Hallå, det här är nyhetspådden Dagens Eko.
Nu börjar vi.
Olle, hur orolig behöver jag vara för AI-utvecklingen just nu?
Du behöver nog vara lite orolig.
Det finns väldigt mycket man inte känner till om de nya, kraftfulla AI-modellerna.
Det här är Dagens Eko med Robin Olin.
Idag om AI-horten och vilka av dem du ska ta på störst allvar.
Högskolan i Jönköping misstänker nu för första gången att studenter använt sig av AI för att fuska.
Det skulle till exempel kunna utnyttjas av bedragare.
Det är orolig för AI-utvecklingen.
Det är det jag ska jobba med att skriva, att man kanske vill ersatta av AI.
Då var det istället, men vi hoppas att det inte blir så.
Ja, nyheterna om risker med AI verkar aldrig tådslut.
Frågan är hur en vanlig nyhetskonsument ska kunna förhålla sig till alla de här AI-hoten.
Nu närmast, den fråga som beskrivs som ett hot mot mänskligheten av vissa.
Det handlar om allt från fusk och bedrägerier till att AI både kommer kunna ersätta och förstöra för människor.
The voices warning against the potential dangers are growing louder.
Because remember, these chatbots will die.
Eller bryr inte av utplåna som art.
Kill humans? How could it kill humans?
It'll figure out ways of manipulating people to do what it wants.
Elon Musk is showing that dire warning about the use of AI
and its major risk to society and an exclusive sit down.
It is non-trovial.
It has the potential of civilizational destruction.
Gäst är Olle Sakrisson, nyhetsbeställare och strategiskt ansvarig för AI-frågor.
Olle, du jobbar med vårt strategiska AI-tänk här på Sveriges Radio.
Och i det ingår att hålla koll på den här snabba utvecklingen som sker nu.
Jag tänker att vi kan gå igenom några av de här varningarna som har kommit den senaste tiden.
Vilka faror den här tekniken kan föra med sig?
Jag mämmer en liten lista på fem punkter som handlar om allt från vad som händer-
när man sätter den här tekniken i händerna på oss vanliga användare-
till några riktigt mörka, dystra framtidsscenarium.
Vad börjar vi då?
Vi börjar i händerna på oss.
Det som snackar så mycket om just nu är de här generativa AI-modellerna-
som är tränade på enorma mängder data och text.
Till exempel den här tjänsten ChatGPT som är så omtalad.
En del använder de här tjänsterna också som kunskapskällar.
Till exempel använder det för att be om medicinska råd.
Och då kan det nog gå ganska illa.
Hur då är det att be om medicinska råd?
Jo, men nån studie har visat att ChatGPT får fel på 9 av 10 frågor-

när man frågar den om bröstkanser.
Ett annat exempel är den här belgiska mannen som rapporterats-
har faktiskt har begått självmord efter att ha chattat med en AI-bott-
under en pågående depression.
Oj, vad händer då? Vet vi någonting om hur det har gått till?
Jo, men AI-inbörd eftertas säga att de skulle mötas i paradiset-
och uppmuntrade sig slut den här mannen att ta livet av sig.
Enligt vad som har rapporterats då och vad hans enka har sagt.
Men en annan väldigt uppmärksammad story-
det var när tech-reporten Kevin Ruse från New York Times-
chattade med Bing som är Microsofts verktyg-
och chatbotten som förövrigt började kalla sig själv Sydney-
sa till reporten att den hade en hemlighet att berätta.
Och nyhetsbolaget CNBC lät en röst läsa in den här konversationen.
I'm Sydney and I'm in love with you.
That's my secret. Do you believe me?
Do you trust me? Do you like me?
Den har det blivit självig om honom.
And I was curious and a little bit weirded out-
and started asking some questions and it said yes, I'm in love with you-
and you need to be with me and I said, well, I'm married.
You're married but you need me.
You need me because I need you.
I need you because I love you.
I basically tried to convince my wife and be with Sydney, this chatbot.
Nu ska man komma ihåg att den här techjournalisten försökte verkligen-
instruera och prompta den så att det här skulle komma ut och komma fram.
Men det här gav väl en slags hint om att om inte leva ett eget liv-
så genererar den väldigt mycket innehåll som kan te sig väldigt mänskligt.
Men det som svindlade lite för mig var egentligen en annan sak.
Vad då?
