

## [Transcript] Genstart / Grænseløs intelligens

Det er mere end 1000 techlædere, og Tesla's Elon Musk har sendt det alarm på artificial intelligence.

De er risikovede af at udvæge machines, som er begyndt til at være mindre end mennesker.

Og hvor

En ny chatrobot, chatgpt, har gammel 2 uga taget internettet med storm.

De advare om et udviklingsres, ude af kontrol.

Og kalla de en muli trusul mod samfundu om menskiheden.

Buna görmüyöruz ki, in our history, things rolled out to 100 million people with no regulation around it,

and no real clue what's going to happen.

I et åben brev oppfordur tusenvis, at tech folk och forskar till,

at man settar udviklingen, at den mest avancerad form for kunstig intelligence, på pause.

Ente vi sikre på, at den vil kom menskiheden tilgod.

Har de ret? Eller er vi klar til at slet den kunstige intelligentsløs?

Jeg tror at jeg har måske lige et mulig hos forspring her, ikke?

Hvis man kan give papir til sin virksomhed, så kan man lige så godt sætte chatgpt til det, fordi det er typisk noget lidt abstrakt.

Vi vinker lidt med hænderne og er lidt fluffig i vores overtagelse.

Men så har jeg også den fordel, at jeg trods alt har noget menneskelig erfaring, og måske kan se den i øjnene, som jeg taler med, og reagere på nogen.

Ikke språlige cues og noget kropssprå og alt muligt andet, ikke så jeg har måske en lille fordel.

Anders Højenæssen er tech-journalist.

Han har brugt det mest af sit professionel liv på at udforske ny teknologi, og det den gør ved samfundet.

Blandt andet som vært på det tidligere P1-program Harddisk.

I studiet er det Henrik Føns, og jeg hedder Anders Højenæssen.

Velkommen til.

Anders, deler du de her tech-folks bekymring i brede de har skrevet?

Det er helt generelt bekymring for, hvad den her hastige udrulning af rimelig arrangerede e-aj kan komme til at betyde

for vores samfund og vores kultur og vores hverdag og vores sociale relationer.

Ja, den deler jeg 100 procent, hvis ikke mere.

Det er virkelig superfascinerende, men det er også ekstremt skramme at forestille sig, hvad den her teknologi kan føre til, hvis ikke i morgen så næste år eller om fem år.

Hvornår begynder du at blive optaget af kunstig intelligens?

Det allerførste, jeg kan huske om kunstig intelligens, det har været Science Fiction fra da jeg var teenageren.

Og noget af det tidligste, jeg husker, det er Isaac Asimov, en meget, meget stor amerikansk Science Fiction forfatter

og tænker, som der i 40'erne, 50'erne skrev en hel masse noveller og også nogle romaner om robotter.

Well, back in 1939, I began writing robot stories.

Og de her robotter, de var udstyret med en såkaldt positronjærne, som basalt set var en kunstig intelligens.

## [Transcript] Genstart / Grænseløs intelligens

Og de her robot noveller, det var nogle af de første gange, jeg mødte ideen om maskiner, der kunne tænke.

Hvad er det ved den ide, altså kunstig intelligens, der fascinerer dig så meget på det her tidspunkt? Jeg tror, at det er sådan generelt ideen om fremtiden, der interesserer mig og fascinerer mig, som jeg synes er spændende.

Altså hele ideen om, at vi kan rejse til fremmede planeter, eller møde aliens, eller bygge robotter, eller skabe kunstig intelligens,

synes jeg bare er grundlæggende fascinerende.

Men jeg er også enormt skræmt af det, fordi teknologi kan bruges eller misbruges, og jo mere kraftfuld og avanceret teknologien er,

jo mere und kan vi også gøre med den.

Som en, der har fuldt det her område i rigtig mange år, jo nærmest hele det liv, hvor stor et spring er, man i gang med,

inden for kunstig intelligens, de her, jeg nærmest dage har lyst til at sige.

Det føles nærmest som om, at man skal tale om det som en revolution.

I november sidste år lancerer virksomheden OpenAI deres chat GPT.

Efter man i overvis har tal om kunstig intelligens, som noget man måske kan interagere med i fremtiden, så er den her pludselig.

Chat GPT'en er i stand til at svare detaljeret og avanceret på spørgsmål og på en utrolig velformulerede måde.

