

AI in Warfare (2024)

Edward Kim

Intelligence, Surveillance, Reconnaissance (ISR)



Ghost Robotics



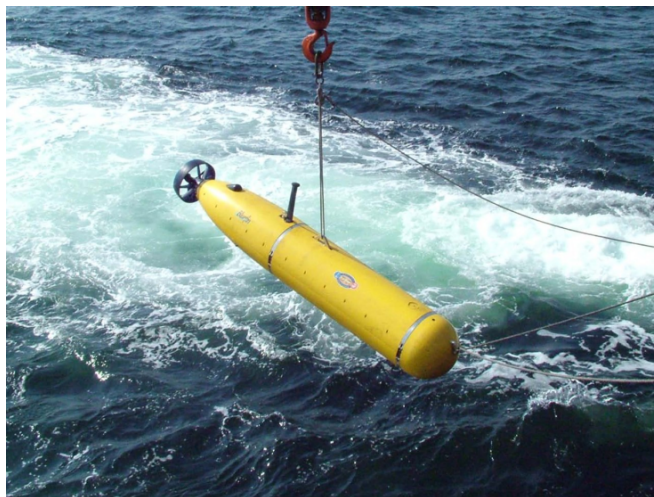
DARPA RACER



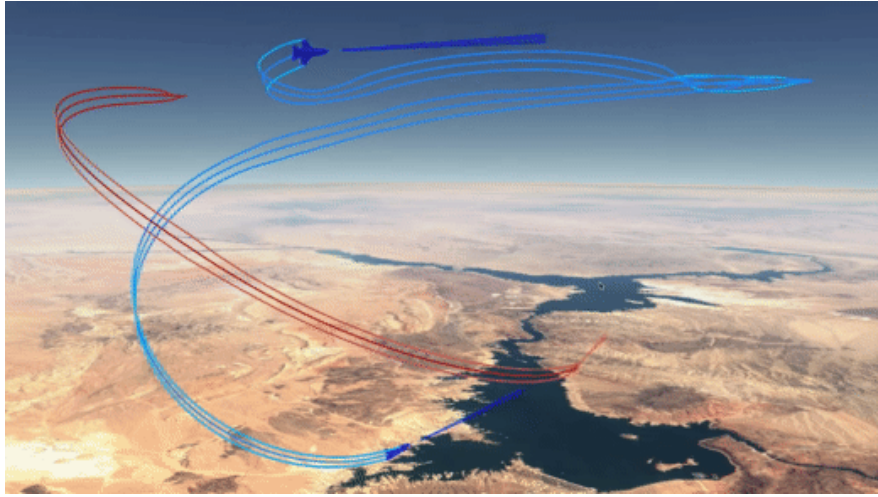
DARPA ACTUV - Sea Hunter



Bluefin Robotics - BPAUV



DARPA Air Combat Evolution



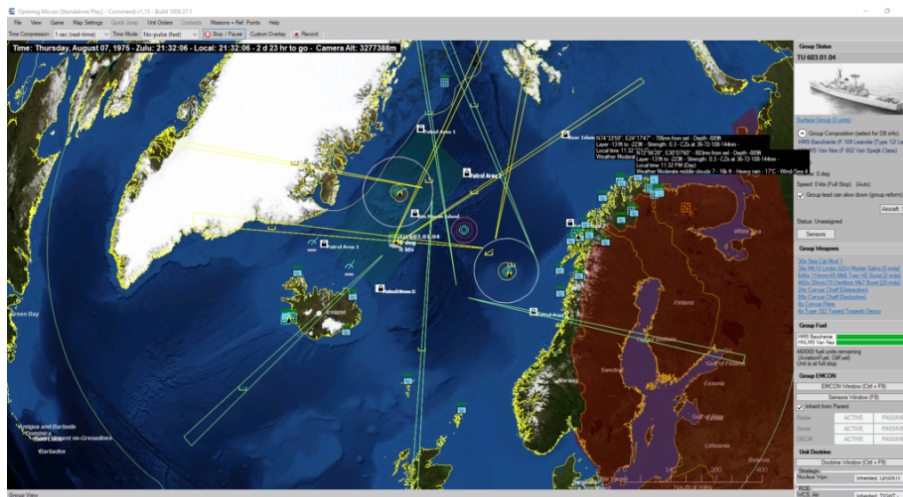
DARPA Air Combat Evolution



Micro-Drones: Insectothopter



DARPA Gamebreaker - Battle Management



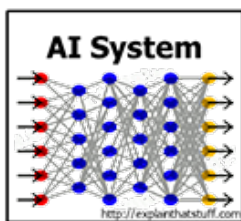
Human Control

Human-in-the-loop: System can prepare for action, but human must pull the trigger.

Human-on-the-loop: System is autonomous, but human can abort an action.

Human-out-of-the-loop: System is fully autonomous, either by design or effectively so due to the pace of the battlefield.

DARPA Explainable AI (2015 - 2021)



- We are entering a new age of AI applications
- Machine learning is the core technology
- Machine learning models are opaque, non-intuitive, and difficult for people to understand



- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

Battlefield Trust through Human-Machine Teaming



AGI in Warfare

Situational Awareness, Leopold Aschenbrenner

Power of Superintelligence

“It’ll mean having **billions of automated scientists and engineers and technicians**, each much smarter than the smartest human scientists, furiously inventing new technologies, day and night.”

Quantitatively superhuman: “Thinking” orders of magnitude faster than humans

Qualitatively superhuman: Completely novel and creative behaviors beyond human understanding

- Move 37 in AlphaGo vs. Lee Sedol
- Beating Minecraft in ?? minutes

Power of Superintelligence



Lead of Months in AGI → Decades of Progress

Imagine, military advances of the 20th century in less than 10 years:

- Horses and rifles and trenches
- 2 years later, modern tank armies,
- 2 years later, supersonic fighter planes, nuclear weapons, and ICBMs
- 2 years later, stealth technology and precision weapons that can knock out an enemy before they even know you're there

Post-Superintelligence Militaries

Aschenbrenner states that, even with a rudimentary form of superintelligence,

- Superhuman hacking to shut down pre-superintelligence militaries and their AI development
- Drone swarms (**including micro-drones**) that threaten instant death for every opposing leader, official, and their families
- **Advanced bioweapons** developed with AlphaFold-style simulation that could target specific ethnic groups (or 'everyone else except us')

What do we do?

Arms race for superintelligence **severely risks AI safety**

Do we have an alternative choice?