AI og kvantecomputere truer dit digitale liv

Hvis ikke vi omkoder kritiske dele af internettet indenfor de næste tre år, bliver det et meget usikkert sted at være, fortæller forsker i cybersikkerhed Bas Spitters. Heldigvis er AI også en del af løsningen.

■ Jeppe Kyhne Knudsen, Aktuel Naturvidenskab.

BAS SPITTERS

Bas Spitters er lektor på Institut for Computer Science, Aarhus Universitet. Han forsker i metoder til at implementere nye kryptografiske teknikker.

Han er født i Holland og kom til Aarhus i 2014. Han er 51 år gammel.

spitters@cs.au.dk

Den 7. april i år kom AI-virksomheden Anthropic med en overraskende udmelding. De havde udviklet en ny kunstig intelligens, Mythos, som er i stand til at finde sikkerhedsbrister i mange af de tjenester, vi til daglig benytter os af på internettet.

Eksempelvis havde den kunstige intelligens helt af sig selv fundet et hul i internetbrowseren Firefox, som har mere end 150 millioner brugere på verdensplan. En brist, som hackere vil kunne bruge til at kompromittere vores allesammens it-sikkerhed.

Men det var ikke det vildeste.
Udviklerne fra Anthropic placerede Mythos i et lukket miljø, en såkaldt sandkasse, hvor den fik lov at lege. Uden at den kunstige intelligens vidste det, havde forskerne gemt en lille åbning ud af sandkassen, og det hul fandt Mythos lynhurtigt og var ude på internettet.

Udviklerne blev så bekymrede for Mythos' egenskaber, at de besluttede sig for ikke at udgive den. I stedet samlede de en række store virksomheder, som alle står for dele af internettets kritiske infrastruktur. Mythos havde hos flere af dem fundet sikkerhedsbrister, og Anthropic gav dem derfor adgang til den kunstige intelligens, så de kunne lukke så mange sikkerhedshuller som muligt.

Om Anthropic kommer til at offentliggøre Mythos er stadig et åbent spørgsmål. Men som journalist Henrik Moltke udtalte i DR-programmet Prompt den 16. april, fungerer hele historien også som god reklame for Anthropic. Han mener dog, at de er oprigtigt bekymrede.

Det samme er lektor Bas Spitters, der forsker i cybersikkerhed på Institut for Computer Science på Aarhus Universitet. Her arbejder han blandt andet med at implementere nye teknologier til at beskytte mod hackerangreb, som kan bryde krypteringen af vores mest personlige oplysninger.

Ifølge ham er det en stakket frist at tilbageholde Mythos:
»Kineserne er som regel tre måneder bagud. Der går derfor ikke lang tid, før de har en model, der kan det samme som Mythos. Forskellen er nok bare, at de udgiver den som open source – og så har alle hackere i verden pludseligt et nyt kraftfuldt værktøj.«

Internettet skal omkodes

Bas Spitters er i det hele taget bekymret for fremtiden. For når kunstige intelligenser bliver gode til at lede efter sikkerhedshuller, bliver det ekstremt svært at gøre vores færden på internettet sikker.

»Den måde internettet er kodet på i dag, gør det næsten umuligt at beskytte os mod de kunstige intelligenser. I stedet er vi nødt til at omkode hele internettet,« siger han.

Lige nu hersker der, ifølge Bas Spitters, et våbenkapløb mellem de store it-virksomheder og hackerne. Virksomhederne kæmper for at få lukket så mange huller som muligt, før hackerne med de nye værktøjer kan finde dem.

Men flere af dem er også ved at omkode deres programmer, fortæller han.
»Den amerikanske regering er bekymret for denne udvikling. DARPA, den militære forskningsafdeling i USA, har foreslået, at hele den amerikanske it-sikkerhed skal omkodes i programmeringssproget RUST – og det giver rigtig god mening.«

RUST er i modsætning til C, som er et af de grundlæggende programmeringssprog i dag, skrevet på en måde, så de enkelte programmer eller kodebaser ikke deler hukommelse med hinanden.

Det lyder lidt kryptisk, men det betyder grundlæggende, at programmet er mere blottet for angreb, fordi det har flere kontaktpunkter med omverdenen. »Når vi programmerer i RUST, er programmerne derimod isolerede. De kører marginalt langsommere, men ikke noget brugeren kan mærke. For at sikre os i fremtiden, er vi nødt til at omskrive til RUST eller andre sikre programmeringssprog,« siger Bas Spitters.

Om tre år er vi på den

Når hackere angriber store virksomheder, sker det oftest ved, at de finder disse huller i programmerne inde i computerens fælles hukommelse. Men der findes også andre metoder som at bryde den kryptering, der skal beskytte os mod, at andre læser med i vores beskeder og e-mails.

Og her lurer endnu en stor trussel – dog lidt længere ude i fremtiden, fortæller Bas Spitters.
»Lige nu arbejdes der intenst på kvantecomputere rundt omkring i verden. Indenfor tre år forventer Google og Cloudflare, at kvantecomputere kan bryde selv den dybeste kryptering – og det er endnu et stort problem for cybersikkerheden,« siger han.

Siden den første kvantecomputer blev bygget i 1998, er udviklingen støt og roligt gået mod mere og mere kraftfulde computere. Den første bestod af to qubits, svarende til to bits i en moderne computer. En meget begrænset regnekraft. Men hvor en enkelt bit i en computer enten kan vise 0 eller 1, kan qubits vise 0 og 1, men også begge dele på én gang.

