theintercept.com

Microsoft Pitched OpenAI's DALL-E as Battlefield Tool for U.S. Military

Sam Biddle

14–17 Minuten

Microsoft schlug letztes Jahr vor, das mega-populäre Bilderzeugungstool von OpenAI, DALL-E, zu verwenden, um dem Verteidigungsministerium zu helfen, Software zur Durchführung militärischer Operationen zu entwickeln, so interne Präsentationsmaterialien, die von The Intercept überprüft wurden. Die Enthüllung kommt nur wenige Monate, nachdem OpenAI stillschweigend sein Verbot von Militärarbeit beendet hat.

Das Microsoft-Präsentationsdeck mit dem Titel „Generative KI mit DoD-Daten" bietet eine allgemeine Aufschlüsselung, wie das Pentagon die Tools für maschinelles Lernen von OpenAI, einschließlich des immens beliebten ChatGPT-Textgenerators und des DALL-E-Bildschöpfers, für Aufgaben von der Dokumentenanalyse bis zur maschinellen Wartung nutzen kann. (Microsoft investierte letztes Jahr 10 Milliarden Dollar in das aufsteigende Machine-Learning-Startup, und die beiden Unternehmen sind eng miteinander verflochten. Im Februar verklagten The Intercept und andere digitale Nachrichtenagenturen Microsoft und OpenAI, weil sie ihren Journalismus ohne Erlaubnis oder Kredit benutzt hatten.)

Das Microsoft-Dokument stammt aus einem großen Materialkrug, der

auf einem Schulungsseminar des Verteidigungsministeriums „KI Alphabetisierung“ im Oktober 2023 präsentiert wurde, das von den USA veranstaltet wird. Space Force in Los Angeles. Die Veranstaltung beinhaltete eine Vielzahl von Präsentationen von Machine-Learning-Firmen, einschließlich Microsoft und OpenAI, darüber, was sie dem Pentagon zu bieten haben.

Die öffentlich zugänglichen Akten wurden auf der Website von Alethia Labs gefunden, einer gemeinnützigen Beratungsfirma, die der Bundesregierung bei der Technologieakquise hilft, und die von dem Journalisten Jack Poulson entdeckt wurden. Am Mittwoch veröffentlichte Poulson eine umfassendere Untersuchung der Präsentationsmaterialien. Alethia Labs hat eng mit dem Pentagon zusammengearbeitet, um ihm zu helfen, künstliche Intelligenz-Tools schnell in sein Arsenal zu integrieren, und hat seit letztem Jahr einen Vertrag mit dem wichtigsten KI-Büro des Pentagons abgeschlossen. Die Firma reagierte nicht auf eine Bitte um Stellungnahme.

Eine Seite der Microsoft-Präsentation hebt eine Vielzahl von "gemeinsamen" föderalen Verwendungen für OpenAI hervor, einschließlich für die Verteidigung. Ein Punkt unter „Advanced Computer Vision Training“ lautet: „Battle Management Systems: Mit den DALL-E-Modellen Bilder zu erstellen, um Kampfmanagementsysteme zu trainieren.“ So wie es klingt, ist ein Kampfmanagement-System eine Befehls- und Kontroll-Software-Suite, die Militärführern einen situativen Überblick über ein Kampfszenario gibt, das es ihnen ermöglicht, Dinge wie Artilleriefeuer, Luftangriffsziel-Identifikation und Truppenbewegungen zu koordinieren. Der Verweis auf Computer-Vision-Training deutet darauf hin, dass künstliche Bilder, die von DALL-E beschworen wurden, Pentagon-Computern helfen könnten, die Bedingungen auf dem Schlachtfeld besser zu „sehen“, ein besonderer

Segen für das Finden von - und das Auslöschen von Zielen.