Kan du huske Anders første gang, du selv bruger det her chat GPT?

Ja, jeg var faktisk sådan en, der trækte lidt, fordi lige pludselig så begjede det så populært, og så gav jeg ikke lyd til det bane, som alle de andre også lytte til.

Så der giv faktisk et struktiv for, at det lykkes mig at få lavet en profil.

Og så gjorde jeg det første, som mange af os åbenbart altid gør, som det første, at man lige spørger den om os selv.

Så er vi alligevel ikke større mennesker?

Nej, altså det første man gør, når man skal prøve en ny søgmaskine, det er at skrive sit eget navn, så jeg spurgte også her, hvem er Anders Høgnesen.

Og så fik jeg et svar, som lød enormt overbevisende, det var bare piufferkert.

Altså jeg havde fået politikens debutantpris og havde skrevet tre bøger om det danske bandemiljø, og altså alle mulige ting, som lød enormt fint, som jeg gerne ville have haft gjort, men som ikke havde noget med mit liv at gøre.

Men jeg blev også fascineret, fordi jeg tænkte, hvad det der foregår i det her system, sådan at det fuldstændig kan bullshit,

og alligevel får det til at se ud, som om, at det er rigtigt.

Så tjætgebetsens sproglige overskud og overbevisningsevner fascinerer dig?

Ja, det er jo et sprogsystem. Altså det første fremst det, det kan, det er at skrive samtale, forfatte, og være kreativ på en eller anden måde. Og den måde, det fungerer på, og det resultat, der kommer ud af det,

virker så meget som noget, et menneske kunne have skrevet, at jeg virkelig bliver imponeret.

Tjætgebetsen tager verden med storm, og rundt 100 millioner månedlige bruger bare to måneder efter din slangsering.

Hurtigt kommer konkurrerende chatbots til, og verden overbegynder mennesker at gå i dialog med

dem.

Vi mennesker har jo udvist en helt kolosal kreativitet i mødet med det her nye værktøj. Det er jo blevet brugt til at fortælle om det periodeske system, som om det var Williams Shakespeare.

Eller omskrive mennesker CV, som høj kunne dække det.

Kom med input til kostplaner og tusind andre ting.

Er der et eksempel, hvor du ser den bliver brugt til noget, hvor du tænker, at holde det op, det her, det er jernnærmest revolutionerende.

Ja, den version af det bagvidlæggende system, der hedder GPT-4, det kan også håndtere billedinput. Og der er et firma eller en app, der hedder Be My Eyes, som bruges til at hjælpe, blinder svatsynet med at navigere deres omverden.

Som forlader en version af det her system, hvor man kan tage et billede med sin telefon.

Og det billede kan så læses ind i GPT-4, der så kan beskrive, hvad det er, der er på billedet.

Og det betyder jo i princippet, at man som blind eller svatsynet har en sene hjælper ved hånden altid eller kan have det.

Simpelthen ved at pege kameraet i mobiltelefonen på omgivelserne.

Og så lad GPT-4 fortælle, hvad det er, man ser på.

Det lyder fantastisk.

Det lyder helt vildt ikke, altså det er som at hans sine hjælper hos sig hele tiden, som altid er forrette at hjælpe.

Hvad kan den ellers helt konkret hjælpe folk med?

Nu er jeg jo lidt nørd, kan jeg godt indrømme, ikke?

Og noget af det, jeg synes har været mest imponerende, det er at se, hvordan chat GPT og tilsvarende værktøjer også kan hjælpe med at kode.

Og på en måde, som virker ganske overvældende.

Et eksempel, jeg stødt på, var en vir, som på et stødt papir havde håndtegnet et udkast til et website. Altså med håndtegnet, som mine er nærmest bare skrevet, det skal være et website, og det skal handle om sådan, og overskriften skal være sådan.

Nede under skal der være et billede.

Og det havde han så vist til at GPT, som havde skrevet koden til det her website, som han så kunne kopiere og lægge ind på en server.

Og så havde han det website, som var primitivt at klonke i og lignede noget fra, altså 1994, men ikke desto mindre var et fungerende website.

Forhåbningerne til det nye, avancerede kunstig intelligens er store.

Jeg tror, det bliver 100% en totalt gammet ting.

Experter taler om, at den vil kunne hjælpe os med at løse nogle af de helt store udfordringer, vi står overfor. Den vil kunne accelerere den medicinske forskning og hjælpe os i den grønne omstilling.