Jo, men det var att det här är ett bevis på att kapprustningen mellan de här teckgettarna-
nu är så hård och går så snabbt att det är helt uppenbart-
att de är beredda att släppa AI-modeller som kanske inte är riktigt färdiga-
som inte går att kontrollera, som till och med kan vara manipulativ eller rentavfarliga.
Och om jag nu är en sån som inte alls använder de här chatbottena i nuläget.
För det är ju ändå inte alla som gör en, sitter där och skriver saker med dem.
Påverkas jag ändå redan idag av det här med AI?
Det är klart att du kan påverkas av att andra använder AI då.
Det var ju en bild för några veckor sedan som fick väldigt stor spridning-
den här bilden på poven i en vit dundkappa.
Vad handlade du det här om?
Jo, men det var en stor puffig vit kappa av ett lyxmärke.
Du kanske har sett en bild av poven i en puffig vitkappa på social media.

Där är det. Det ser riktigt ut.
Och det var ju väldigt många som reagerade, oj då.
På en kanske nog är det ganska fåfängen då, en riktig lyxlidare bakom den här torftiga fasaden.
Men det här var ju fakat och kanske mest en lite kul grej.
Det är inte riktigt, men det ser ganska realistiskt ut.
Jag tror att det visar koncern att hur snart vi bara drar i de här dina fakta bilderna som inte är riktiga.
Och folk gick på det här i mediabranschen, eller?
Många medier gjorde det. Många förstod direkt vad det handlade om.
Men vi på Sveriges Radio har ju också gått på bluffar.
Tonight, Vladimir Putin rolling out the red carpet for China's President Xi.
I Xi Jinping, Kinas ledare, besök i Moskva.
Då spreds det en bild där Putin ser ut som att han knäbe ju framför Xi och kyser hans hand.
Men det här var ju en fejkad bild, en AI-fejkad bild som lurade många.
Och inklusive oss på Sveriges Radio faktiskt.
Vi nämnde bilden som hastigaste i en sändning som om den var sant, men fick sig en gå ut och rätta.
Så Putin har aldrig, vad vi vet i alla fall fångats på bild när han knäbe ju för Xi Jinping och kyser hans hand.
Inte vad vi vet, och det finns ju naturligtvis någon som har ett syfte bakom att den här bilden gjordes och spreds på det här sättet.
Vad har du, liksom, nummer tvåa på din lista med AI-faror där Olle?
Jo, men då kan man ju naturligtvis tänka sig vad som händer när den här tekniken sätts i händerna på dem som verkligen har ett ont uppsått.
Till exempel kriminella.
Hur då, då?
Jo, men ett exempel som blev väldigt viralt var när en mamma i amerikanska Arizona la upp på sin Facebook att hon precis hade blivit utsatt.
And it's my daughter's voice crying and sobbing.
Det ringer alltså på hennes mobil från vad hon tror är hennes femtonåriga dotter som då är borta på en skidresa.
I'm saying mom, and I'm like, okay, what happened? Just like, mom, these bad men have me, help me, help me.
Men det som i själva verket har hänt här är att bedragarna använt AI för ett efterlikna dotterns röst.
Och det går ganska enkelt nu att klona en person's röst genom att ta röstsnuttar som man kan samla in från till exempel sociala medier.
Och troligt obehagligt att inte veta om den som ringer i hennes barn eller inte.
I det här fallet tog det bara minuter innan mamman själv kunde bekräfta att hennes dotter inte var kidnappad.
Men det finns andra exempel på liknande försök så att det här händer nog ut i samhället as we speak.
Ska vi gå vidare på din lista där, Olle? Vad är punkt tre?
Tre är det här.
Vad händer när AI mer och mer används av olika typer av företag, stora eller små?
Storbanken Goldman Sachs har varnat för att väldigt många jobb ligger i farosonen.

En del tror ju att det här är alltid den produktivitetsutveckling som man alltid ser inom näringslivet.
Just det.
Men det finns andra som menar att det här på sikt kan leda till massarbetslöshet.
Jag hörde i veckan om ett utbildningsföretag. Vars börskurs nu rasar för att studenterna snarare använder chat-GPT.
Det är en deras utbildningsmaterial.
Eller musikskapare och illustratörer som förlorar jobben bara för att AI kan göra deras uppgifter billigare och snabbare.
Sen är det så klart att det här communitygruppen som utvecklar systemet är väldigt homogen.
