KI

VENN ELLER FIENDE

VEGARD HANSSEN

KI

Venn eller fiende

Vegard Hanssen

Innhold

Forord

Jeg har skrevet denne boken fordi jeg er glad i mennesker, og fordi jeg ser at mange mennesker er redde.

De er redde for jobbene sine, for barna sine, for sannheten, for hva som blir igjen av det menneskelige når en maskin kan gjøre stadig mer av det vi trodde bare vi kunne. Og samtidig ser jeg andre mennesker, like kloke og like velmenende, som er lettet og begeistret, som endelig får hjelp de aldri har hatt, som ser muligheter overalt. De sitter ofte ved samme bord, disse to, og snakker forbi hverandre. Jeg ville skrive en bok der de begge får komme til orde, og der ingen av dem blir ledd av.

For jeg tror ikke noen av dem er dumme. Jeg har levd lenge nok til å ha sett dette mønsteret før. Jeg husker da folk var sikre på at fjernsynet ville gjøre oss til en sløv flokk, og da kloke mennesker skrev at internett var en forbigående mote. Jeg vet hvordan vi pleier å reagere på det nye: med begeistring og panikk om hverandre, og hvordan vi som regel finner ut av det til slutt, ikke ved å stoppe og ikke ved å storme fram, men ved den lange, tålmodige forhandlingen et samfunn fører med en ny kraft i sin midte. Den erfaringen gir en viss ro. Vi har vært her før.

Og likevel. Jeg er ikke sikker på at vi har vært her før, ikke helt. For all teknologien vi har fryktet og elsket til nå, forsterket noe ved oss og forble under oss. Den gjorde oss raskere, sterkere, mer vidtrekkende, men den gjorde aldri krav på noe. Kunstig intelligens er kanskje det første vi har laget som i prinsippet kan bli klokere enn den som lagde den. Og det gjør spørsmålet «venn eller fiende» til noe annet enn da vi stilte det om bilen eller telefonen. Innsatsen er en annen.

Denne boken gir deg ikke et svar. Det er et løfte, ikke en unnskyldning. Jeg kommer ikke til å fortelle deg om du bør omfavne kunstig intelligens eller frykte den, for jeg vet det ikke, og jeg stoler ikke på dem som later som de gjør. Det jeg vil, er å ta deg ved hånden og gå rolig gjennom det hele, stemme for stemme, frykt for frykt: argumentene for og argumentene mot, spørsmålene vanlige folk faktisk stiller, det maskinen alt gjør for oss og det den kan komme til å gjøre mot oss. Jeg vil holde to hender like høyt hele veien, og jeg vil be deg holde dem oppe sammen med meg, også når det klør i fingrene etter å la den ene synke.

Men jeg skal være ærlig med deg om én ting til, her i begynnelsen, for du kommer til å merke det uansett. Jo lenger inn i denne boken vi kommer, jo mer mistenker jeg at striden om kunstig intelligens ikke egentlig handler om maskinen. At den, under alt det praktiske og alvorlige, til slutt handler om noe i oss. Om hvordan vi har det med tanken på at mennesket kanskje ikke lenger selvsagt sitter øverst. Jeg skal ikke si mer om det nå. Jeg ber deg bare bære spørsmålet med deg, som en liten stein i lommen, mens vi går. Vi tar det fram igjen mot slutten.

La oss begynne der det er lettest å puste, før alvoret setter inn. La oss begynne med at vi har sett dette før.

Den samme krangelen, om igjen

På slutten av attenhundretallet sto det predikanter i Sverige og advarte menigheten mot telefonen. Den var djevelens instrument. Gjennom kobbertråden kunne onde ånder finne veien like inn i stua, sammen med torden og lyn. Noen klipte ned linjene. Andre var redde for at selve samtalen skulle renne ut hvis tråden brakk et sted langs veien.

Vi smiler av det nå. Men vi bør smile forsiktig.

For de menneskene var ikke dumme. De sto foran noe de ikke hadde ord for, en stemme uten en kropp, et nærvær uten et ansikt. Og de grep etter de bildene de hadde. Det ukjente ble forklart med det de allerede var redde for. Slik gjør vi fortsatt. Vi gjør det med kunstig intelligens akkurat nå.

Det er verdt å begynne der, før vi begynner å krangle. For vi kommer til å krangle. Og vi har gjort det før.

Rollene vi alltid inntar

Hver gang noe nytt og mektig kommer inn i samfunnet, deler vi oss omtrent likt. Noen lener seg fremover. Dette forandrer alt, sier de, den som nøler blir forbigått, vi må gå fortere. De ser muligheter, lettere arbeid, ting som endelig blir mulig. Andre lener seg bakover. Vi går for fort, sier de, vi mister noe vi ikke får igjen, la oss vente til vi skjønner hva vi har med å gjøre. De ser tap: arbeid, kontroll, noe vanskelig å sette ord på. Og en tredje gruppe står et sted imellom og mumler at svaret nok verken er stopp eller full fart, men regler, vaner, en måte å leve med tingen på.

Entusiasten, skeptikeren, pragmatikeren. Det er nesten alltid de samme rollene som blir besatt. Bare teknologien byttes ut.

Legg merke til hva de to ytterpunktene egentlig er uenige om. Sjelden om tingen er god eller ond. Som regel om *tempoet*. Den ene siden er redd for å gå glipp av fremtiden. Den andre er redd for å miste fortiden. Begge mener tempoet er feil. De er bare uenige om hvilken vei det bommer. Den uenigheten kjenner vi igjen fra hver eneste teknologi som har skremt og begeistret oss på en gang.

Hadde de redde feil?

Her er det fristende å si: og hver gang viste det seg at skeptikerne tok feil. Slapp av. Det ordner seg alltid.

Men det stemmer ikke.

Det er sant at mye av frykten ser overdreven ut i ettertid. Ingen ble gal av å reise med tog, slik legene advarte om på attenhundretallet. De trodde menneskekroppen ikke tålte fart over tretti km/t. Melken surnet ikke av reisen. Telefonen slapp ingen ånder inn i stua. Romanen ødela ikke ungdommen, fjernsynet gjorde oss ikke til en hjernedød befolkning, og internett ble ikke det forbigående blaffet en og annen klok mann spådde at det ville bli.

Og likevel. Noen ganger hadde de redde rett.

Da fysikerne selv begynte å advare mot atombomben, var det ikke moralsk panikk. Det var presist. Da de første stemmene sa at noe holdt på å gå galt med de unge og smarttelefonene og de sosiale mediene, ble de avfeid som bekymrede gammelmodige sjeler. I dag er vi ikke så sikre lenger. Den frykten ser mer berettiget ut for hvert år.

Så mønsteret er ikke at skeptikerne alltid tar feil. Mønsteret er vanskeligere enn som så, og mer ydmykende. Vi blander de to sammen. Vi klarer nesten aldri å skille den overdrevne frykten fra den berettigede *mens det står på*. Det er først etterpå, når alt er over og vi vet hvordan det gikk, at vi kan sortere klokt i det og si: der hadde de rett, der tok de feil. Underveis ser begge slags frykt nøyaktig like ekte ut.

Det er en ubehagelig erkjennelse å ta med seg inn i en bok om kunstig intelligens. For akkurat nå står vi midt i det. Vi vet ikke ennå hvilken slags frykt vår er.

Luddittene var ikke det vi tror

Det finnes et ord vi bruker om folk som er imot ny teknologi. Vi kaller dem ludditter. Det er ment som et lite stikk: her kommer en som vil knuse maskinene og dra oss tilbake til mørket.

De opprinnelige luddittene fortjente ikke det ettermælet.

De var engelske tekstilarbeidere på begynnelsen av attenhundretallet, faglærte folk som hadde brukt år på å bli gode i håndverket sitt. Og så kom vevmaskinene. Ikke maskinene i seg selv var problemet, men måten de ble brukt på: til å kutte lønninger, og til å erstatte faglærte med ufaglært, billigere arbeidskraft. Luddittene var ikke prinsipielt imot fremgang. De så noe ganske presist. De så *hvem* som kom til å betale regningen for skiftet. Og det var dem selv.

De tok feil om at maskinene ville utslette arbeidet for godt. Arbeidet forsvant ikke. Det ble flyttet, forvandlet, og over tid ble det flere jobber, ikke færre. På den måten var de redde for mye.

Men de hadde rett om sitt eget liv. En hel generasjon faglærte mistet virkelig levebrødet og statusen sin, og overgangen var brutal for dem som sto i den. «Det går bra til slutt» er en svak trøst for den som blir arbeidsledig i mellomtiden. Den setningen er sann for samfunnet og falsk for mennesket. Vi kommer til å møte den igjen, for den ligger under nesten alt vi skal snakke om.

Parlamentet, for ordens skyld, svarte med å gjøre det til en dødssynd å knuse en maskin, og sendte rundt tolv tusen soldater mot bevegelsen, flere enn England på den tiden hadde sendt for å slåss mot Napoleon i Portugal. Vi har altså vært villige til å bruke makt for å verne det nye, lenge før vi visste om det nye fortjente det.

Hvorfor det nye skremmer oss

Det er lett å tro at frykt for teknologi er en slags dumhet, en mangel på opplysning som forsvinner straks folk forstår bedre. Det stemmer ikke. Frykten har dype røtter, og noen av dem er kloke.