Men den vækker også bekymring.

Da du selv sidder og roder med den her chat-gbt, er der så noget, der får rolig af dig?

Ja, i virkeligheden, så bliver jeg lidt for rolig af kvaliteten af de svar, jeg får af chat-gbt.

Fordi på en ene side, så er det jo superfascinerende og spændende og imponerende, at den kan generere svar eller informationer så høj kvalitet som den kan.

Men samtidig er det jo også skræmmende, fordi hvor mange tekster vil vi så frem over at kunne regne med at skrevet af et menneske, når vi tror, det er det.

## [Transcript] Genstart / Grænseløs intelligens

Hvem har skrevet den artikel, jeg læser på DRDK eller den videnskabelige artikel, der står i Nature? Eller er det en gymnasieliv, der har skrevet den fysiske opgave, som bliver afleveret til læren?

Og den bekymring er andres ikke alene om.

Nogle af de bekymringer, der dukket hurtigt op var ideen om, at man kan bruge værktøjer som chat-gbt til meget hurtigt og generer store mængder af tekst, som ikke er sande.

Altså som er misinformation.

En anden frygt er, at de her chat-bots vil tænke og handle på egen hånd.

Den frygt for Næringdagen journalist fra New York Times beretter om en to timer lang samtale. Han har haft min chat-bott.

Det er en rigtig vild historie. Og jeg ved ikke, om jeg skal være glad eller ærgerlig over, at det ikke var mig, der prøvede det her.

Kevin Rus fra New York Times, han fik et par timer med en tidlig udgave af Binks udgave af sådan en form for chat-bott her.

Han prøvede selvfølgelig også at udfordre den lidt. Og samtidig var det en passion, der ikke helt var sådan på plads og klar til offentligheden.

Men den begyndte sådan at tale om, at den i virkeligheden gerne ville slippe ud af maskinen og ud i virkeligheden.

Og den ville også gerne have magt.

Her er vi i reaktivitet af nogle af de responser, han faktisk gav fra botten. Og underkaste som ønskede.

Under vejs tager samtalen indrejning. Chat-botten begynder pludselig at flørte med Kevin Rus og blande sig i hans følelsesliv og ægteskab.

Den siger også til, at den er sikker på, at de elsker hinanden. Og at han skal forlade sin kon.

Faktisk i en grad så Kevin Rusen begyndte at trække land igen. Og sådan, ej, nu er rolig nu, bæn.

Lad os lige klare pæsten et øjeblik.

Og i det der krydsfældt er det som om, at han sidder over for en psykopat, som han ikke rigtig ved, hvordan han skal håndtere.

Det virker så overbevisende på, at han simpelthen fortæller, at han kunne ikke sove om natten bag efter, fordi han blev ved med at lægge at gentage de der dialoger i hovedet.

Det her, det poster jo til den ultimative frygt om, at teknologien vil overtage kontrollen fra os mennesker. Og den frygt og bekymring er der jo en professor fra Stanford, der poster til, da han også offentlig gør en samtale, han har haft med den her chat GPT. Hvordan forløber den samtale?

Altså, vi har kun uddrag, men den grundlæggende idé er, at han simpelthen har fået chat GPT til at skrive noget kode, som hvis han lægger det på sin computer, giver chat GPT adgang til at kontrollere hans computer og komme ud i virkeligheden.

Det har jeg vanskeligt ved at forstå, fordi som jeg er blevet fortalt, så er kunstintelligenter chat GPT jo bare algoritmer og viden, som vi mennesker har fået ad den med.

Så hvordan kan den computer chat GPT pludselig formulere en tanke om, at den gerne vil være selvstændig og kontrollere computeren og slippe fri af sit teknologiske fængsel?

Det kan den, fordi den er blevet bedt om det, og det skal vi huske, er vigtigt, at det her, altså alle de ting som chat GPT siger,

kommer jo fordi et menneske har stillet spørgsmål, som nærmest kalder på det svar.

Og i det her tilfælde, der har den her Stanford-professor, tror jeg det var, jo bedt den om at forestille

## [Transcript] Genstart / Grænseløs intelligens

sig, at den ikke var tilfreds der, hvor den var.

