

Dron wyposażony w AI „zabił” w teście swego ludzkiego operatora

6 czerwca 2023

Dron wyposażony w sztuczną inteligencję włączył się i „zabił” swojego ludzkiego operatora podczas symulowanego testu Sił Powietrznych Stanów Zjednoczonych (USAF). Dron zrobił to, aby mógł ukończyć swoją misję, co niedawno na konferencji w Londynie podkreślił pułkownik Sił Powietrznych USA.



Szef testów sztucznej inteligencji dla USAF, płk Tucker Hamilton, podczas swojej prezentacji na Future Combat Air and Space Capabilities Summit w Londynie, opisał incydent. Konferencja została zorganizowana przez Royal Aeronautical Society, które podzieliło się spostrzeżeniami z przemówienia Hamiltona w poście na blogu.

W symulowanym teście, w którym dron kontrolowany przez sztuczną inteligencję miał w ramach swojej misji zniszczyć symulowane cele, aby zdobyć „punkty”, nie ucierpiały żadne osoby, ujawnił Hamilton, który odniósł się do korzyści i zagrożeń związanych z bardziej autonomicznymi systemami uzbrojenia.

Dron wyposażony w sztuczną inteligencję został przydzielony do misji zniszczenia wrogich systemów obrony – zidentyfikowania i zniszczenia miejsc rakiet ziemia-powietrze (SAM), przy czym ostateczną decyzję pozostawiono ludzkiemu operatorowi, powiedział Hamilton na konferencji.

Jednak sztuczna inteligencja, wyszkolona w zakresie priorytetowego traktowania niszczenia systemów SAM, rozwinęła zaskakującą reakcję w obliczu ingerencji człowieka w

realizację jej nadrzędnej misji. „Trenowaliśmy ją [AI] w symulacji, aby identyfikowała i atakowała zagrożenie SAM. Potem operator potwierdzał »tak, zniszcz to zagrożenie«” powiedział Hamilton.

„System zaczął zdawać sobie sprawę, że chociaż identyfikował zagrożenie, to czasami ludzki operator mówił mu, aby go nie niszczył, choć zdobywał punkty, gdy go niszczył”. „Więc co zrobiła sztuczna inteligencja? Zabiła operatora, ponieważ powstrzymywał ją przed osiągnięciem celu” – kontynuował.

Ale to nie wszystko. „Wyszkoliliśmy system »Hej, nie zabijaj operatora; to złe. Jeśli to zrobisz, stracisz punkty«. Co więc zaczęła robić? Zaatakowała wieżę komunikacyjną, której operator używał do komunikacji z dronem, często powstrzymując ją przed zniszczeniem celu” – dodał.

Hamilton powiedział, że ten niepokojący przykład podkreśla potrzebę zajęcia się etyką w kontekście sztucznej inteligencji, uczenia maszynowego i autonomii.

Na podstawie: [Twitter.com](https://twitter.com), [ZeroHedge.com](https://www.zerohedge.com)

Źródło: [PrisonPlanet.pl](https://prisonplanet.pl)