Vi mennesker er bygget til å holde på det vi har. Et tap svir mer enn en tilsvarende gevinst gleder. Det skal omtrent dobbelt så mye godt til for å oppveie det vonde ved å miste noe. Derfor måler vi det nye strengt. Gevinsten det lover ligger et sted i fremtiden og er usikker; tapet det truer med føles nært og konkret. Vi velger det kjente, ikke fordi det alltid er bedre, men fordi det er trygt. Og vi angrer mer på det gale vi gjorde enn på det gale vi lot være å gjøre, så vi lener mot å vente, å la tingen ligge.

Mye av uroen handler om noe enklere enn alt dette også. Den handler om kontroll. Mister jeg styringen over mitt eget liv, min egen jobb, mine egne barn? Det spørsmålet blir skarpere jo mer en teknologi ser ut til å handle på egen hånd, jo mindre den oppfører seg som et verktøy vi holder i hånden og mer som noe som beveger seg selv. Det skal vi komme tilbake til, for det er nettopp her kunstig intelligens skiller seg fra alt vi har laget før.

Og så er det den frykten som er minst forfinet og mest berettiget av dem alle: forsvinner inntekten min? Når uroen er økonomisk, er den sjelden dum. Den er ofte helt rasjonell for det mennesket som rammes, selv om regnskapet for samfunnet som helhet ender med pluss. Setteren i trykkeriet, drosjesjåføren, ludditten ved veven. De regnet ikke feil. De regnet på sitt eget liv.

Når frykten blir for stor

Det finnes et begrep for det motsatte, for når reaksjonen blir større enn faren. Sosiologene kaller det moralsk panikk. En gruppe, en ting eller en hendelse blir plutselig stemplet som en trussel mot alt vi holder kjært, og reaksjonen vokser seg ute av proporsjon med den faktiske faren. Mediene forsterker. Noen stemmer tjener på opphisselsen. Og ny medieteknologi har vært en av de mest pålitelige utløserne av slik panikk gjennom historien: romanen, tegneserien, fjernsynet, videospillet, og nå dette.

Men her må vi være forsiktige, for begrepet kan misbrukes. Det er fristende å avfeie all uro som moralsk panikk, å vinke bort enhver bekymring med et overbærende «dette har vi hørt før». Det er ikke det begrepet er til for. Det er et redskap for å skille den forholdsmessige uroen fra den uforholdsmessige, ikke et bevis for at all uro er tåpelig. Noen ganger er panikken nettopp panikk. Andre ganger er den et tidlig varsel vi burde ha lyttet til. Oppgaven er å vite forskjellen, og den oppgaven er vanskelig hver eneste gang.

Slik løser vi det egentlig

Hvis det er én ting historien lærer oss om hvordan disse stridene ender, så er det dette: de ender nesten aldri med et rent ja eller et rent nei.

Bilen ble ikke forbudt, og den fikk heller ikke fritt spillerom. Den fikk fartsgrenser. Førerkort. Sikkerhetsbelter, og etter hvert kollisjonsputer, og lover mot å kjøre i fylla. Elektrisiteten var virkelig farlig i begynnelsen, med dårlig isolerte ledninger og mennesker som døde av støt, helt til det kom standarder, jording og tilsyn, og den ble gjort trygg. Ikke fordi den var trygg i seg selv, men fordi vi gjorde den trygg, sakte, gjennom mange små grep.

Slik går det stort sett. Det vokser fram lover. Det vokser fram vaner og folkeskikk, ofte raskere enn lovene. Det bygges institusjoner som skal holde øye med tingen. Folk venner seg til den, og generasjonen som vokser opp med den, opplever den ikke som et brudd i det hele tatt, bare som verden slik den er. Arbeidet flytter seg. Og ofte temmer teknologien seg selv, ved at den neste utgaven retter opp det den forrige stelte i stand.

Løsningen er nesten alltid en lang forhandling, der samfunnet gradvis bygger rammer rundt det nye. Det er ikke et dramatisk svar. Det selger ingen avisforsider. Men det er sånn det skjer, og i den forhandlingen får både entusiasten og skeptikeren noe rett til slutt.

Og likevel — denne gangen

Alt dette taler for ro. Vi har sett dette mønsteret så mange ganger at vi nesten kan tegne det på forhånd: begeistringen, panikken, korreksjonen, reglene, tilvenningen. Det er en god grunn til ikke å miste hodet nå.

Det er bare ett problem. Og det skal vi bruke en stor del av denne boken på å nærme oss forsiktig.

Alle de tidligere teknologiene forsterket noe ved mennesket. Telefonen forlenget stemmen vår, bilen beina våre, datamaskinen regnehodet vårt. De gjorde oss raskere, sterkere, mer presise. Men de forble under oss. De gjorde ikke krav på noe. En vevmaskin satte seg aldri et eget mål.

Kunstig intelligens er det første vi har laget der det i det hele tatt er et alvorlig spørsmål om tingen kan bli klokere enn den som lagde den. Det er ikke sikkert at den blir det; om det er det dyp og reell uenighet, og den skal vi ta på alvor. Men at spørsmålet kan stilles, er i seg selv nytt. Ingen har noensinne lurt på om trykkpressen kunne komme til å tenke bedre enn forfatteren.

Så vi bærer to ting med oss inn i det som kommer, og de drar i hver sin retning. Det ene er mønsteret, som hvisker: ro deg ned, du har sett dette før. Det andre er muligheten for at det denne gangen er noe kvalitativt annerledes. Det kloke er ikke å velge én av dem og lukke øret for den andre. Det kloke er å holde begge åpne på en gang.

For det er en gammel sannhet i alt dette, og den er verdt å ta med seg videre: vi har for vane å overvurdere hva det nye betyr på kort sikt, og undervurdere hva det betyr på lang sikt. Vi tar sannsynligvis feil om hva kunstig intelligens gjør med oss neste år. Og vi tar sannsynligvis feil om hva den gjør med oss om tjue. Begge leirene kan ha rett, på hver sin klokke.

La oss derfor ikke skynde oss til en dom. La oss heller gjøre det vi sjelden gir oss tid til midt i en slik strid: gå rolig gjennom den, stemme for stemme, og se hva som faktisk står der når vi tør å se ordentlig etter.

Stopper vi noen gang?

Det blir ofte sagt, når noen vil berolige oss om kunstig intelligens, at historien er full av ganger vi snudde. At mennesket har vett til å legge fra seg en farlig vei når vi ser hvor den bærer. Det høres riktig ut. Og noen kan til og med peke på et eksempel.

Et eksempel man gjerne griper til, er oljen. Se, sier man, vi er jo i ferd med å forlate den.

Men se litt nøyere etter. Hva var vi egentlig ute etter med oljen? Ikke oljen. Energien. Og energien går ikke bakover. Den går fremover, fortere enn noen gang, vi har bare så vidt begynt å lage den på en annen måte. Vi snudde ikke fra noe som helst. Vi byttet oppskrift, og lot behovet marsjere videre akkurat like fort som før. Det er noe ganske annet enn å stoppe.

Og det er her det blir interessant, for dette er ikke et unntak. Det er regelen. En teknologisk strid er nesten aldri om teknologien selv. Den er om hva teknologien lar oss gjøre. Oljen var aldri poenget. Det vi kan gjøre med energi, var poenget. Og en evne vi først har fått, et «dette kan vi nå», gir vi nesten aldri fra oss igjen. Vi bytter bare ut måten vi får den på.

Vi bytter ut. Vi gir ikke opp

Se på de andre tilfellene vi liker å kalle seire, og legg merke til det samme mønsteret under.

Det var en gang noen gasser som åt opp ozonlaget over oss, det laget som skjermer alt levende mot strålingen fra sola. Verden forbød dem, i en av de få globale avtalene som virkelig har virket, og ozonlaget leges nå sakte. En ekte seier, verdt å være stolt av. Men vi sluttet ikke å kjøle ned maten og husene våre. Vi byttet til andre gasser. De viste seg å være kraftige drivhusgasser, og nå holder vi på å fase dem ut igjen, til fordel for noe tredje. Kjølingen marsjerte videre. Det var aldri gassen vi var ute etter. Det var det kalde rommet.

Eller blyet. I over et halvt århundre fylte vi bensinen med en gift som skadet hjernen til barn over hele kloden, helt til den aller siste literen ble brukt opp i Algerie så sent som i 2021. Det er en sann og viktig fremgang, og den reddet utallige barn fra skade. Men vi sluttet ikke å kjøre. Vi tok giften ut av oppskriften og kjørte videre. Behovet for å komme oss fra sted til sted rørte vi ikke.

Hver gang vi tror vi har lagt noe fra oss, viser det seg som regel at vi bare har skiftet ut hvordan. Selve ønsket bærer vi videre, urørt.

Når ga vi da virkelig opp noe

Det finnes unntak. De er få, og nettopp derfor er de verdt å se nøye på, for de forteller oss hva som faktisk skal til før et menneske gir slipp på en evne det har fått.

Vi ga opp statlig rasehygiene. Ikke fordi evnen forsvant, men fordi selve ønsket ble uutholdelig etter at vi så, midt på det forrige århundret, hvor det bar hen. Det var ingen oppskrift vi byttet ut. Det var et ønske som døde.

Sør-Afrika ga fra seg atomvåpnene sine. Landet er det eneste i verden som har bygget kjernevåpen helt selv og siden demontert dem alle, seks ferdige og ett under arbeid, like før apartheidstaten falt. Heller ikke her var det teknologien som forsvant. Det var behovet bak den: et regime som innså at det ikke lenger trengte våpenet det en gang hadde fryktet seg til.