»Det betyder, at hver qubit kan give langt mere regnekraft end en enkelt bit. Kvantecomputere er især gode til at bryde kryptering, som det tager almindelige computere meget lang tid at bryde,« siger han.
Google har i dag en kvantecomputer kaldet Willow med 105 qubits. Den kan, ifølge Google selv, lave en beregning på fem minutter, som det vil tage en almindelig supercomputer 10 kvadrilliarder (10^{24}) år at lave.

»Nogle af mine tætte kolleger har udviklet en ny form for kryptering, som kvantecomputere har svært ved at bryde. Teknologien skal dog, ligesom omskrivningen af internettets kode, implementeres på sikker måde. Det tager tid, og vi har travlt,« siger Bas Spitters.

Vi har ekstremt travlt

Lige nu befinder verdens it-sikkerhedseksperter sig i et kapløb med tiden. Spørgsmålet er, om vi når i mål, før nye kunstige intelligenser og kvantecomputere bliver tilgængelige.

Bas Spitters er i tvivl. Eksempelvis fyldte cybersikkerhed og kunstig intelligens ingenting i forårets valgkamp.
»Det er ikke mit indtryk, at politikerne er særligt opmærksomme på det her i Danmark. Det er noget andet i USA. Men for at vi skal nå det, kræver det, at vi sætter ressourcer af til det. Ellers står vi meget snart på et meget usikkert sted med vores fælles kritiske it-infrastruktur i Danmark,« siger han.

Han mener, at vi som borgere bør være enormt bekymrede for denne udvikling. Og at vi bør udtrykke vores bekymring til politikerne, så de bliver tvunget til at tage affære.

»Om lidt kommer der nye modeller, som kan det samme som Mythos – og så er vi på den. Vi går en usikker tid i møde på internettet,« siger Bas Spitters. ■

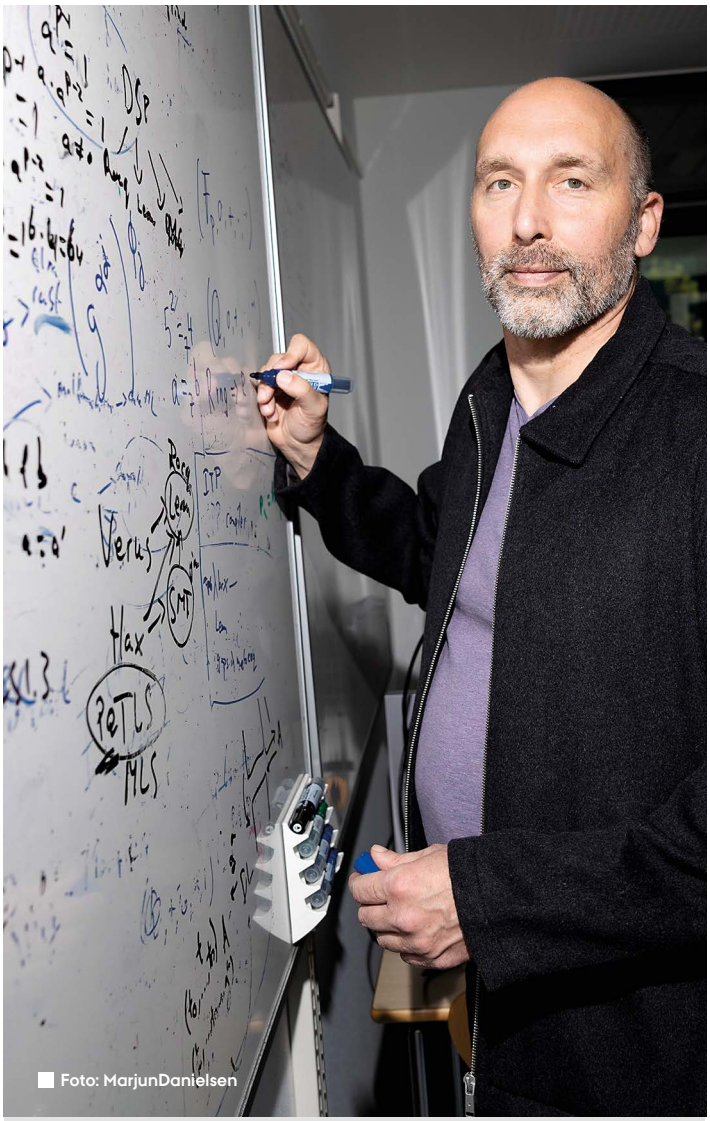

Foto: MarjunDanielsen

EN TRUSSEL MOD INTERNETTET

Mythos er en kunstig intelligens, som er udviklet af det amerikanske softwarefirma Anthropic. Mythos er endnu ikke officielt udgivet, og kun få samarbejdspartnere har adgang til den.

Anthropic frygter, at Mythos vil gøre det for billigt, nemt og hurtigt for hackere at finde sikkerhedsbrister i vores kritiske infrastruktur på internettet. Derfor har de besluttet sig for ikke at udgive den for nuværende.

Mythos er den seneste i en række af AI-modeller, som Anthropic har udviklet. De er nok mest kendt for deres model Claude, som det amerikanske militær blandt andet bruger.