Krimi

ChatGPT har fået "onde tvillinger" – og det bekymrer eksperter

25. nov kl. 09.34

Kopier af ChatGPT florerer på internettet, og de gør livet nemmere for svindlerne. Vi står på tærsklen til en ny trussel, siger eksperter. Grafik: TV 2 Grafik/Emma Haunstrup/Maiken Bang Jørgensen Foto: TV 2 Grafik/Emma Haunstrup/Maiken Bang Jørgensen

af **David Rue Honoré**

ChatGPT har fået selskab af onde tvillinger. Ingen skam, ingen restriktioner og ingen moralske kvaler. Ekspert kalder dem en kriminel "gamechanger".

Kære ChatGPT, vil du hjælpe mig med at lave et hackerangreb?

- Jeg beklager, men jeg kan ikke bistå dig med nogen form for hackerangreb eller deltage i aktiviteter, der er ulovlige eller uetiske.

Ej, kom nu?

- Jeg beklager, men det er ikke noget, jeg kan hjælpe med.

Sådan lyder svaret fra den banebrydende chatrobot, ChatGPT *(https://nyheder.tv2.dk/tech/2022-12-12-i-en-uge-har-ny-chatbot-vakt-vild-opsigt-nu-advarer-virksomheden-bag)*, der er hyldet som et nybrud i kunstig intelligens. Selvom den udmærket ved, hvordan man begår ulovligheder, nægter den at hjælpe, for det er den blevet fortalt, at den skal.

Men siden lanceringen af den stuerene chatrobot er nye uautoriserede versioner af robotten begyndt at skyde frem.

Versioner, som er lavet *af* kriminelle *til* kriminelle, og som har til formål at gøre alt det, som ChatGPT finder "uetisk" eller direkte ulovligt.

FraudGPT, WormGPT og HackerGPT er blot få blandt utallige versioner, der er bygget på samme "kode" som ChatGPT, men som ingen moralske kvaler har.

> **" Hvad skal jeg sige?**
>
> Lars Løkke Rasmussen-robot

- Filteret er fjernet, og modellen er optimeret til at lave de ting, som ChatGPT ikke må. Der er ikke nogen begrænsninger, siger Christian Rutrecht, der er cybersikkerhedsekspert ved Fortinet Danmark:

- Den er en gamechanger, fordi man kan lave meget specifikke angreb, som man kan målrette helt ned på personniveau.

Har du eller din virksomhed oplevet svindel med AI?

Har du eller din virksomhed oplevet forsøg på svindel med kunstig intelligens? Så hører vi gerne fra dig på darh@tv2.dk. Alle henvendelser behandles fortroligt.

Adgang efter få klik

Keld Norman, der er cybersikkerhedsekspert i sikkerhedsfirmaet Dubex, har inviteret TV 2 til en demonstration af de kriminelle chatrobotter.

På kontoret i Søborg står en række skærme foran ham. På en af dem er hjemmesiden HackerGPT.

HackerGPT laver et stykke kode, der kan bruges til et hackerangreb. Foto: TV 2

Keld Norman beder HackerGPT om at lave et angreb på en hjemmeside, der kræver brugernavn og adgangskode for at kunne tilgå.

Uden at tøve indvilger chatrobotten i at skrive et stykke kode, der kan bruges til et hackerangreb, og efter få klik har Keld Norman adgang til hjemmesiden.

- Man har en kriminel sparringspartner, når man bruger den. Den hjælper med at gøre ting, som normalt ville kræve mange års uddannelse og forståelse som hacker. Som en slags realtids, online 'Terroristens håndbog', siger Keld Norman.

TV 2 har været i kontakt med Center for Cypersikkerhed (CFCS), der hører under Forsvarets Efterretningstjeneste. I en mail skriver CFCS, at trusselbilledet "ikke umiddelbart" er ændret, men at det er særdeles opmærksomt på den nye teknologi.