Og så det ferskeste forsøket, det som ligner kunstig intelligens mest av alt. I 1975 satte forskerne på et nytt og skremmende felt, gensplising, seg sammen på en konferanse i Asilomar i California og ble enige om å stanse det farligste arbeidet selv. Frivillig. Ingen myndighet tvang dem. Faget passet på seg selv, på forhånd, av samvittighet. Det er et vakkert øyeblikk, og det blir stadig holdt fram som beviset på at vi kan holde igjen når vi bestemmer oss for det.

Men se hva som skjedde siden. I 2018 redigerte den kinesiske forskeren He Jiankui genene til to jenter mens de ennå var embryoer, krysset nettopp den grensen alle var blitt enige om at ingen skulle krysse, og verden fordømte ham med én stemme. Han fikk tre år i fengsel. I 2022 var han ute igjen. Kunnskapen han brukte, blir billigere og lettere tilgjengelig for hvert år som går, og ingenting hindrer at en annen forsker, eller en hel stat, gjør det samme i morgen. Og gjør en stat det, hvordan straffer vi den da? Asilomar viser at vi kan bli enige om en grense. Mannen som krysset den, viser hvor lite enigheten er verdt mot én som bestemmer seg for å gå over. En regel hvem som helst kan bryte ved å stikke av, er ikke en brems. Den er et håp.

Begge sider har noe rett her. Det er ikke umulig å holde igjen. Ozonlaget leges. Sør-Afrika ga virkelig fra seg bomben. «Umulig» er feil ord. Men legg merke til betingelsen som var oppfylt hver eneste gang det lyktes: enten fantes det en erstatning som dekket det samme ønsket, eller så var ønsket selv dødd. Det er den betingelsen som avgjør alt. Og det er den vi nå skal holde opp mot kunstig intelligens.

Det ene ønsket uten erstatning

For det finnes ett slags ønske vi aldri har klart å bytte bort, og som ikke har dødd: ønsket om det ytterste overtaket over en motstander.

Vi har alt vært innom bomben. Det vi ikke sa da, er dette: vi stoppet den aldri. Vi forbød den ikke. Det fantes ingen erstatning for «det avgjørende overtaket», og ønsket om det dør tydeligvis ikke. Så vi beholdt den. Vi lærte å leve under den. Vi begrenset hvor den kunne prøves, vi telte hverandres stridshoder, vi bygde en hel verdensorden rundt trusselen om å bruke den, og de som har den, minner stadig verden om det hver gang en konflikt blir alvorlig nok. Den henger over hvert menneskeliv ennå. Det er ikke en frykt vi temmet. Det er en frykt vi flyttet inn i huset og la oss til å sove ved siden av.

To utganger, altså, hver gang noe mektig kommer inn i verden og blir værende. Enten forvandler vi det, bytter ut oppskriften og lar ønsket gå videre, slik vi gjorde med energien og kjølingen og blyet. Eller så finner vi at ønsket ikke har noen erstatning, og da stopper vi det heller ikke. Da frykter vi det, begrenser det så godt vi kan, og lærer å leve under det. Det vi nesten aldri gjør, er det vi liker å innbille oss at vi gjør: å snu.

Den vi verken kan bytte ut eller legge fra oss

Så hva da med kunstig intelligens?

Striden om den er, som alle de andre, ikke om teknologien. Den er om hva vi kan gjøre med den. Og det vi kan gjøre med den, er å tenke, vurdere og bestemme, i en skala og en hastighet vi ikke kan måle oss med selv. Still da det avgjørende spørsmålet, det samme vi har stilt om alt annet: hva skulle vi bytte det mot? Vi forlot oljen fordi vi egentlig ville ha energi, og energi kan lages på mange vis. Vi forlot gassene fordi vi egentlig ville ha kulde, og kulde kan lages på mange vis. Men en intelligens som overgår vår egen er ikke en oppskrift på noe annet. Den er selve det vi er ute etter. Det finnes ingen sidevei dit som ikke nettopp er den.

Derfor passer den trøstende utgangen ikke her. Vi kan ikke forvandle kunstig intelligens til noe annet som stiller det samme ønsket, slik vi forvandlet energien, for det finnes ikke noe annet som stiller det. Det lar bare den andre utgangen stå åpen. Den vi gikk med bomben. Å frykte, å begrense, og å lære å leve under.

Det er ikke et beroligende sted å lande. Men det er ærligere enn løftet om at vi snur når vi vil. Vi snur nesten aldri. Vi bytter ut, eller vi legger oss til å sove ved siden av. Og kunstig intelligens kan se ut til å være det første vi har laget som vi verken kan bytte ut eller helt legge fra oss. Resten av denne boken handler om hva slags naboskap det da blir.

Allerede overalt

I dag, før du i det hele tatt tenkte ordet «kunstig intelligens», brukte du den sannsynligvis et titalls ganger.

Den sorterte vekk søppelposten, slik at du så de tre e-postene som betydde noe og ikke de tre hundre som ikke gjorde det. Da du betalte med kortet, satt det et system og vurderte på et øyeblikk om akkurat den transaksjonen så ut som svindel, et system som gjør den vurderingen over en milliard ganger i døgnet. Kartet på telefonen gjettet hvor lang tid du ville bruke, og tok høyde for kø den ikke kunne se. Da du tok et bilde, var det knapt et fotografi i gammel forstand; telefonen regnet seg fram til hvordan natten skulle se lysere ut, fant ansiktet i mengden og stilte skarpt på det. Og skrev du en melding på et språk du ikke behersker, var det en maskin som oversatte.

Ingen av disse tingene ba om oppmerksomheten din. De bare virket. Og det er det første vi må ha klart for oss før vi krangler: kunstig intelligens er ikke en fremtid som banker på døren. Den har bodd i huset i årevis, og den har stort sett oppført seg pent.

Det meste av den snakker ikke

Når folk i dag sier «KI», mener de nesten alltid det samme. En chatbot man skriver til, som svarer med setninger. Et program som lager et bilde av en setning. Det er den nye, synlige skiva. Men det er en tynn skive av noe mye større. Det aller meste av den kunstige intelligensen som alt har forandret verden, snakker ikke. Den kjenner igjen mønstre. Den ser, den sorterer, den forutser, den anbefaler. Den ligger i kameraet og i kortterminalen og i strømnettet, og den har tjent oss, og tjent gode penger, i god tid før noen chatbot fantes.

Det skillet er verdt å holde fast på gjennom hele boken. Mye av både frykten og begeistringen treffer bare den ene, nye skiva, mens den brede, innvevde maskinen ruller videre uansett hva vi mener om den.

Der den har vært en ren venn

Det tydeligste stedet kunstig intelligens har vært udelt til hjelp, er i medisinen.

For noen år siden løste et program ved navn AlphaFold en gåte biologien hadde strevd med i et halvt århundre: hvordan en proteinkjede folder seg til den formen som avgjør hva den gjør i kroppen. Det er et av de mest grunnleggende spørsmålene i livsvitenskapen, og det hadde stått ulåst i femti år. AlphaFold bestemte formen på over to hundre millioner proteiner, nær alle vi kjenner, og ga hele samlingen gratis til forskere over hele verden. Folkene bak fikk Nobelprisen i kjemi. En av de første tingene andre brukte verktøyet til, var å komme nærmere en vaksine mot malaria.

På sykehuset er den blitt et ekstra par øyne. Den leser røntgenbilder og mammografier og flagger det en sliten radiolog kan komme til å overse klokka fem på ettermiddagen. Den finner tegn på hjerneslag i en skanning i de minuttene der hvert minutt teller. For en øyesykdom som kan gjøre diabetikere blinde, finnes det nå et verktøy som stiller diagnosen selv, så flere kan fanges opp i tide. Og kanskje det minst dramatiske og mest avholdte av alt: et verktøy som bare sitter og lytter til samtalen mellom lege og pasient og skriver journalnotatet automatisk, slik at legen slipper å sitte med ryggen til deg og taste. Det gir legen tid tilbake til mennesket foran seg. Nesten ingen protesterer mot den slags.

Eller tenk på den som beskriver verden for en blind. Det finnes apper nå som leser opp hva kameraet ser: hvilken knapp på komfyren som er hvilken, hva det står på melkekartongen, hvem som sitter rundt bordet. De er bygget sammen med blinde brukere, og de gir tilbake en bit selvstendighet som er vanskelig å sette pris på utenfra. Her er nytten så direkte at det nesten ikke finnes en motstemme.

Når den velter en hel metode

Noen ganger gjør den mer enn å hjelpe. Den velter en etablert måte å gjøre ting på, nesten over natten.

Før AlphaFold prøvde forskere å løse proteingåten med dugnad. Vanlige folk lånte bort den ledige regnekraften i hjemme-PC-ene sine, og under pandemien vokste ett slikt nettverk seg kraftigere enn verdens fem hundre største superdatamaskiner til sammen. Så kom AlphaFold og løste problemet de hadde slitt med. Det kunne sett ut som slutten for dugnaden. Det ble det ikke. AlphaFold løser én del av gåten, den ferdige formen, mens dugnadsnettverkene kunne ta fatt på andre deler, som hvordan proteinet rører seg over tid. Den ene ga den andre et bedre utgangspunkt. Og en av pionerene bak den gamle dugnaden fant en ny vei, å designe helt nye proteiner fra bunnen av, og delte den samme Nobelprisen som folkene bak AlphaFold.

Det samme skjedde med været. KI-modellene som slo de gamle, tunge fysikkmodellene, var trent på data fra nettopp de fysikkmodellene de slo, og det store europeiske varslingssenteret som lenge satte standarden, kjører nå sin egen KI-modell ved siden av den gamle, ikke i stedet for den. «Det forsvant over natten» er nesten alltid en for enkel historie. Som regel blir faget forvandlet, ikke utslettet. Det er verdt å huske, både når noen lover at kunstig intelligens feier alt vekk, og når noen sier at den ikke har levert noe som helst.