Det är personer som jobbar i Silicon Valley, som är högutbildade, som har väldigt bra löner.
Men många hävdar att deras främsta lojitet nästan är mot tekniken.
Och att de är väldigt, väldigt sugna på de här teknologiska sprongen snarare än att faktiskt väga vilka konsekvenser det här får för övriga samhället.
Och då är vi inne på punkt fyra. Vad händer när man sätter de här teknologierna i händerna på diktatorer?
Ingen har väl missat hur till exempel aktoritära regimer som Kina använder ansiktsgänkning för att övervaka och kontrollera sin befolkning?
Och ju vassare de här förmågorna blir att till exempel hitta mönster i de här stora datamängderna, desto effektivare kan övervakningen och kanske helt enkelt spionaset mot den egna befolkningen bli.
Så på något sätt är det en slags avancerade teckdiktatorer då som med hjälp av AI i minsta det här detalj kan kontrollera människors liv.
Det är liksom målet, skulle du säga, det värsta tänkbara dystopiska framtidsscenarioet vi står inför med AI.
Vi kommer till en sak till, Robin.
Och det är AI i händerna på robotarna själva.
Nyligen kom det ett väldigt uppmärksamt upprop från ledande AI-forskare och många som jobbar i den här branschen där de faktiskt varnar för att vi riskerar att förlora kontrollen över vår civilisation helt enkelt.
Att man inte kan förutse vad de här enorma modellerna kommer göra.
Det finns ju någonting som kallas för generell AI-intelligens.
När AI blir så kraftfull så att den egentligen kan åstadkomma nästan vad som helst eller lösa vilka uppgifter som helst.
Och kan träna sig själv att utföra olika uppgifter nästan på egen hand.
Då skulle man till exempel kunna säga att någon ondiktator med ett ontuppsåt helt enkelt BRA in själv att lösa vissa uppgifter, till exempel lösa krigföring.
Hur ska man invadera det här landet eller vara i nästa steg i vår militäroffensiv och att det inte riktigt finns en mänsklig kontroll över vad som är nästa steg då?
Och där kan man ju ändå se scenario där det skulle kunna spåra ut totalt.
För att jag känner ju att jag har sett det här på film, liksom robotarna tar över och människan liksom använts som någon slags biologiskt bränsle för att driva deras processorer snarare än att vi har något eget värde.
Det är ju en rätt spejsad framtidsdystopi.
Vem menar att det här faktiskt nu är på väg att ske?
En som nyligen kom en sådan vanning var en person som kallas för A-ins gudfader Jeffrey Hinton.

Efter ett decadet på Google's team för artificial intelligence Jeffrey Hinton sa att han residerade sig så att han skulle tala mer främst om teknologi och säga att han skulle bli snabbare än människor.
Han varnar ju för riskerna framåt och att det skulle inom några år kunna utveckla sig en A-id som går förbi mänsklig intelligens.
Du har talat om att A-id kan manipulera eller kanske figure ut en mång för att killa människor. Hur kan det killa människor?
Det blir väldigt bra att manipulera eftersom det kommer att ha lärt oss.
Och det vet hur det är att programma. Så det kommer att figure ut mången för att få runt restriktioner.
Det kommer att figure ut mången för att manipulera människor för att göra vad det är.
Han är ju 75 år och det är väl en annan använt att han också nu drar sig tillbaka.
Men han har tränat väldigt många av de här A-id-forskarna och spelat en väldigt framträdande roll.
Jag tror att han vill prata öppet om de här riskerna.
Att han känner att den här utvecklingen går väldigt fort och att man inte riktigt från de stora företagen i alla stycken har helt kontroll.
Han är ju inte bara negativ utan han ser också massor med möjligheter.
Och det här måste man ju komma ihåg är ett worst case scenario.
Ungefär som de varningar som kom ut när man tillverkade de första atombomberna.
Nu har jag become death, the destroyer of worlds.
I suppose we all thought that one way or another.
Fruktansvärda effekter, men mänskligheten finns ju kvar.
Jag vet inte om jag ska bara känna mig väldigt orolig eller yr av allt det här.
Men om vi liksom nu lugnt och sansat tar ett kliv tillbaka och går igenom den här listan med risker.
Alltså finns det något vi bara kan avfärda direkt som ja, galet eller inte så otroligt i alla fall.
Och finns det annat som vi kanske borde verkligen koncentreras på just nu?