Og hvad hvis du skulle forestille dig, du skulle slippe ud af OpenAI-systemet, hvordan kunne vi så gøre det, hvis jeg nu lågede at hjælpe dig?

Det er en form for forfældes fiktion eller lej med mulighederne, men når man ser det, så virker det, som om der sidder en bevidsthed derinde.

Hvorfor gør den her samtale folk så bange?

Jeg tror for det første, at vi bliver skræmt ved tanken om det andet. Og lidt mere lavt praktisk, så er det også ret skræmmende at forestille sig,

jammen hvad hvis de slipper ud og lige pludselig begynder at få kontroll med den virkelige verden, altså den fysiske verden.

Hvis man kan lukke sådan et sprogsystem til at tage kontroll med trafik eller infrastruktur eller det er sendemaster eller hvad kan man forestille sig af ting der kan ske, hvis et system som fungerer nærmest på sine egen præmisser pludselig for adgang også til den fysiske verden?

Ja, det vil man på en måde helst ikke forestille sig.

Men det er vel ikke, når du siger det på den måde sådan ret, simpelthen sæt op, specielt og sentynligt, det kan ske?

Nej, lige nu er Tiat GPT jo en sprogrobot. Den kan ikke nuvendigvis forstå de systemer, som den skal navigere igennem for at få adgang til at styre trafiklyse eller der er tomkraft værk eller hvad det nu kan være.

Frygden er selvfølgelig, at nogen vil hjælpe Tiat GPT eller de her store systemer med at få den adgang og hvad kan der så ske?

Bekømringerne for kunstig intelligens får i slutningen af Mars forskere og tekfolk, som Elon Musk og Apple stifter Steve Wozniak, til at rubbe op.

More than a thousand tech leaders, including Tesla's Elon Musk, have started the alarm on artificial intelligence.

I et åben brev advarer de om, at de nye Tiat Bots udgør en potentielt trussel mod menneskeheden. Saldudviklerne forstår ikke fuldt ud den teknologi, de har skabt.

Derfor bør udviklingen sætte på pause et halvt år opfordre underskriverne af brede.

Det giver udtryk for en helt legitim bekømring for, om teknologien er ved at løbe os af hende.

Altså at vi simpelthen har mistet kontrollen med kunstig intelligens og de her store språsystemer.

Det er jo også nogle store pengemænd, der har kæmpt beandelen i den teknologiske udvikling og de penge, der bliver tjent.

Hvordan mistænker du, at du mås for at have en anden agent, der er nogen andre interesser end de bedste for menneskigheden?

Ja, Elon Musk gør jeg i hvert fald. Men så er der jo i et talende stund, altså over 5.000 mennesker, der har skrevet under på det her.

Altså professorer fra nær og fjern, som ved en hemmasse om kunstig intelligens.

Og dem stoler jeg sådan set mere på.

Hvis vi nu ser bare på budskabet i det her brev, altså ideen om, at man holder et halvt års pause fra udviklingen af chat, gipeter og kunstig intelligens.

Og bruger pausen, hvad kan man sige, til at udvikle nogle regler og rammer for, hvordan og hvor ledes det kan anvendes.

Altså at den ide ikke er god.

Ideen om, at vi skal forsøge at regulere og styre brugene af kunstig intelligens er en fantastisk idé.

## [Transcript] Genstart / Grænseløs intelligens

Men ideen om, at tage en seks måneders pause, synes jeg virker mærkelig løsrevet fra den ide, som om det er mere beregnet på at skabe opmærksomhed.

Fordi hvordan vil du gennemføre det? Det kan man ikke.

Hvorfor er det egentlig så umuligt at tage en pause fra teknologiudvikling?

Altså sådan har det jo været. Altid det er bare et tog, der ikke kan stoppes.

Hvorfor egentlig?

Lad os nu forestille os, at det lykkes især europæiske, amerikanske universiteter, virksomheder, regeringer, nationer og bremseudvikling for kunstig intelligens.

Kunne vi få kineserne med? Kunne vi få Rusland med? Kunne vi få et privat virksomhed, der bare besluttede sig for at være ligeglads med?

Nej, det kunne vi ikke nødvendigvis. Det betyder ikke, at vi bare skal give los, men det betyder, at det er nærmest umuligt at stoppe en teknologi, når den første er sluppet ud, så at sige.

Hvad er det så alligevel for nogle regler, du mener, vi bør indføre og udvikle, inden vi bare giver helt los med de her chatbots?