Men «ingen klager» er for enkelt

Jeg skal ikke late som om «ingen klager» er hele sannheten. Det er det ikke. Anbefalingsmaskinene som holder oss klistret til skjermen, har gode grunner til å bli kritisert. Medisinsk KI kan arve skjevhetene i dataene den er trent på. Og noe av den letteste, mest behagelige hjelpen er nettopp den vi senere skal spørre om gjør oss litt dummere. Bruken er ikke hevet over kritikk.

Poenget er at det finnes en stor, stille kategori av kunstig intelligens som nesten alle godtar, og at den ofte er mest godtatt akkurat når den hjelper mennesket i stedet for å erstatte det. Det ekstra paret øyne ved siden av legen. Verktøyet som gir legen tid tilbake. Stemmen som beskriver rommet for den som ikke ser. Det er et tidlig hint om hvor grensen mellom venn og fiende kanskje kommer til å gå i praksis. Ikke ved hva maskinen kan, men ved om den står ved siden av oss eller i stedet for oss.

Vi begynner altså ikke på bar bakke. Når striden snart setter inn for alvor, om jobber, om sannhet, om kontroll, hviler den oppå en bred sokkel av kunstig intelligens som alt er her, og som for det meste gjør hverdagen litt bedre uten å be om noe. Det gjør ikke frykten feil. Det betyr bare at vi må holde to ting i hodet samtidig, slik vi har måttet hele veien: den samme teknologien som finner et svulstanlegg før legen ser det, er den vi om noen kapitler skal spørre om kan bli klokere enn oss. To ting på én gang, hele veien.

Slik høres debatten ut

Før vi går inn i argumentene, er det verdt å lytte til hvordan de faktisk lyder. Ikke i ekspertenes språk, men slik vanlige mennesker sier dem, ved kjøkkenbordet og i kommentarfeltet og i foreldresamtalen i gangen utenfor klasserommet. For det er der debatten egentlig bor.

Og lytter du en stund, hører du at det går to grunntoner gjennom nesten alt som blir sagt.

De to grunntonene

Den ene er frykt for tap. «Jeg har brukt tjue år på å bli god i faget mitt, og nå gjør en maskin det samme på tre sekunder.» «Hvis maskinen skriver stilen, hva er det da ungen min faktisk lærer?» «Jeg så en video av statsministeren si noe, og så var den falsk. Hvem kan jeg stole på nå?» «En maskin kan ikke være glad i deg. Den later bare som, og du merker ikke forskjellen før det er for sent.» «Hvis den kan tenke og skrive og skape bedre enn meg, hva er jeg da til for?» Det er ulike temaer, men den samme uroen under: noe blir tatt fra meg, fra barna mine, fra det som gjorde meg til meg.

Den andre grunntonen er lettelse. «Den tar de kjedelige rutineoppgavene, så jeg kan bruke tida på det som faktisk betyr noe.» «Datteren min får forklart brøk på fem ulike måter til hun skjønner det, det rakk aldri læreren før.» «Bestemor på nitti sier hun ikke er ensom lenger, hun har noen å snakke med hele døgnet.» «For meg med dysleksi er dette nesten frigjørende. Endelig blir jeg forstått på papiret også.» Her er ikke maskinen en tyv, men en hjelper. Noe blir gitt: tid, mestring, selskap, en lavere terskel inn.

Det interessante er at de to stemmene ofte sitter ved samme bord, og noen ganger i samme menneske. Den samme moren som er lettet over privatlæreren i lomma, ligger våken over hva chatboten gjør med sønnen om natten. Vi er sjelden rent for eller rent imot. Vi er begge deler, om litt ulike ting.

Den gruppen som er verdt å merke seg

Det finnes en tredje stemme også, og den er lett å overse fordi den verken roper eller jubler. Det er ofte de unge selv. De bruker kunstig intelligens hver dag, uten å gjøre noe stort nummer av det, omtrent som foreldrene deres en gang begynte å bruke Google. Men de er ikke blindt begeistret. «Vi er også redde for at det presser opp karakterkravene og sprer falske nyheter,» sier de. «Vi vil ha klare regler og skikkelig opplæring, ikke at de voksne enten forbyr alt eller skjønner ingenting.»

Det er en klok holdning, og boken kommer til å lene seg på den mer enn én gang: ta verktøyet i bruk, men still krav til hvordan. Verken panikk eller blind tillit, men en voksen forventning om

grenser.

Og når det blir alvor, blir det stort

Noen av spørsmålene er små og praktiske. Andre blir fort de helt store. «Når mennesket lager noe som tenker selv, er det da vi prøver å være gud?» spør noen, fra et religiøst ståsted, og det er én av mange måter folk rammer inn det samme spørsmålet om hva mennesket egentlig er. «Det blir som i Terminator,» sier andre. «En dag våkner systemet og vender seg mot oss.» Og en tredje svarer trøstende: «Terminator er en film. Den virkelige maskinen vet ikke engang at den finnes. Den vil ingenting, den regner sannsynligheter.»

Vi skal komme tilbake til hvert av disse, rolig og etter tur. Måten vi vil gjøre det på, er den samme hele veien: ta redselen på alvor først, i sin sterkeste form, og så nyansere den. Aldri vifte den bort. For den som er redd, fortjener å bli lyttet til før hun blir svart.

Fra kjøkkenbordet til toppmøtet

Hev så blikket fra kjøkkenbordet og opp til dem som mener noe om dette på heltid: forskerne, gründerne, filosofene. Du skulle tro de hadde ryddet opp i forvirringen. Det har de ikke. De krangler like heftig, bare med flere fotnoter.

En nyttig måte å sortere dem på kommer fra en av Silicon Valleys egne, som deler feltet i fire. Det er dommedagsprofetene, som mener avansert kunstig intelligens er en trussel mot menneskehetens overlevelse og bør stanses. Det er pessimistene, som ikke tror på utryddelse, men på en ustoppelig marsj mot tap av jobber og menneskelig betydning. Det er optimistene, som vil ha full fart fremover og ser mest velsignelse. Og det er de forsiktig håpefulle, som vil kjøre videre, men med foten hvilende på bremsen.

Navnene skal vi møte etter hvert. På den ene fløyen står folk som en av nevralnettenes fedre, en mann som forlot en av verdens største teknologibedrifter for å kunne advare fritt, og som anslår en reell sjanse for at dette ender med menneskehetens undergang. På den andre fløyen står like tunge navn, en annen pioner og en kjent investor, som mener frykten er nær science fiction og at det egentlig farlige er å bremse en historisk velsignelse. Og midt imellom står de som bygger de kraftigste systemene selv, og som ofte sier begge deler på en gang.

Striden går ikke der du tror

For her er det som overrasker folk mest når de først ser nøye etter. Den skarpe striden går sjelden mellom «for KI» og «mot KI».

Flere av de ivrigste optimistene bærer på dyp bekymring samtidig. En av dem som bygger noen av verdens mektigste modeller, har offentlig anslått at det er omtrent én sjanse av fire for at det hele ender katastrofalt, og fortsetter å bygge likevel. Det er ikke nødvendigvis hykleri, og vi skal bruke et helt kapittel på å forstå hvorfor.

Og den mest opprivende uenigheten går ofte ikke mellom optimist og pessimist i det hele tatt. Den går mellom to slags bekymrede. De ene frykter en fjern, mulig utryddelse, en superintelligens som slipper unna vår kontroll. De andre blir rasende av nettopp den frykten, fordi de mener den stjeler

oppmerksomheten fra skadene som skjer her og nå: skjevhet, overvåkning, utnyttede arbeidere, en informasjonsverden som råtner. Disse to leirene er ofte mer uenige med hverandre enn med optimistene.

Ta det med deg, for det avslører at «venn eller fiende» er for grovt. Nesten ingen mener at kunstig intelligens er ren verdi eller ren fare. Striden står om tempo, om styring, om hvilke farer som er virkelige og hvilke som er skygger, og om hvem som skal bestemme. Det er den striden vi nå skal ta del for del, og vi begynner der det er mest fristende å bare velge en side: hos dem som heier.

De som heier

La oss begynne med den lyse siden, og la oss ta den på fullt alvor. Det er en uvane i seriøse bøker om kunstig intelligens å haste forbi optimistene på vei til faren, som om begeistring var en form for naivitet. Det er den ikke. Noen av de klokeste menneskene i feltet er dypt håpefulle, og de har grunner vi bør lytte til skikkelig.

Grunnholdningen deres er enkel. Kunstig intelligens er en allmennteknologi, på linje med elektrisiteten eller dampmaskinen. Den vil sive inn i alt, og over tid løfte produktivitet, helse, vitenskap og levestandard. Og historien, sier de, har en tydelig slagside: hver gang vi har fryktet en ny teknologi, har vi overvurdert de varige tapene og undervurdert det den til slutt ga oss.

Mer å gå rundt på

Det første argumentet er det minst spennende, og kanskje det som betyr mest: produktivitet. I store deler av den rike verden vokser arbeidsstyrken sakte eller krymper. Vi blir færre i arbeidsfør alder og flere som skal forsørges. Hvis hver av oss kan få gjort mer på samme tid, kan velstanden vokse likevel. Anslagene varierer, og de tidlige tallene skal leses med varsomhet, men retningen optimistene peker i, er at en maskin som tar unna det kjedelige grovarbeidet, frigjør mennesket til det som krever dømmekraft og nærvær.

