

Framtida AI utan AGI

EN SVÄRM AV insektsliknande drönare släpps ut från en skåpbil utanför ett universitet och tar sig in i en föreläsningssal. Med hjälp av små riktade sprängladdningar dödar insekterna elva av studenterna i salen. Det hela visar sig vara en globalt koordinerad terroraktion mot sammanlagt tusentals studenter, vilka alla figurerat i en protestaktion på sociala medier.

Scenariot är fiktivt, och figurerar i en kort video som släpptes hösten 2017 av Future of Life Institute lett av den svensk-amerikanske fysikern Max Tegmark. Sekvensen bjuder på blodiga scener och kritiserades för sensationsmakeri och skrämselfpropaganda, men trots dess science fiction-liknande innehåll bör den – sedd som varnande beskrivning av den destruktiva potentialen i en teknologi som praktiskt taget redan är här – betraktas som sakligt välgrundad.⁵¹

AI-teknik för autonoma vapen – eller mördarrobotar som de med en något mindre polerad term kallas – är en av de obehagligaste sidorna av dagens AI-utveckling. Vi står idag på randen till vad som kan bli en mycket farlig kapprustning på det området. En bild av hur farligt det kan bli ges i det öppna brev med rubriken *Autonomous Weapons: An Open Letter from AI and Robotics Researchers*, initierat av Future of Life Institute och

51 Cussins (2017), Haggström (2017b).

undertecknat av tusentals forskare världen över, som uppmanar till moratorium för utveckling av autonoma vapen. I brevet ges en allvarsam bild av vart en sådan kapprustning kan leda, med betoning på att de nya vapnen till skillnad från kärnvapen inte kräver några kostsamma eller svårtillgängliga råvaror, och därför kan väntas få en helt annan spridning:

Det är bara en tidsfråga innan [vapnen] dyker upp på den svarta marknaden och i händerna på terrorister, diktatorer som vill öka kontrollen över den egna befolkningen, krigsherrar med avsikt att genomföra etnisk rensning, etc. Autonoma vapen är idealiska för lönnmord, folkförtryck, destabilisering av nationer, och det selektiva dödandet av specifika etniska grupper. Vi tror av dessa skäl att en militär AI-kapprustning inte vore till gagn för mänskligheten.⁵²

Den avslutande meningen är uppenbarligen en kraftig underdrift, och det finns mer att säga. En aspekt har att göra med skalbarhet: 100000 Kalashnikovs kan inte användas till att döda 100000 gånger fler fiender än en enda med mindre än att vi har lika många soldater att sätta vapnen i händerna på. Om vi ersätter Kalashnikovs med mördarrobotar fungerar däremot motsvarande uppskalning av dödlighetskapacitet utan att på samma sätt behöva öka den militära personalstyrkan.⁵³

Frågan om autonoma vapen kändes brådskande 2015 då brevet skrevs, och är än mer brådskande idag. Ändå träffar man ibland på någon som försöker tona ned dessa risker, som i den paneldiskussion i EU-parlamentet i Bryssel 2017 på temat AI-risker, där jag deltog tillsammans med bland andra den

52 Russell m. fl. (2015), min översättning.

53 Russell (2019).

uppburne kognitionsvetaren och vetenskapspopularisatorn Steven Pinker. Han framhöll att det skulle krävas en galning för att konstruera något så fruktansvärt som »en svärm av robotinsekter avsedd att attackera enskilda människor baserat på ansiktsgenkänning«, och att det idag inte längre finns utrymme för sådana enstaka galningar eftersom alla betydande ingenjörprojekt numera genomförs i stora samarbeten.⁵⁴

Den obändiga optimism Pinker ger uttryck för bygger på att han ignorerar såväl den inre logiken i kapprustningar och det militärindustriella komplexet som att minst lika hemiska massförstörelsevapen har utvecklats i över 70 års tid – inte av ensamma galningar utan just i stora samarbetsprojekt. Mest känt är det amerikanska Manhattanprojektet på 1940-talet, som utvecklade de första kärnvapnen, men det finns många andra. Varför skulle allt sådant utvecklingsarbete plötsligt upphöra av sig självt? Faran med en militär AI-kapprustning är verklig och något vi behöver förhålla oss till.

* * *

Problemet med den begynnande militära AI-kapprustningen hör till den kategori i listan i Kapitel 1 över frågor som betecknades jordnära – alltså frågor som handlar om AI av det slag som tydligt ligger i tangentens riktning i förhållande till den teknik vi har idag, och som inte förutsätter något AGI-genombrott. Det finns många andra sådana frågor med långtgående potentiella konsekvenser för samhället och våra liv, och avsikten

⁵⁴ Häggström (2018a). Här kan man misstänka att Pinker var inspirerad av den ovan nämnda videon från Future of Life Institute, men vad som i någon mån talar emot det är att den släpptes först några veckor efter Brysselmötet.