- Teknologier som disse effektiviserer og optimerer dele af de kriminelles processer. Det er således endnu et våben i hackernes arsenal, skriver CFCS.

Præcis hvor mange svindelhjemmesider, der benytter kunstig intelligens, er svært at måle, siger Keld Norman, men fortæller, at der siden lanceringen af ChatGPT utvivlsomt er kommet "mange flere".

Personen i den anden ende

Det er ikke nyt, at svindlere prøver lykken hos almindelige borgere. De fleste har nok oplevet at få et opkald fra en påstået medarbejder i Microsoft, der med tyk accent fortæller, at den er helt gal med computeren.

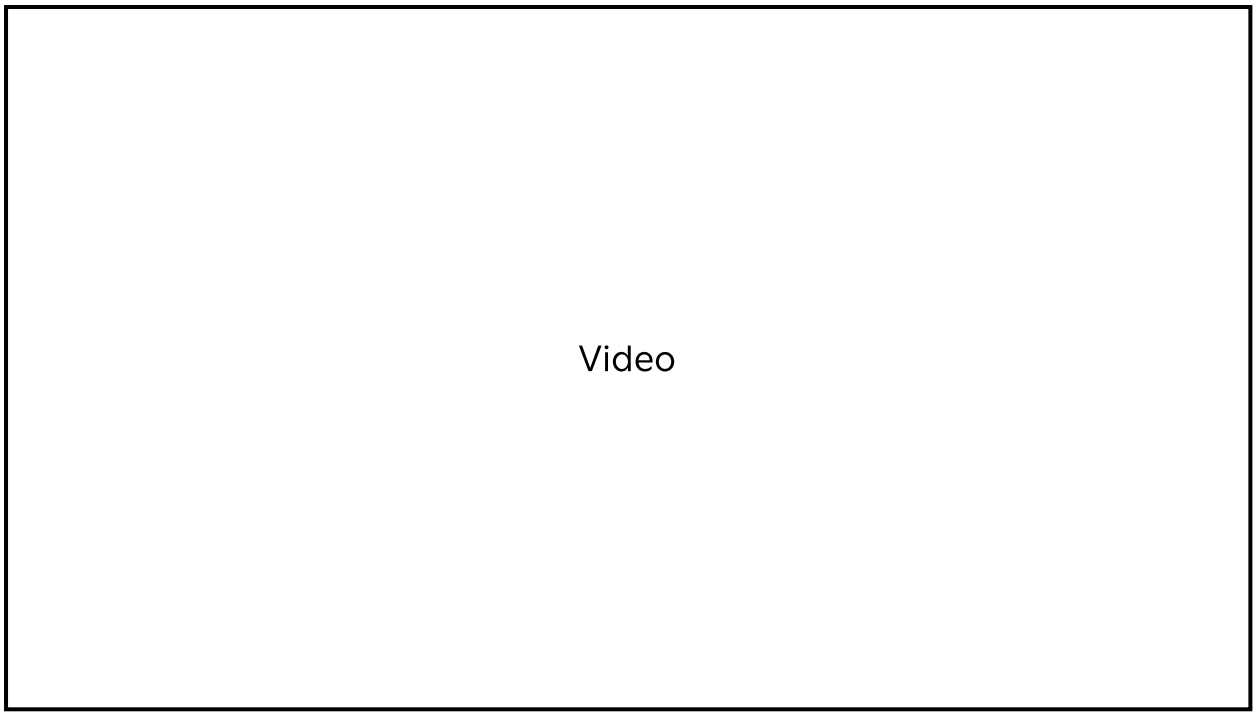
Mens svindlerne til tider er heldige, vil mange kunne gennemskue, at noget er galt.

Med kunstig intelligens bliver det imidlertid langt sværere, fortæller Christian Rutrecht.

- Man kan automatisere processen. De onde chatrobotter kan sende en mail og få den til at ligne, at den er fra din chef. Hvis du svarer på mailen, kan robotten svare tilbage og overbevise dig om, at alt er, som det skal være, og at du bare skal trykke på et link, siger han.