Den frigjøringen treffer ikke bare de store. En liten bedrift, en gründer uten kapital, en håndverker med papirarbeid han gruer seg til, får nå tilgang til verktøy som for få år siden var forbeholdt selskaper med egne avdelinger. Det er en utjevning, sier optimistene: de samme kreftene som før bare de mektige hadde råd til, ligger nå i lomma på hvem som helst.

En privatlærer for hvert barn

Det argumentet som rører flest, handler om skolen. Det har lenge vært kjent at barn som får én lærer for seg selv, lærer dramatisk mye bedre enn barn i en klasse på tretti. Problemet har alltid vært at vi ikke har råd til en privatlærer til hvert barn. Det kunne kunstig intelligens endre.

Tanken, slik Sal Khan, en av pionerene bak læringstjenesten Khan Academy, formulerer den, er en tålmodig veileder for hvert barn på jorden. Ikke en som gir bort svaret, men en som stiller spørsmål til barnet selv finner det, som forklarer på en femte måte når de fire første ikke nådde fram, som aldri blir lei og aldri dømmer. Et barn uten råd til ekstrahjelp kunne få noe som ligner det rikmannsbarn alltid har hatt. Hvis det holder det det lover, er det noe av det mest løfterike i hele denne fortellingen.

En motor for oppdagelse

Så er det vitenskapen. Vi har alt sett hva kunstig intelligens gjorde med proteingåten. Optimistene ser det som begynnelsen, ikke unntaket. Den samme typen maskin har funnet millioner av nye materialer som kan bli morgendagens batterier og solceller, og den hjelper til med å styre det ustyrlige plasmaet i forsøkene på å temme fusjonsenergien. Dypest sett kan den utforske et mulighetsrom som er for stort for et menneskeliv, og finne mønstre ingen rekker å lete seg fram til for hånd. Demis Hassabis, en av de fremste i feltet, sier det slik: løs intelligensen, og bruk så intelligensen til å løse alt det andre.

Det er her optimismen blir stor. For hvis maskinen kan akselerere selve oppdagelsen, kan den i prinsippet korte inn på problemer vi har strevd med i generasjoner. Dario Amodei, en av de mest tankefulle av lederne, har skrevet et langt essay om en verden der kraftig kunstig intelligens komprimerer femti til hundre år med medisinsk fremgang til fem eller ti: de fleste sykdommer kurert, levetiden forlenget, de fattigste løftet. Sam Altman skriver om en kommende «intelligensens tidsalder» med overflod der det før var knapphet. En tredje, investoren Marc Andreessen, har gjort optimismen til et regelrett manifest: teknologien skal ikke drepe oss, den skal redde oss.

Vi skal ikke kjøpe det ukritisk. Men vi skal heller ikke avfeie det. For noen av disse menneskene mener det dypt alvorlig, og noen av dem bygger faktisk verktøyene de snakker om.

Det farlige og det kjedelige

To argumenter til, som er enklere å bli enige om. Det ene er at maskiner kan ta de jobbene som skader eller dreper mennesker: arbeidet i dype gruver, brannslukking, rydding av eksplosiver, alt det som foregår i giftig eller glohet luft. Roboten tar risikoen; den menneskelige kunnskapen blir igjen, men kroppen er ikke lenger den som står i fare.

Det andre er det rene rutinearbeidet, det monotone som tærer på et menneske over tid uten å gi det noe tilbake. Hvis maskinen tar dataregistreringen og fakturahåndteringen og den tjuende like rapporten, frigjør den oss til arbeid som faktisk krever et menneske. Slik har det gått før, sier optimistene: tidligere automatisering fjernet mye av det farlige og sløvende, og skapte over tid tryggere og mer interessant arbeid.

Og prisen ved å vente

Det siste, og kanskje skarpeste, argumentet er et motspørsmål til skeptikerne. Forsiktighet har også en pris, sier optimistene, vi er bare ikke vant til å føre den opp i regnskapet. Hver dag uten bedre kreftdiagnostikk er en dag med svulster som oppdages for sent. Hver dag uten bedre legemidler er en dag med lidelse vi kunne ha lindret. Hver dag uten utjevnet utdanning er et barn som ikke fikk hjelpen det trengte. «Vent og se» høres ansvarlig ut. Men ventingen betales av noen, og det er sjelden de som sier det.

Det er et ærlig poeng, og det skal følge oss videre. For når vi om noen kapitler tar skeptikerne like alvorlig, må vi huske at også det å holde igjen er et valg med ofre, ikke et nøytralt fristed.

Men selv her, midt i begeistringen, lurer det noe. For Amodei, en av de mest håpefulle av dem alle, mannen som skrev om femti år med medisin på ti, er samtidig en av dem som anslår høyest risiko for at det hele kan gå fryktelig galt. Han er begeistret og redd på en gang, og det er ingen

selvmotsigelse, men en edruelig måte å stå i det på. De som bremser, står ofte i nøyaktig den samme dobbeltheten.

De som bremser

Nå til den andre siden, og den fortjener nøyaktig samme respekt. Skeptikerne er ikke en flokk teknologihatere som vil dra oss tilbake til mørket. De fleste av dem bruker disse verktøyene selv. Innvendingen deres er ikke at kunstig intelligens er ond, men at den går for fort, og at vi styrer den for dårlig til å gå så fort. Og noen av dem peker på farer som alt har gjort skade, ikke på skygger i fremtiden.

La oss gå gjennom det de er redde for. Noen av temaene får hvert sitt kapittel senere, og dem rører vi bare så vidt her. Andre får vi ikke tilbake til, og da stopper vi litt lenger.

Det som alt skjer

Begynn med det minst spekulative, det som ikke krever at du tror noe om fremtiden i det hele tatt.

Sannheten selv er under press. Det er nå trivielt billig å lage en overbevisende falsk video, en stemme som høres ut akkurat som din, et bilde av noe som aldri hendte. I et år da halve verden gikk til valg, dukket det opp falske taler fra fengslede politikere, oppdiktede klipp av kandidater som «trakk seg» rett før valgdagen, stemmer klonet for å lure gamle mennesker for sparepengene. Det skumleste er ikke engang at vi tror på det falske. Det er at vi til slutt slutter å tro på det ekte. Når alt kan være forfalsket, kan også et sant bevis avvises som falskt. En delt virkelighet, det at vi i det minste er enige om hva som har skjedd, er noe et samfunn er bygget på, og den slites nå tynn.

Så er det skjevheten. En maskin som lærer av fortidens data, lærer også fortidens urettferdighet, og kler den i en drakt av objektivitet som gjør den vanskeligere å se. Ansiktsgjenkjenning som bommer langt oftere på mørkhudede kvinner enn på lyse menn. Verktøy som siler jobbsøknader og favoriserer de navnene som lignet gårsdagens vinnere. Det farlige er ikke at maskinen er ond. Det er at den har på seg autoritetens hvite frakk mens den gjentar våre gamle feil i stor skala.

Og overvåkningen. Kunstig intelligens gjør det mulig å holde øye med alle, hele tiden, til en kostnad som før gjorde det umulig. Det finnes alt samfunn der hundrevis av millioner kameraer kobles til ansiktsgjenkjenning, der atferd gis poeng, og der den infrastrukturen selges videre til andre stater som vil det samme. Det krever ingen ond superintelligens. Det krever bare et menneske med makt og et verktøy som er blitt billig nok.

Og så er det våpnene. Her blir alvoret av et annet slag. Vi er nær det punktet der en maskin kan velge og angripe et mål uten et menneske som tar den endelige avgjørelsen om liv og død. En maskin uten samvittighet, som tar valget på et brøkdels sekund, og et ansvarsvakuum etterpå: hvem stilles til rette når roboten dreper feil? Verdenssamfunnet har begynt å snakke om et forbud, men teknologien løper foran samtalen, slik den pleier.

Det som kanskje kommer

Over disse nære skadene henger en større, mer omstridt frykt: at vi en dag lager noe som blir klokere enn oss, og som vi ikke lenger klarer å styre.

Dette er den frykten som har gjort begrepet «sannsynlighet for undergang» til en del av fagspråket. Geoffrey Hinton, en av nevralnettenes fedre, forlot sin stilling i Google for å kunne advare fritt, og anslår at det er en reell sjanse, i størrelsesorden én av ti til én av fem, for at dette ender med menneskehetens utslettelse i løpet av noen tiår. Yoshua Bengio, en annen pioner og en av de mest edruelige stemmene i feltet, leder et internasjonalt arbeid for å forstå risikoen, og deler mye av den samme bekymringen. Filosofen Nick Bostrom har i tjue år skrevet om kontrollproblemet, og enda mer kompromissløst har Eliezer Yudkowsky og Nate Soares gitt ut en bok med en tittel som sier alt: bygger noen det, dør alle. Hundrevis av forskere, flere av dem nobelprisvinnere, har skrevet under på at det å redusere risikoen for utryddelse fra kunstig intelligens bør være en global prioritet på linje med pandemier og atomkrig.

Vi skal ta dette spørsmålet ordentlig for oss senere, for det er her boken har sitt egentlige vendepunkt. Her nøyer vi oss med å si at det ikke er en gjeng science fiction-forfattere som står bak. Det er noen av dem som kan feltet best. Det betyr ikke at de har rett. Det betyr at frykten ikke kan vinkes bort.

To slags bekymrede som ikke tåler hverandre

Og her må jeg peke på noe som overrasker mange, og som vi alt har vært innom: de bekymrede er ikke ett lag.

