Administracija američkog predsjednika Joea Bidena spremna je otvoriti novo bojište u svojim nastojanjima da zaštiti američku umjetnu inteligenciju od Kine i Rusije s preliminarnim planovima za postavljanje zaštite oko najnaprednijih modela umjetne inteligencije.

Istraživači iz vladinog i privatnog sektora zabrinuti su da bi američki protivnici mogli koristiti modele, koji pretražuju goleme količine teksta i slika za sažimanje informacija i generiranje sadržaja, za kibernetičke napade ili čak proizvodnju snažnog biološkog oružja.

Evo nekih prijetnji koje predstavlja AI:

'DEEPFAKES' I DEZINFORMACIJE

Deepfakeovi - realistični, ali izmišljeni videozapisi koje su izradili algoritmi umjetne inteligencije na temelju velikog broja objavljenih snimaka - pojavljuju se na društvenim mrežama, zamagljujući činjenice i fikciju u polariziranom svijetu američke politike.

Iako su takvi sintetički mediji prisutni već nekoliko godina, tijekom prošle godine pojačani su nizom novih "generativnih AI" alata kao što je Midjourney koji čine jeftinim i lakim stvaranje uvjerljivih deepfakeova.

Alati za izradu slika koje pokreće umjetna inteligencija tvrtki uključujući OpenAI i Microsoft mogu se koristiti za izradu fotografija koje bi mogle promicati dezinformacije povezane s izborima ili glasanjem, unatoč tome što svaki od njih ima politiku protiv stvaranja obmanjujućeg sadržaja, rekli su istraživači u izvješću u ožujku.

Neke kampanje dezinformiranja jednostavno koriste sposobnost umjetne inteligencije da oponaša stvarne novinske članke kao sredstvo za širenje lažnih informacija.

Dok su glavne platforme društvenih medija poput Facebooka, X-a (bivšeg Twittera) i YouTubea uložile napore u zabranu i uklanjanje deepfakeova, njihova učinkovitost u nadzoru nad takvim sadržajem varira.

Na primjer, prošle godine, stranica s vijestima pod kontrolom kineske vlade koja koristi generativnu AI platformu progurala je prethodno cirkuliranu lažnu tvrdnju da Sjedinjene Države vode laboratorij u Kazahstanu za stvaranje biološkog oružja za upotrebu protiv Kine, navodi Ministarstvo domovinske sigurnosti u svojoj procjeni domovinske za 2024.

Savjetnik za nacionalnu sigurnost Jake Sullivan, govoreći na skupu o AI-u u Washingtonu, rekao je da problem nema lakih rješenja jer kombinira kapacitet AI s "namjerom državnih i nedržavnih aktera da koriste dezinformacije u velikim razmjerima, poremete demokracije, unaprijede propagandu, oblikuju percepciju u svijetu."

"U ovom trenutku napad uvelike nadmašuje kapacitete obrane," upozorio je.

BIOLOŠKO ORUŽJE

Američka obavještajna zajednica, think tankovi i akademici sve su više zabrinuti zbog rizika koje

predstavljaju strani akteri koji dobivaju pristup naprednim AI sposobnostima. Istraživači Gryphon Scientific i Rand Corporation primijetili su da napredni AI modeli mogu pružiti informacije koje bi mogle pomoći u stvaranju biološkog oružja.

Gryphon je proučavao kako se veliki jezični modeli (LLM) - računalni programi koji izvlače iz ogromnih količina teksta za generiranje odgovora na upite - mogu neprijateljima koristiti za nanošenje štete u domeni znanosti o životu i otkrio je da oni "mogu pružiti informacije koje bi mogle pomoći pri stvaranju biološkog oružja pružanjem korisnih, točnih i detaljnih informacija u svakom koraku na ovom putu."

Otkrili su, na primjer, da LLM može pružiti znanje na postdoktorskoj razini za rješavanje problema pri radu s virusom sposobnim za pandemiju.

Randovo istraživanje pokazalo je da bi LLM mogli pomoći u planiranju i izvršenju biološkog napada. Otkrili su da doktor medicine može, na primjer, predložiti metode isporuke botulinum toksina aerosolom.

NOVI NAPORI ZA RJEŠAVANJE PRIJETNJI

Ministarstvo domovinske sigurnosti (DHS) priopćilo je da će cyber akteri vjerojatno koristiti umjetnu inteligenciju za "razvijanje novih alata" za "omogućavanje većih, bržih, učinkovitijih cyber napada" protiv kritične infrastrukture uključujući cjevovode i željeznice, u svojoj procjeni domovinske prijetnje za 2024.

Kina i drugi protivnici razvijaju AI tehnologije koje bi mogle potkopati kibernetičku obranu SAD-a, smatra DHS, uključujući generativne AI programe koji podržavaju napade zlonamjernog softvera.

Microsoft je u izvješću iz veljače rekao da je pratio hakerske grupe povezane s kineskom i sjevernokorejskom vladom, kao i ruskom vojnom obavještajnom službom i iranskom Revolucionarnom gardom, dok su pokušavali usavršiti svoje hakerske kampanje koristeći velike jezične modele.

Dvostranačka skupina zastupnika predstavila je nacrt zakona koji bi Bidenovoj administraciji olakšao nametanje kontrole izvoza modela umjetne inteligencije, u pokušaju da zaštiti cijenjenu američku tehnologiju od stranih loših aktera.

Prijedlog zakona, koji su sponzorirali republikanci u Zastupničkom domu Michael McCaul i John Molenaar te demokrati Raja Krishnamoorthi i Susan Wild, također bi Ministarstvu trgovine dao izričitu ovlast da zabrani Amerikancima da rade sa strancima na razvoju sustava umjetne inteligencije koji predstavljaju rizik za nacionalnu sigurnost SAD-a.

Tony Samp, savjetnik za politiku umjetne inteligencije u DLA Piperu u Washingtonu, rekao je da kreatori politike u Washingtonu pokušavaju "poticati inovacije i izbjegavati oštre propise koji guše inovacije" dok se nastoje pozabaviti brojnim rizicima koje donosi tehnologija. No upozorio je da bi "suzbijanje razvoja umjetne inteligencije kroz regulaciju moglo spriječiti potencijalne proboje u područjima kao što su otkrivanje lijekova, infrastruktura, nacionalna sigurnost i druga,

te ustupiti teren konkurenciji u inozemstvu". (Hina)