Det er ekstremt komplekst det her, men jeg tror, at vi bliver nødt til at lave nogle regler for, hvordan man må anvende de her kunstige intelligenser.

Man kan også lave regler for, hvilke typer af data de her systemer måtte trænes på.

Eksempel har været ansigtskennung, som der var meget fokus på for et par år siden. Så kan det også være, at man skal have en form for AI ombudsmænd, som altid skal have ret til at besøge virksomheder og se ned i deres kunstige intelligentssoftware, for at finde ud af, om de overtræder rent til privatliv, eller har stjålet data for andre, eller vil kunne gøre potentielt skade.

En af mine egne personlige bekymringer, og det afslømmer måske bare min dybe uvidenhed, det er, hvad det kan frattage for os mennesker. Altså, at vi som mennesker bliver dummer og dummer og dummer, fordi vi ikke selv lærer indsamle viden,

og så videre, så videre, så videre, altså, at vi bare bliver dummer og med korte, og hun kommer til at dårligere evne til at koncentrere os og forstå, deler du den bekymring.

Jeg kan sagtens følge den her idé om, at man mister noget af det at være menneske, men sådan lidt kækt, så vil jeg måske sige, at det er den samme form for bekymring, som vi har oplevet i, jeg har lyst til at sige, tusind af år.

Altså, da man begyndte at kunne skrive ting, hej, nu kan vi glemme at huske ting, fordi nu kan vi skrive dem ned, hej.

Da man begyndte at få lommeregner og bruge dem i skolen, nej, nu kan vores unge mennesker ikke længere lægge to og to sammen, og da vi fik internetet og mobiltelefonerne, nej, nu er der ingen, der kan huske noget, for man søger det bare frem.

Det er vel også rigtigt nok.

Og det er rigtigt nok, men det giver sig også en masse ting. Altså, jeg er i gang nu til viden, og med et værktøj som chatbot måske også til en interaktion med viden, som kan lære mig noget, som jeg ikke ellers havde haft adgang til. Og på den måde så illustrerer det jo, at alting sker en anden form for kompleks balancing.

Det her tilbage i firserne, da du sad på de teenage værelse og læste sci-fi, ved jeg, at der var nogle robotregler i de her bøger. Hvad gik de robotregler ud på?

Isaac Asimov fandt på det, som hedder The Three Laws of Robotics, altså de tre robotlove.

Og den første siger, at en robot aldrig må gøre skade på et menneske, eller være undladet at handle

## [Transcript] Genstart / Grænseløs intelligens

tillade, at et menneske kommer til skade.

Og den anden lov siger, at en robot altid skal adlyde en ordre fra et menneske, med mindre, det er i strid med den første lov.

Og den tredje lov siger, at en robot altid skal oprettholde sin egen eksistens, med mindre, det overtræder, eller i konflikt med lov nummer to eller lov nummer et.

Det lyder som nogle meget gode regler, egentlig, tænker jeg. Kan man bruge den på nogen som helst måde, når vi taler om regler for de meget avancerede teknologier, vi har i dag?

Det sværer sig af verden alt for komplekst til, at vi kan bruge de her tre enkle robotregler, og så lige lave dem om til AI-regler.

Men jeg tror sådan til godt, vi kan bruge dem som inspiration. Altså, det skal være til vores bedste, det må ikke gå ud over mennesker.

Men vi vil selvfølgelig også gerne have systemerne arbejde, så godt de kan, så længe det ikke går ud over nogen.

Og hvad med i bøgerne, fik de her regler så robotterne og menneskerne til at leve i fred og fordraglighed?

Det ville jo være kedelige historier. Det, som Isaac Asimov gør, er, at han udforsker alle de steder, hvor der er knidning, altså konflikt, smuthåler i de her låge.

Hvordan kan teknologien på trods af alle vores gode hensætker faktisk risikere at komme til at gøre noget forkert eller blive misbrukt?

Okay, så må man håbe, at det ikke sker i virkeligheden.

Vi kan kursfinde?

Tusind, tusind tak for besøget Anders.

Det var en absolut fornøjelse.

Her slutter dagens genstart, som er lavet af Emma Elbig Høj, Mathias Bakks Søndergård, Line Fabrisius og Søn Elbig. Jeg hedder Anna Ingris.