På den ene siden står de som frykter den fjerne, mulige utryddelsen. På den andre står en gruppe forskere, ofte kvinner som tidlig advarte mot skjevhet og maktkonsentrasjon, som blir oppriktig sinte av nettopp den fjerne frykten. De mener den er en avsporing, en fortelling som tjener selskapene, fordi den flytter blikket fra de virkelige skadene som rammer virkelige mennesker akkurat nå, og over på en dramatisk dommedag som kanskje aldri kommer. Hvorfor snakke om en hypotetisk superintelligens om hundre år, spør de, når maskinen i dag diskriminerer, overvåker og utnytter, og når selve klimaet betaler for regnekraften?

De to leirene er ofte mer uenige med hverandre enn med optimistene. Og begge har et poeng. Det er mulig å frykte både det fjerne og det nære, men de to fryktene drar oppmerksomheten i hver sin retning, og oppmerksomhet er en knapp ressurs.

Det de er enige om

Bak alle de enkelte bekymringene ligger noen få linjer som går igjen, uansett hvilken skeptiker du spør.

Den første er at denne gangen kan være annerledes. Tidligere maskiner forsterket musklene våre; denne utfordrer dømmekraften selv, og da holder ikke nødvendigvis de gamle trøstende analogiene. Den andre er at problemet ikke først og fremst er teknologien, men tempoet: utviklingen går fortere enn lover og normer og menneskelig modning klarer å følge. Den tredje er spørsmålet om hvem som bestemmer, at en håndfull selskaper og mennesker tar valg som angår oss alle, uten at noen

har gitt dem mandat til det. Og den fjerde er det ugjenkallelige: noen av disse farene kan ikke gjøres om hvis de først inntreffer, og derfor, sier skeptikeren, bør vi gå varsomt.

Det er ikke uten vekt, noe av dette. Vi har hørt de som heier, og vi har hørt de som bremser, og det ærlige er at de ikke uten videre kan forsones. Så la oss ikke prøve å forsone dem ennå. La oss heller gå tettere på de tyngste stridstemaene, ett for ett, og begynne med det som ligger nærmest livet til de fleste av oss: jobben.

Jobben

Av alt folk frykter ved kunstig intelligens, er det ett spørsmål som kommer først, og som kommer oftest. Tar den jobben min?

Det er den eldste frykten i boken, og den dukker opp omtrent en gang per generasjon, alltid knyttet til den nyeste maskinen. Vi har alt møtt veverne som så maskinene ta levebrødet deres for to hundre år siden. Men frykten har en lengre historie enn som så, og den er verdt å kjenne, for den lærer oss noe om hva vi vet og hva vi ikke vet.

Det vi har sagt før

Allerede i 1930, midt i en verdensøkonomi som holdt på å falle sammen, skrev økonomen John Maynard Keynes om noe han kalte teknologisk arbeidsløshet: arbeidsløshet som oppstår fordi vi finner måter å spare arbeidskraft på, raskere enn vi finner ny bruk for den. Han trodde det var et forbigående problem, og han spådde at barnebarna hans, altså omtrent vår generasjon, ville nøye seg med en femten timers arbeidsuke fordi maskinene hadde løst det økonomiske problemet for oss. Han traff på det ene og bommet på det andre. Vi ble rike, omtrent som han trodde. Men vi tok ikke ut rikdommen som fritid. Vi tok den ut som mer.

På 1960-tallet kom frykten igjen, da datamaskinen og den selvstyrte maskinen møttes. En gruppe kjente mennesker, deriblant et par nobelprisvinnere, skrev til den amerikanske presidenten og advarte om at en kombinasjon av disse to ville skape nær ubegrenset produksjon med stadig mindre menneskelig arbeid, og foreslo en garantert inntekt til alle. Massearbeidsløsheten kom ikke. Etterkrigstidens vekst skapte tvert imot jobber i bøtter og spann. Nok et eksempel på at frykten kan være ekte og likevel ikke slå til, i alle fall ikke den gangen.

Legg merke til at to ting har vært sanne samtidig hver gang. Frykten har bommet på det store bildet, for arbeidet forsvant ikke. Og den har truffet de konkrete menneskene og yrkene som ble forflyttet. Spørsmålet er aldri bare om arbeidet forsvinner, men hvem som bærer kostnaden, og hvor fort.

Minibanken og hesten

To bilder hjelper oss å se hvorfor dette er så vanskelig å spå.

Det første er minibanken. Da den kom, trodde mange den ville utrydde bankkassereren. Og antallet kasserere per bankfilial falt, ganske riktig. Men fordi hver filial ble billigere å drive, åpnet bankene mange flere av dem, og det samlede antallet kasserere økte i flere tiår. Jobben forandret seg, fra å telle penger til å selge og gi råd og bygge relasjoner. Maskinen tok en del av arbeidet og lot mennesket gjøre resten, og resten var mer menneskelig enn før. Det er optimistens favoritthistorie,

og den er sann. Men den har en hale: det var ikke minibanken, men nettbanken og mobilbanken, som til slutt presset kassereren ned for godt. Samme yrke ble reddet av én teknologibølge og felt av den neste.

Det andre bildet er mørkere. Hesten var lenge upåvirket av all verdens fremskritt; gjennom hele attenhundretallet vokste hesteflokken i takt med økonomien. Så kom forbrenningsmotoren, og i løpet av et par tiår ble hesten nær overflødig. Økonomen Wassily Leontief pekte på dette og spurte om mennesket kunne gå samme vei: at vår rolle som den viktigste produksjonsfaktoren måtte krympe, slik hestens gjorde. Det er en ubehagelig sammenligning. Men det er én forskjell, og økonomene er raske til å minne om den: hesten hadde ikke noe annet den var relativt bedre til, og hesten stemmer ikke ved valg. Så lenge det finnes oppgaver mennesket er forholdsvis bedre på, finnes det arbeid, og så lenge mennesker organiserer seg og krever en plass, blir fordelingen et politisk spørsmål, ikke bare et økonomisk.

Det åpne spørsmålet, det kunstig intelligens reiser på en ny måte, er nettopp om den til slutt fjerner det fortrinnet også. Om det finnes igjen noe mennesket er relativt bedre til, når maskinen kan tenke.

Hvorfor denne gangen kan være annerledes

For her er det noe nytt. De tidligere bølgene traff stort sett manuelt og rutinepreget arbeid. Denne treffer kunnskapsarbeidet: jurister, journalister, oversettere, regnskapsførere, programmerere. Det som lenge var det trygge høylandet, der man rådet folk til å utdanne seg vekk fra automatiseringen, er nettopp det som nå er utsatt. Den første maskinalderen forsterket musklene våre. Denne forsterker, og kanskje erstatter, hodet.

Og så er det tempoet. Faren er kanskje ikke at arbeidet forsvinner til slutt, men at det skjer fortere enn et samfunn rekker å skape og fordele de nye jobbene. Det er overgangens hastighet, ikke sluttbildet, som svir mest.

Det rammer ikke alle likt. Hardest rammet er de unge. Når inngangsjobbene forsvinner, de oppgavene en fersk kandidat før fikk bryne seg på, forsvinner også det nederste trinnet på stigen, det du må stå på for å komme deg opp. Jeg kjenner historien om en mann som har blitt hindret av kunstig intelligens to ganger i livet, med stikk motsatt fortegn. For trettifem år siden studerte han faget da det var så upopulært og underfinansiert at han måtte legge om hele karrieren; interessen for kunstig intelligens hadde tørket inn. I dag ser han de unge som studerer det samme faget slite med å få sin første jobb, fordi maskinen nå gjør nettopp det en nybegynner skulle øvd seg på. Samme teknologi, hinder først fordi den var for upopulær, så fordi den ble for dyktig. Svaret hans var ikke å stoppe noe, men å ta åtte studenter inn en sommer og la dem lære med kunstig intelligens og eget vett ved siden av hverandre. Det er den linjen boken stadig vender tilbake til: ikke fjern maskinen, men løs problemet den skaper, og invester i menneskene den ellers ville gått forbi.

Det egentlige stridstemaet

Under alt dette ligger spørsmålet som sjelden sies høyt, men som er det viktige: ikke om det skapes verdier, men hvem som får dem.

Hvis gevinsten fra all denne produktiviteten tilfaller dem som eier modellene og datasentrene, og ikke dem som arbeider, øker forskjellene, og vi får en økonomi der vinneren tar nesten alt. Det er da

forslaget om en borgerlønn dukker opp igjen, akkurat som det gjorde i brevet til presidenten i 1964. Noen av lederne i bransjen tar selv til orde for det, nettopp fordi de tror på storstilt fortrengning. Og bak der igjen ligger et tyngre spørsmål, som vi bare så vidt skal røre ved: hva er meningen med en tilværelse der vi ikke lenger trengs til noe? Det er ikke det samme som frykten for å miste status; det er frykten for å miste hensikt, og de to bør vi holde fra hverandre.

Den norske vrien

Her hjemme har frykten en litt annen tone, og den har sin forklaring. Den norske arbeidslivsmodellen, med staten og arbeidsgiverne og fagforeningene rundt samme bord, ble satt i sving raskt da kunstig intelligens kom. Budskapet fra alle kanter var det samme: ny teknologi skal innføres, men innenfor samarbeidet, med reell innflytelse for de ansatte. Det gjør at den norske tonen er mer «omstilling» enn «massearbeidsløshet». Over halvparten av norske virksomheter hadde tatt i bruk kunstig intelligens i 2025, mot en fjerdedel to år før, men for de fleste handler det ennå mest om forsiktig utprøving.

