow AI have been demonstrated to bluff and use deception in simple board games. So what happens if the machines get smart enough that they decide not to trust the human through either self-preservation out of fear of the human’s performance or the machine becoming biased towards an individual because of its understanding of the human’s character? Could the machine refuse to hand over control? Will it manufacture situations where the pilot makes a mistake (such as not informing the pilot when they need to conduct checklists), breaks a rule/order (there are so many rules to be followed and some are contradictory), demonstrates poor performance (such as handing over manual flight control to a pilot under pressure for a complex situation or in bad weather conditions) or speaks ill of others? This situation only becomes worse when the video and voice recording is controlled by the machine with a direct means to report infractions to command. Will command get the full story of what happened and who would they believe more; the recollection of the human or the hard evidence provided by the machine?

The requirements for reporting built into the machine needs to be known and understood by the human in the team. If every minor infraction is reported and tracked then it is unlikely that the human will ever develop trust with the machine. This would also be akin to the Chinese Social Credit System (Donnelly, 2023). Command responsibility requires superiors to know when people under their command are not meeting the required standards. Command responsibility was a key issue raised in the Afghanistan Inquiry (Australian Defence Force, n.d.). There needs to be a degree of reporting on poor performance or rule breaking within the team. Finding a workable middle ground will be complex. It has to be somewhere between the extremes of an overzealous machine monitoring a social credit system and an unsupervised team that has the potential to go rogue.

Trust is an essential component of every team and the human-machine team will be no different. More importantly, this trust with a machine owned by your employer has to exist in the personal space as well as the work environment. While the humans need to trust the machine to do its job, they also need to trust that the machine won’t end their job. How both elements of personal and work are integrated into the human-machine team is going to be essential for their success.

References

Australian Defence Force. (n.d.). Afghanistan Inquiry. Retrieved from Australian Department of Defence: https://www.defence.gov.au/about/reviews-inquiries/afghanistan-inquiry

Donnelly, D. (2023, April 6). China Social Credit System Explained – What is it & How Does it Work? Retrieved from Horizons: https://nhglobalpartners.com/china-social-credit-system-explained/

Konaev, M., & Chahal, H. (2021, February 18). Building trust in human-machine teams. Retrieved from Brookings: https://www.brookings.edu/articles/building-trust-in-human-machine-teams/