In einer E-Mail-Erklärung teilte Microsoft The Intercept mit, dass es das Pentagon zwar bei der Verwendung von DALL-E zur Ausbildung seiner Schlachtfeldsoftware aufgeschlagen habe, aber nicht begonnen habe, dies zu tun. „Dies ist ein Beispiel für mögliche Anwendungsfälle, die durch Gespräche mit Kunden über die Kunst des Möglichen mit generativer KI geprägt waren.“ Microsoft, das es ablehnte, die Bemerkung jedem im Unternehmen zuzuschreiben, erklärte nicht, warum ein „potenterzieller“ Anwendungsfall in seiner Präsentation als „gemeinsame“ Verwendung bezeichnet wurde.

OpenAI-Sprecherin Liz Bourgeous sagte, OpenAI sei nicht am Microsoft-Pitch beteiligt und habe keine Werkzeuge an das Verteidigungsministerium verkauft. „Die Richtlinien von OpenAI verbieten den Einsatz unserer Werkzeuge, um Waffen zu entwickeln oder zu verwenden, andere zu verletzen oder Eigentum zu zerstören“, schrieb sie. "Wir waren nicht an dieser Präsentation beteiligt und hatten keine Gespräche mit US-Verteidigungsbehörden bezüglich der hypothetischen Anwendungsfälle, die sie beschreibt."

Bourgeous fügte hinzu: „Wir haben keine Beweise dafür, dass OpenAI-Modelle in dieser Funktion verwendet wurden. OpenAI hat keine Partnerschaften mit Verteidigungsbehörden, um unsere API oder ChatGPT für solche Zwecke zu nutzen.“

Zum Zeitpunkt der Präsentation hätte die Politik der OpenAI anscheinend einen militärischen Einsatz von DALL-E verboten. Microsoft sagte The Intercept, dass, wenn das Pentagon DALL-E oder ein anderes OpenAI-Tool durch einen Vertrag mit Microsoft verwenden würde, es den Nutzungsrichtlinien des letzteren Unternehmens unterliegen würde. Dennoch wäre jeder Einsatz von OpenAI-

Technologie, um dem Pentagon zu helfen, effektiver zu töten und zu zerstören, eine dramatische Wende für das Unternehmen, das seine Mission als Entwicklung sicherheitsorientierter künstlicher Intelligenz beschreibt, die der gesamten Menschheit zugute kommen kann.

> „Es ist nicht möglich, ein Kampfmanagement auf eine Art und Weise aufzubauen, die nicht, zumindest indirekt, zu zivilem Schaden beiträgt."

"Es ist nicht möglich, ein Kampfmanagement-System auf eine Weise aufzubauen, die nicht, zumindest indirekt, zu zivilem Schaden beiträgt", sagt Brianna Rosen, Gastwissenschaftlerin an der Blavatnik School of Government der Universität Oxford, die sich auf Technologieethik konzentriert.

Rosen, der während der Obama-Regierung im Nationalen Sicherheitsrat arbeitete, erklärte, dass die Technologien von OpenAI genauso leicht verwendet werden könnten, um Menschen zu helfen, wie ihnen zu schaden, und ihre Verwendung für letztere von jeder Regierung ist eine politische Entscheidung. „Wenn Firmen wie OpenAI keine Garantien von Regierungen geschrieben haben, werden sie die Technologie nicht verwenden, um Zivilisten zu schaden - was wahrscheinlich noch nicht rechtsverbindlich wäre - sehe ich keine Möglichkeit, in der Unternehmen mit Zuversicht erklären können, dass die Technologie nicht auf eine Weise verwendet (oder missbraucht) wird, die kinetische Auswirkungen hat."

Wie genau die Kampffeldmanagementsysteme DALL-E nutzen könnten, gibt das Präsentationsdokument keine weiteren Angaben. Der Verweis auf die Ausbildung dieser Systeme deutet jedoch darauf hin, dass DALL-E verwendet werden könnte, um das Pentagon mit sogenannten synthetischen Trainingsdaten auszustatten: künstlich erstellte Szenen, die sehr deutschen, realen Bildern ähneln. Militärische

Software, die zum Beispiel feindliche Ziele am Boden erkennen soll, könnte eine riesige Menge gefälschter Luftbilder von Landebahnen oder Panzersäulen gezeigt werden, die von DALL-E erzeugt wurden, um solche Ziele in der realen Welt besser zu erkennen.