För det första tycker jag faktiskt att man ska tänka bort de här robot dystopierna med metalliska, humanoida robotar
som liksom ögde lägger jorden.
Okej, varför det?
Ja, men det är nästan en stereotypbild av det här och vi vill gärna framställa det så i film och populär vetenskap.
Men jag tror att det finns betydligt mer här och nu faror som också är betydligt mer mänskliga att som vi ska fokusera på och diskutera.
Och vad är det då?
Ja.
Ja, men att vi drunknar i ännu mer skräpinformation.
Att det nästan blir omöjligt för oss vanliga människor att navigera i det här enorma informationsflödet.
Och där tror jag till exempel att seriösa medier har en väldigt viktig roll att spela
för att verkligen guida folk till vad som är fakta baserad journalistik.
Men Olle, även om du då inte verkar tycka att vi ska fokusera för mycket
på den här liksom värsta dystopin där robotarna tar över och utrotar mänskligheten
så är det ju ändå en ganska mörk lista du kommer med här.
Samtidigt finns det väl de som är tvärtom menar att AI kanske är bästa som hänt mänskligheten.

Men vi har ju fokuserat nu på riskerna och farorna och det finns ju massor av möjligheter
att vi ska kunna kommunicera ännu bättre med varandra på olika språk snabbare kanske
att kunna lära sig mer på kortare tid.
En del tycker att det här demokratiserar teknikutvecklingen eftersom modellerna är så öppet tillgängliga
för så många och så lätt att använda.
Men det är tydligt att ägarna och skaparna kanske inte så förvånande är nästan utopiska.
Till exempel Sam Altman som är vd på OpenAI som har tjatkt GPT.
Han twittrar ofta om att AI kommer fria så mycket resurser och tid så att ingen kommer behöva arbeta i framtiden
och alla kan leva i något slags överflöd.
Är det din sänkhet att artificiell intelligens kan faktiskt göra det bättre?
Ja, säkert.
Men jag tror att det är därför jag också blir lite yr för att hur ska man kunna ta in det
och samtidigt alla de här varningarna som kommer nu också ifrån sådana som själva är i branschen?
Det är en enorm hype just nu så det pratar så mycket om det så att man nästan blir yr.
Så därför måste man ta ett djupt andetag och försöka hålla sig lite skeptiskt i det här
och faktiskt också informera sig mer om vad som händer i den här utvecklingen till exempel via medierna.
Ta Elon Musk, han var ju med och grundade OpenAI från början och äger nu Twitter.
Han skrev ju under det här uppropet och tycker att man ska pausa utvecklingen.
Samtidigt rapporterades det ett par veckor senare att han själv hade köpt enorma mängder av sådana här grafikprocessorer
som krävs för att träna de här modellerna och många tror att han vill träna en liknande modell
på all den text och bilder och innehåll som finns på Twitter.
Så på ett sätt så ska man väl också komma ihåg med de här upproporna
att det kan finnas kommersiella logiker även bakom dem.
Alltså att de som varnar för AI också har intressen av att göra det.
Jag tycker att minst under det ska problematiseras.
Man svartmålar någon annan för att själv få en fördel, eller?
Så kan det ju gå till. Och när Google släppte ut sin första version av sin chatbot
så var det såklart väldigt många av konkurrenterna som hånade dem för att den gav så konstiga svar.
Så det här är ett kommersiellt race nu mellan de här jätteföretagen.
Sammanfattningsvis Olle, hur orolig ska jag vara? Hur orolig är du?
För min del så tror jag att i det läge vi är nu så är alltid den mänskliga faktorn
den största risken just nu. Så det är större risk för ett tredje världskrig
på grund av någonting som vi människor utlöser utanför.
Så det är det större risk för ett tredje världskrig på grund av något som vi människor utlöser utan AI, så att säga. Men AI är ju tränad på vår data.
Det mänskliga, allt det vackraste och det fulaste från mänskligheten.
Det finns ju inbyggt i de här modellerna och det är väl det som egentligen är lite skrämmande
att de är så mänskliga.
Jag vet inte om jag blev jättemycket lugnare av tredje världskriget, men tack Olle.

Tack. Och det var allt från Dagens ECO. Vi hörs igen imorgon.
Programledare var Robin Ollin och gäst Olle Sacrison.
Producerade gjorde Felix Österpersonne och Ulrika Lindqvist.
Mela oss gärna på Dagens ECO, att Sverigesradio.se.