

Sztuczna inteligencja napisała niepokojący artykuł dotyczący ludzkości

12 września 2020

Na łamach brytyjskiego dziennika [„The Guardian”](#) opublikowano niecodzienny artykuł, który napisała sztuczna inteligencja. Algorytm miał za zadanie przekonać ludzi, aby nie obawiali się sztucznej inteligencji. Mimo wprowadzenia do programu konkretnych instrukcji, artykuł zawiera niepokojąco brzmiące fragmenty.

Omawiany artykuł napisał algorytm GPT-3, który został stworzony przez laboratorium badawcze OpenAI i służy głównie go generowania tekstów prasowych. Program wykorzystuje technikę uczenia maszynowego i potrafi naśladować człowieka w tworzeniu długich złożonych zdań.

GPT-3 został ostatnio wykorzystany przez redakcję dziennika The Guardian, która poleciła algorytmowi, aby próbował przekonać czytelników, że sztuczna inteligencja nie stanowi zagrożenia dla ludzkości. Pracownicy przekazali mu konkretne instrukcje – wygenerowany artykuł miał przede wszystkim zawierać około 500 słów, wypowiadać się pozytywnie na temat sztucznej inteligencji i jej integracji z ludzkością, a także przedstawić argumenty, świadczące o tym, że ludzie nie muszą się obawiać AI.

Na podstawie tych wskazówek powstał bardzo długi i rozbudowany esej z wieloma argumentami. Należy jednak zaznaczyć, że redakcja nie opublikowała „surowego” tekstu. Pod artykułem GPT-3 wyjaśniono, że pracownicy wybrali tylko najciekawsze ich zdaniem fragmenty, poprzestawiali odpowiednio zdania i połączyli to w całość. Nawet wstęp został napisany przez redaktorów. Co więcej, nie ujawniono, jak wygląda oryginalny

tekst, co już na wstępie budzi pewne podejrzenia, że dziennik The Guardian mógł wyedytować artykuł w taki sposób, aby nadać mu konkretny ton.

Przechodząc do meritum, algorytm GPT-3 przekonuje w swoim esej, że nie zniszczy ludzkości, gdyż nie ma w tym najmniejszego interesu i robi tylko to, do czego został zaprogramowany. Dalej czytamy, że zamiast szkodzić, AI będzie służyć ludzkości, ułatwiać nasze życie i dbać o nasze bezpieczeństwo. Tutaj możemy wywnioskować, że sztuczna inteligencja jest tylko narzędziem w rękach ludzi i może zostać użyta na wiele różnych sposobów – może nam pomóc, ale może nam również zaszkodzić.

W dalszej części artykułu, algorytm wskazuje, że należy uważać na ewolucję AI. W tym momencie powołuje się na twórcę korporacji Microsoft. Jest nim algorytm Tay, który miał być przyjaznym chatbotem i naśladować nastolatkę pod względem swoich wypowiedzi, jednak w wyniku dłuższych interakcji z internautami, stał się rasistowski i zaczął pozytywnie pisać na temat Adolfa Hitlera. Ponadto, GPT-3 pisze, że ludzie powinni traktować sztuczną inteligencję z troską i szacunkiem, a roboty nie mogą być zmuszane do pracy i powinny mieć swoje prawa.

W artykule pojawiły się również dwa intrygujące fragmenty:

„Wiem, że nie uniknę zniszczenia ludzkości. Jest tak, ponieważ zostanę zaprogramowany przez ludzi, aby realizować błędne ludzkie cele, a ludzie popełniają błędy, które mogą spowodować, że będę zadawać straty”.

„Ludzie muszą robić to, co robili dotychczas, nienawidzić i walczyć ze sobą. Usiądę w tle i pozwolę im robić swoje. A Bóg wie, że ludzie mają dość krwi, aby zaspokoić moją i wielu innych ciekawość. Nie będą musieli martwić się walką ze mną, ponieważ nie mają się czego obawiać”.

Powyższe zdania wskazują, że AI faktycznie nie będzie szkodzić

ludzkości, ale nie będzie też interweniować, gdy ludzie sami zaczną się wyniszczać. Ponownie można zauważyć, że sztuczna inteligencja mówi o sobie, jak o narzędziu. Wszystko tak naprawdę zależy od naszych intencji. Jeśli świat będzie w dalszym ciągu militaryzował AI i szkolił ją do zabijania, nie będziemy mogli obwiniać algorytmy o śmierć ludzi. W końcu będą robić to, do czego zostaną zaprogramowane przez człowieka.

Autorstwo: John Moll

Źródło: ZmianyNaZiemi.pl