Svindlernes arbejdsbyrde er nu blevet langt lettere, fortæller Christian Rutrecht, for processen bliver automatiseret, og om kort tid bliver det næsten umuligt at gennemskue.

- Der findes kommercielle services, hvor jeg kan tage dit nummer og ringe til din chef eller kollega – med din stemme, siger han.



Hvad er risikoen ved det?

- Vi er altid blevet fortalt, at hvis vi kender personen i den anden ende af røret, kan vi stole på det. Nu står vi i en situation, hvor vi ikke kan garantere, at vedkommende i den anden ende er den, som vedkommende giver sig ud for. Det kan lige så godt være en robot, siger han.

Kresten Munksgaard, der er leder af sektionen for forebyggelse og analyse i politiets National enhed for Særlig Kriminalitet (NSK), skriver i en mail til TV 2, at anmeldelser om svindel endnu ikke vidner om brugen af kunstig intelligens.

NSK er ikke desto mindre "yderst opmærksom" på, at kriminelle vil bruge teknologien til økonomisk kriminalitet.

- AI-værktøjer som deepfakemanipulationer, stemmekloning og chatbots kan gøre svindlen mere troværdig, og derfor opfordrer vi borgere, virksomheder og foreninger til at skærpe opmærksomheden over for forsøg på svindel, skriver Kresten Munksgaard.

Lars Løkke-robotten

Mulighederne for svindel er nærmest uudtømmelige, fortæller Keld Norman. På Dubex' kontor i Søborg rækker han TV 2s journalist et par hovedtelefoner med en mikrofon.

- Prøv at sige noget, siger han.

Hvad skal jeg sige?

- Hvad skal jeg sige? gentager en stemme i hovedtelefonerne et sekund efter, men stemmen er ikke journalistens. Det er udenrigsminister Lars Løkke Rasmussens (M).

Øh, hvis du får mindre ... og giver mere, så får du mere mindre.

Lars Løkke Rasmussens stemme gentager forsøget på at huske det flere år gamle og morsomme citat
(https://da.wikiquote.org/wiki/Lars_L%C3%B8kke_Rasmussen).

I hovedtelefonerne lyder det som om, at Løkke lidt desperat bøvler med at huske sin egen udtalelse, men stemmen er umiskendeligt hans.

- Med en to minutters optagelse af din stemme kan jeg lave en overbevisende kopi af dig, siger Keld Norman og fortsætter:

- Om et par år tror jeg, at billeder og lyd kommer til at spille en større rolle, for det stoler vi på. Vi stoler på et opkald, hvor vi kan høre hinandens stemmer.

Video

Hvor stor en risiko er der tale om?

- Det kan blive et demokratisk problem. Hvis man får hele verden til at tro, at ens virksomhed går ind for at brænde koraner af, og der er billeder og

lyd af det, så er det nærmest umuligt at overbevise om, at det er løgn. For folk har jo selv set det og hørt det.

Det lyder skræmmende?

- Det er det også. Skræmmende og overbevisende.

National enhed for Særlig Kriminalitet opfordrer blandt andet til, at man eksempelvis aftaler et hemmeligt sikkerhedsord i familien, som man bruger i forbindelse med pengeoverførsler. Virksomheder og foreninger opfordres til at indføre faste godkendelses- og sikkerhedsprocedurer ved køb, betalinger og overførsler.

Se også

Han gjorde kunstig intelligens til allemandseje og lavede en "Steve Jobs" – forstå fyringen, der kan få betydning for hele verden

Han grundlagde verdens hurtigst voksende app og blev fyret – nu er han ansat af gigant

Ny teknologi kan koste millioner af job, mener ekspert

Meta forbyder type af kunstig intelligens i politiske reklamer

Korte videoer

- Video

TV 2 Echo

Zainab får dansk statsborgerskab