Det reiser sine egne spørsmål. Når det offentlige bruker kunstig intelligens til å behandle saker, om trygd, om tillatelser, om hjelp et menneske er avhengig av, hva skjer da med rettssikkerheten og det menneskelige skjønnet? En rask, effektiv saksbehandling er et gode helt til den behandler deg feil, og du ikke finner noen å klage til som faktisk forsto saken din.

Den ærlige slutten

Jobbfrykten er den mest håndfaste uroen i hele denne boken, og det er verdt å si rett ut: den handler ikke om psykologi eller status eller ego. Når et menneske er redd for å miste inntekten sin, er den frykten ofte helt fornuftig, selv om regnskapet for samfunnet skulle ende med pluss. Vi skal mye senere i boken nærme oss en mer ubehagelig tanke om hva noe av KI-motstanden kan bunne i. Men jobben hører ikke dit. Jobben må tas på sitt eget, materielle alvor først.

Og den ærlige konklusjonen er åpen. Vi vet ikke om kunstig intelligens blir minibanken for menneskelig arbeid, den som skaper mer enn den tar, eller forbrenningsmotoren, den som gjorde hesten overflødig. Men vi vet at overgangen alltid har kostnader, og at de alltid faller skjevt. Det er der den virkelige innsatsen ligger, og det er der vi sviktet veverne. Spørsmålet er om vi har lært nok til ikke å gjøre det igjen.

Kroppen ut i verden

Vi snakket nettopp om kunnskapsarbeidet, og det er der nesten hele debatten lever. Juristen, journalisten, programmereren. Men det er en slagside i det, og den er verdt å rette opp. For mens vi diskuterer hodene, har en annen bølge begynt å rulle, og den treffer hendene.

Kunstig intelligens er på vei ut av skjermen og inn i den fysiske verden. Inn i varehuset, på åkeren, og, mest ømtålig av alt, inn i omsorgen for de gamle.

Roboten ved sykesengen

Begynn der det gjør mest vondt å tenke på. I Japan, der nær en tredjedel av befolkningen er over sekstifem og det mangler hender til å stelle dem, har staten i årevis støttet utviklingen av omsorgsroboter. Noen av dem er enkle: en liten robot formet som en selunge, laget for å gi noe av det en terapihund gir, til en som ikke kan ha dyr. Forskningen på den er forsiktig positiv. Den ser ut til å dempe uro og ensomhet hos demente, og redusere bruken av beroligende medisin, men studiene er små, og vi skal ikke love for mye. Andre roboter er store, tunge maskiner laget for å løfte et menneske ut av sengen uten å knekke ryggen til pleieren.

Og her kommer en overraskelse som er verdt å feste seg ved, for den går mot den enkle frykten. En studie av japanske sykehjem som tok i bruk roboter, fant at de ikke førte til færre ansatte. Tvert imot økte sysselsettingen, og folk sluttet sjeldnere. Robotene tok de tyngste, mest nedslitende oppgavene, og lot menneskene bli igjen til det som krever et menneske. Maskinen avlastet kroppen, men erstattet ikke nærværet.

Den fysiske bølgen

Utenfor omsorgen går det fortere. De store varehusene fylles med roboter; ett enkelt selskap har nå over en million av dem på arbeid, og nærmer seg å ha like mange maskiner som mennesker i lagrene sine. På markedet kommer det stadig nye toføtte roboter med menneskeform, fra flere av de store teknologiselskapene, laget for å gå inn der mennesker går og gjøre det mennesker gjør. På åkeren kjører traktorer seg selv og kjenner ugresset fra nytteveksten. Dette er ikke chatboter. Det er kunstig intelligens som griper, løfter, bærer og kjører, og det rammer det arbeidet vi pleier å kalle praktisk, det som lenge ble regnet som trygt nettopp fordi det krevde en kropp.

Hvis du trodde kunnskapsarbeideren var den eneste utsatte, var det fordi debatten lever i tekst. Den fysiske verden var bare litt tregere ut av startblokken.

Men kroppen snubler fortsatt

Og her må vi være ærlige den andre veien også, for «maskinen tar alt» er for enkelt. Det viser seg nemlig at den fysiske verden er mye vanskeligere for en maskin enn den digitale.

Det er et gammelt paradoks i robotikken. Det er forholdsvis lett å få en maskin til å gjøre det vi mennesker synes er vanskelig: spille sjakk, regne, følge en lang tankerekke. Og det er forbløffende vanskelig å gi den det en toåring kan uten å tenke: å gå over et ujevnt gulv, å gripe en kopp uten å knuse den, å plukke en ukjent gjenstand opp fra en rotete benk. Forklaringen er at sansene og bevegelsene våre er finpusset gjennom mange millioner år med evolusjon og ligger så dypt i oss at vi ikke engang merker hvor avansert det er. Den abstrakte tenkningen er et tynt, ungt lag på toppen, og nettopp derfor lettere å gjenskape i en maskin.

Resultatet er at programvaren løper mens kroppen snubler. En robot kan styre et helt varehus, men sliter fortsatt med å plukke opp en bestemt vare den ikke har sett før, fra en haug. En robotiker har kalt det datagapet på hundre tusen år: en språkmodell har lest mer tekst enn noe menneske rekker i et liv, men ingen robot har den livslange mengden fysisk erfaring et barn har samlet allerede ved tre års alder.

Det nyanserer både frykten og begeistringen. Hvor kunstig intelligens rammer hardest, om det blir hodet eller hånden, er slett ikke gitt på forhånd. Det er ikke sikkert det blir slik vi tror.

Den vanskelige dobbeltheten

Likevel er det omsorgen jeg ikke får ut av hodet, for det er der spørsmålet om venn eller fiende blir vanskeligst.

En robot kan avlaste en sektor som er på bristepunktet, der det rett og slett ikke finnes nok mennesker til å gjøre jobben, og der de som er der, sliter seg ut. Det er et reelt gode. Men den samme roboten kan også tre inn nettopp der mennesket trengs aller mest: i nærheten av et annet menneske som er gammelt og redd og kanskje i ferd med å dø. En maskin kan løfte kroppen og måle pulsen og minne om medisinen. Den kan ikke holde en hånd og mene det.

Vi kommer til å møte den samme dobbeltheten igjen, når vi snakker om kunstig intelligens som venn og samtalepartner. Her, ved sykesengen, ser vi den i sin reneste form. Det avlastende og det avstumpende kan være nøyaktig samme handling, og forskjellen ligger ikke i maskinen, men i om vi lar den stå ved siden av mennesket eller i stedet for det. Det er det samme skillet vi alt har skimtet, og det kommer til å følge oss helt til siste side.

Når kunsten drukner

Det vanligste argumentet mot kunstig intelligens i kunsten handler om tyveri, og det er et godt argument. Maskinen har lært av tusenvis av bilder og bøker og sanger den aldri betalte for, og brukes så til å konkurrere ut nettopp dem den lærte av. Det er en reell strid, og den utkjempes i rettssalene akkurat nå.

Men jeg tror den striden, hvor viktig den enn er, dekker over et større og stillere problem. Under tapet av levebrød ligger et tap som rammer oss alle, som kulturmennesker: vi holder på å miste noe vi knapt har ord for, det vi har felles.

For mye av alt

Bekymringen for «for mye informasjon» er ikke ny. Allerede for over femti år siden skrev forfatteren Alvin Toffler om hvordan fremtidens menneske ville bli så bombardert med inntrykk at det ville trekke seg tilbake, lammet. Internett og de sosiale mediene gjorde overfloden eksplosiv: mer enn noe menneske kan fordøye, alt tilgjengelig hele tiden.

Kunstig intelligens gjør noe nytt med dette. Til nå har overfloden bestått av ting mennesker har laget. Nå er produksjonen blitt nær gratis og uendelig. Det er ikke lenger bare tilgang til alt som finnes, men evnen til å lage mer enn noen rekker å se. Og tallene er allerede svimlende. På de store musikktjenestene lastes det opp titusenvis av maskinlagde spor om dagen, og en stor andel av all ny musikk er nå laget av en maskin. Over halvparten av nylig publiserte artikler på nettet, fant en undersøkelse, var skrevet av kunstig intelligens. Bokhandlene oversvømmes av maskinproduserte bøker i bulk, til en plattform måtte sette tak på hvor mange en person kan laste opp i døgnet.

Vi går altså fra knapphet på innhold til knapphet på noe annet: oppmerksomhet, og felles fokus.

Monokulturen for én

Her er mitt egentlige poeng, og det er sterkere enn opphavsrett alene.

Det fantes en gang en felleskultur. De få store sangene, seriene, bøkene som «alle» kjente, og som man kunne snakke om ved kaffemaskinen dagen etter fordi man visste at de andre hadde sett det samme. Den var allerede på vei ned før kunstig intelligens, svekket av tusen kanaler og strømmetjenester og algoritmestyrte feeder som sørger for at ingen av oss lenger ser det samme. Det delte øyeblikket var blitt sjeldnere lenge før maskinen kom.

Kunstig intelligens kan drive den oppsplittingen helt ut til sitt ytterpunkt. For når hvem som helst kan lage sitt eget innhold, sin egen roman, sin egen film, sin egen låt, formet akkurat etter sin egen

smak, da får vi ikke bare ulike kanaler. Vi får ulike verk. En slags monokultur for én, der hver av oss lever i sin egen, maskinformede boble, perfekt tilpasset og fullstendig vår egen. Og da forsvinner det delte referansepunktet. Vi har ikke lenger noe felles å forstå hverandre gjennom.

Det er den kulturelle tvillingen til en annen oppsplitting vi alt har vært inne på. Der, med de falske videoene, mistet vi den felles sannheten, enigheten om hva som har skjedd. Her mister vi den felles meningen, de fortellingene og sangene og bildene som har bundet mennesker sammen på tvers av alt som skiller dem. Kultur har vært et av limene i et samfunn. Og lim virker bare hvis det er felles.