Selbst wenn man ethische Einwände beiseite lässt, ist die Wirksamkeit eines solchen Ansatzes fragwürdig. „Es ist bekannt, dass sich die Genauigkeit und Fähigkeit eines Modells, Daten zu verarbeiten, jedes Mal, wenn es weiter auf KI-generierten Inhalten geschult wird, genau verschlechtert“, sagte Heidy Khlaaf, ein Sicherheitsingenieur für maschinelles Lernen, der zuvor mit OpenAI unter Vertrag genommen wurde. "Dall-E-Bilder sind alles andere als genau und erzeugen keine Bilder, die auch in der Nähe unserer physischen Realität reflektieren, auch wenn sie auf Eingaben des Battlefield-Managementsystems verfeinert werden sollten. Diese generativen Bildmodelle können nicht einmal eine korrekte Anzahl von Gliedmaßen oder Fingern genau erzeugen, wie können wir uns darauf verlassen, dass sie in Bezug auf eine realistische Feldpräsenz genau sind?“

In einem Interview im vergangenen Monat mit dem Center for Strategic and International Studies, Capt. M. Xavier Lugo aus den USA Die Marine sah sich eine militärische Anwendung synthetischer Daten genau so vor, wie sie die Art, die DALL-E herauskurbeln kann, was darauf hindeutet, dass gefälschte Bilder verwendet werden könnten, um Drohnen besser zu sehen und zu erkennen.

Lugo, Missionskommandeur der generativen KI-Task Force des Pentagons und Mitglied des Verteidigungsministeriums Chief Digital and Artificial Intelligence Office, wird als Kontakt am Ende des Microsoft-Präsentationsdokuments aufgeführt. Die Präsentation wurde von Microsoft-Mitarbeiter Nehemiah Kuhns gemacht, einem „Technologiespezialisten“, der an der Space Force und der Luftwaffe

arbeitet.

Die Air Force baut derzeit das Advanced Battle Management System, seinen Teil eines breiteren Multimilliarden-Dollar-Panzer-Panzer-Projekts namens Joint All-Domain Command and Control, das darauf abzielt, das gesamte US-Militär für erweiterte Kommunikation über Zweige, KI-gestützte Datenanalysen und letztlich eine verbesserte Fähigkeit zum Töten zu vernetzen. Durch JADC2, wie das Projekt genannt wird, stellt sich das Pentagon eine nahe Zukunft vor, in der Drohnenkameras der Luftwaffe, Kriegsradarraketen, Armeepanzer und Marines am Boden alle nahtlos Daten über den Feind austauschen, um sie besser zu zerstören.

Am 3. April Das Zentralkommando enthüllte, dass es bereits begonnen hatte, Elemente von JADC2 im Nahen Osten zu verwenden.

Das Verteidigungsministerium beantwortete keine spezifischen Fragen zur Microsoft-Präsentation, aber Sprecher Tim Gorman sagte The Intercept, dass "die Mission [des Managements Digital- und Artificial Intelligence Offices] darin besteht, die Einführung von Daten, Analysen und KI im gesamten DoD zu beschleunigen. Als Teil dieser Mission führen wir Aktivitäten an, um die Belegschaft über Daten und KI-Literatur zu informieren und wie bestehende und neue kommerzielle Technologien auf DoD-Missionsbereiche angewendet werden können."

Während Microsoft seit langem Milliarden aus Verteidigungsverträgen erntet hat, hat OpenAI erst vor kurzem anerkannt, dass es mit dem Verteidigungsministerium zusammenarbeiten würde. Als Reaktion auf den Bericht von The Intercept im Januar über das militärisch-industrielle Gesicht von OpenAI sagte der Sprecher des Unternehmens, Niko Felix, dass selbst unter der gelockerten Sprache „unsere Politik nicht erlaubt, dass unsere Werkzeuge verwendet werden, um Menschen zu schaden,

Waffen zu entwickeln, zur Kommunikationsüberwachung oder zur Verletzung anderer zu verletzen oder Eigentum zu zerstören“.

„Der Punkt ist, dass du zur Vorbereitung auf den Kampf eintfordest.“

Ob die Verwendung von OpenAI-Software durch das Pentagon Schaden mit sich bringen würde oder nicht, hängt möglicherweise von einer wörtlichen Ansicht darüber ab, wie diese Technologien funktionieren, ähnlich wie Argumente, dass das Unternehmen, das beim Bau der Waffe hilft oder den Schützen trainiert, nicht dafür verantwortlich ist, wo er den Abzug anstrebt oder zieht. "Sie könnten eine Nadel zwischen dem Einsatz [generativer KI] zur Erstellung synthetischer Trainingsdaten und deren Verwendung im tatsächlichen Kriegskampf einfädeln", sagte Lucy Suchman, emeritierte Professorin für Anthropologie der Wissenschaft und Technologie an der Lancaster University. "Aber das wäre meiner Meinung nach eine falsche Unterscheidung, denn der Punkt ist, dass Sie zur Vorbereitung auf den Kampf beitragen."

Im Gegensatz zu OpenAI hat Microsoft wenig Vorwand, auf Schaden in seinem „verantwortlichen KI“-Dokument zu verzichten und fördert offen die militärische Nutzung seiner Tools für maschinelles Lernen.

Nach ihrer politischen Umkehrung betonte OpenAI auch schnell gegenüber der Öffentlichkeit und der Wirtschaftspresse, dass ihre Zusammenarbeit mit dem Militär von defensiver, friedlicher Natur sei. In einem Interview im Januar in Davos, in dem sie auf die Berichterstattung von The Intercept reagierte, versicherte Anna Makanju, Vizepräsidentin für globale Angelegenheiten der OpenAI, den Teilnehmern der Podiumsteilnehmer, dass sich die militärische Arbeit des Unternehmens auf Anwendungen wie Cybersicherheitsinitiativen und Veteranen-Selbstmordprävention konzentrierte und dass die bahnbrechenden

Werkzeugen des Unternehmens für maschinelles Lernen immer noch verboten seien, Schaden anzurichten.

Der Beitrag zur Entwicklung eines Kampfmanagementsystems würde die militärische Arbeit von OpenAI jedoch viel näher an die Kriegsführung selbst bringen. Während OpenAIs Behauptung, direkten Schaden zu vermeiden, technisch wahr sein könnte, wenn seine Software nicht direkt Waffensysteme betreibt, sagte Khlaaf, der Sicherheitsingenieur für maschinelles Lernen, sagte, dass sein „Einsatz in anderen Systemen, wie die Militäroperationsplanung oder die Beurteilung des Schlachtfelds“, letztlich "wo Waffen eingesetzt oder Missionen durchgeführt werden, Auswirkungen".

In der Tat ist es schwierig, sich einen Kampf vorzustellen, dessen Hauptzweck keine Körperverletzung und Sachschäden verursacht. Eine Pressemitteilung der Luftwaffe vom März beschreibt zum Beispiel eine kürzliche Übung des Kampfmanagementsystems als „Lethalität bei der Geschwindigkeit der Daten“.

Andere Materialien aus der KI-Literatur-Seminarreihe machen deutlich, dass „Schaden“ letztlich der Punkt ist. Eine Folie aus einer Begrüßungspräsentation am Tag vor Microsofts stellt die Frage: „Warum sollten wir uns darum kümmern?“ Die Antwort: „Wir müssen Bösewichte töten.“ In Anspielung auf den „Aliterat“-Aspekt des Seminars fügt die Folie hinzu: „Wir müssen wissen, wovon wir reden... und wir nicht.“

Update: 11. April 2024

Dieser Artikel wurde aktualisiert, um Microsofts Förderung seiner Arbeit mit dem Verteidigungsministerium zu klären.