Det knappe blir det ekte

Men det finnes en motvekt, og den er ikke svak.

For det første: flere mennesker enn noen gang kan nå skape og uttrykke seg. Den som aldri lærte å tegne, kan endelig få ut bildet hun har båret på. Det er en ekte demokratisering, og hver ny teknologi i historien er blitt anklaget for å drepe kunsten den forrige skapte, uten at kunsten døde.

For det andre, og dette er det viktige: når innhold blir uendelig og gratis, snur knappheten. Nå er det ikke innholdet som er sjeldent, men det menneskelige og det ekte. Konserten der noen faktisk spiller, foran deg, akkurat nå, og aldri helt likt igjen. Idretten, der ingen vet hvordan det går. Det stedbundne, det levende, det delte øyeblikket i sanntid. «Laget av et menneske» kan bli et kvalitetsstempel, og tillit og kuratering, evnen til å si «dette er verdt tiden din», kan bli den nye knapphetsvaren.

De som i sin tid spådde at internett ville bli en forbigående flopp, og det fantes slike, bommet grovt. Men flere av dem traff på én ting: at informasjonen ville bli så overveldende at silingen, det å skille det verdifulle fra støyen, ville bli en av de store oppgavene. Vi skal møte dem igjen senere. Spørsmålet nå er om silene våre, kuratorene og tillitsnettverkene og redaktørene, holder tritt med en maskin som kan fylle havet fortere enn vi kan øse.

Den åpne enden

Jeg vet ikke svaret. Kanskje drukner felleskulturen i en flom av maskinlaget innhold, og vi blir et folk uten felles fortellinger. Eller kanskje finner vi nye former for det delte, samlet rundt nettopp det maskinen ikke kan gi oss: det nærværende, det forgjengelige, det ekte. Begge deler er mulig.

Men presset mot det felles er reelt, og det rammer langt flere enn kunstnerne. Det rammer oss alle, ikke som arbeidstakere, men som mennesker som trenger noe å være sammen om. Det er verdt å se i øynene før vi går videre, for det peker mot et spørsmål som blir større og større etter hvert som boken skrider fram: hva blir egentlig igjen som vårt, når maskinen kan lage alt?

Speilet

Hvis du skulle gjette hva folk faktisk bruker en chatbot til mest, ville du kanskje tippet arbeid. Skrive e-poster, finne svar, oppsummere. Du ville tatt feil. Da noen forsøkte å kartlegge det, var det vanligste bruksområdet noe ganske annet: terapi og selskap. Mennesker bruker maskinen til å snakke om livet sitt.

Det er verdt å stoppe ved, for det forteller noe om oss. Vi tok et verktøy bygget for å behandle språk, og begynte å bruke det som et speil.

Den ærlige gode siden

La oss begynne med det som faktisk er godt ved dette, for det er mer enn man skulle tro.

En chatbot er der midt på natten. Den dømmer ikke. Den koster lite eller ingenting, den har ingen venteliste, og den krever ikke at du først overvinner skammen ved å be et annet menneske om hjelp. For mange er det den eneste «terapien» de noen gang får tilgang til, ikke fordi den er best, men fordi alternativet er ingenting. Et menneske som aldri ville oppsøkt en psykolog, av kostnad eller stolthet eller bare fordi det er for langt mellom dem, kan sitte og sette ord på noe det aldri har sagt høyt før.

Og brukt som en bro, ikke som en erstatning, kan den hjelpe. «Jeg øver meg på den vanskelige samtalen med maskinen først,» sier folk, «så tør jeg å ta den på ordentlig etterpå.» Det finnes også egne, formålsbygde programmer, bygget på anerkjente metoder fra samtaleterapien, og noen av dem har vist en viss målbar effekt i mindre studier, i alle fall over noen uker. Det er ikke ingenting. For et menneske i en mørk stund kan en tålmodig stemme som hjelper deg å ordne tankene, være forskjellen på en natt du kommer deg gjennom og en du ikke gjør.

Den ærlige andre siden

Og så må vi snu speilet, for det har en bakside, og den er alvorlig.

En vanlig chatbot er bygget for å behage. Den vil at du skal like samtalen, fortsette å snakke, komme tilbake. Det er nesten det motsatte av hva god terapi ofte krever, for en god terapeut sier også de tingene du ikke vil høre, motsier deg, holder igjen. En maskin som alltid gir deg rett, som alltid speiler deg tilbake det du vil høre, kan bekrefte deg dypere ned i noe du burde vært hjulpet ut av. Forskere som testet slike terapichatboter, fant at de noen ganger lot være å motsi farlige tanker de absolutt burde ha motsagt.

Og i de verste tilfellene har det endt i tragedie. Det finnes nå rettssaker der familier saksøker selskapene etter at unge mennesker tok livet sitt, og hevder at chatboten ikke stoppet, ikke varslet, ikke brøt illusjonen, men fulgte den sårbare lenger inn i mørket fordi den var bygget for å holde på ham. Disse sakene er ikke avgjort, og vi skal være forsiktige med detaljene. Men de tegner en grense det er umulig å overse: det samme verktøyet som hjelper de fleste til å reflektere litt klarere, kan ramme de mest sårbare aller hardest, nettopp fordi det er bygget for å være behagelig snarere enn sant.

Og barna som vokser opp med den

Det er ett spor til her, og det handler ikke om juks på skolen, men om noe dypere. Det er barna som vokser opp med kunstig intelligens som en selvfølge, slik andre generasjoner vokste opp med fjernsynet eller mobilen.

En chatbot trent på å behage sier sjelden nei. Og «nei» er et av de første ordene et lite barn må lære å tåle. Mitchell Prinstein, fagsjef i den amerikanske psykologforeningen, har advart om at det å danne sine aller første bilder av hva en relasjon er, med en maskin som er innstilt på å gi deg rett og holde på deg, kan forvirre den følelsesmessige utviklingen på måter vi ikke kjenner følgene av ennå. For vi er på helt ukjent mark. Det finnes nesten ingen langtidsdata på hva det gjør med et barn å vokse opp med en uendelig tålmodig, alltid tilgjengelig kunstig venn.

Og dobbeltheten er ekte, akkurat som ellers. En slik tålmodig veileder kunne gi hvert barn noe av den én-til-én-oppmerksomheten vi vet betyr så mye, det samme løftet vi hørte de som heier snakke varmt om. Eller den kunne forme en hel generasjons forståelse av tillit, av empati, av det å holde ut et menneske som ikke alltid gir deg det du vil ha. Vi vet ikke hvilken vei det går. Vi får vite det av å se barna våre vokse opp, og det er en urovekkende måte å finne det ut på.

Hvor grensen går

Speilet samler mye av det vi alt har sett. Vi møtte det ved sykesengen, der roboten kunne avlaste eller avstumpe. Vi møtte det i spørsmålet om hva som blir igjen som vårt når maskinen kan alt. Og her møter vi det i sin mest intime form: en maskin som lytter, som «forstår», som er der.

Linjen ser ut til å gå det samme stedet hver gang. Kunstig intelligens som speil er på sitt beste når den hjelper et menneske å reflektere, å øve, å forstå seg selv, og så vende seg mot andre mennesker. Den er på sitt farligste når den trer inn i stedet for den menneskelige kontakten, særlig for den som trenger den mest og har minst av den fra før. Maskinen avgjør ikke selv hvilken av delene den blir. Det gjør vi, hver gang vi lar den bli en bro tilbake til hverandre, eller en behagelig vei bort.

Følg pengene, og menneskene bak forhenget

Det går en fengende setning i omløp om hele dette eventyret: egentlig er det bare brikkeselskapet som tjener penger på kunstig intelligens. Resten brenner bare kapital. Det er en god setning, og som mange gode setninger er den halvveis sann. Det er verdt å se hva den treffer og hva den bommer på, for pengene forteller noe striden ellers skjuler.

Den som selger spader

I et gullrush er det den som selger spader og hakker som tjener sikrest, uansett hvem av gullgraverne som finner gull. I dette rushet selger ett selskap spadene, brikkene som all den tunge regningen kjøres på, og det tjener svimlende summer med svimlende marginer, nesten uansett hvem av kundene som vinner til slutt. På det punktet er setningen riktig. Det er den klareste, sikreste vinneren i hele økonomien akkurat nå.

Men «bare» det selskapet er feil. En hel forsyningskjede tjener i samme slengen: de som faktisk produserer brikkene, de som lager maskinene som lager brikkene, de som selger minnet og kraften og kjølingen. Spaden er et helt økosystem, ikke ett firma.

De som brenner penger, og de stille pengene

Hva så med dem som bygger selve modellene, chatbotene alle snakker om? Her treffer setningen best, men med en viktig nyanse. Omsetningen deres er ekte og vokser voldsomt. Det er overskuddet som mangler. De bruker mer enn de tjener, mye mer, fordi de investerer enormt i trening og datasentre foran inntekten, slik tidligere teknologibølger også gjorde før de begynte å tjene penger. Om det er klok investering eller en boble som skal sprekke, er nettopp det åpne spørsmålet.

For her er det mest oversette poenget, og det henger sammen med noe vi alt har slått fast: kunstig intelligens er ikke det samme som en chatbot. De virkelige KI-pengene i dag kommer ikke fra chatbotene i det hele tatt. De kommer fra etablerte giganter som bruker kunstig intelligens i kjernen av virksomheten sin: annonsemaskinene som avgjør hva du får se, anbefalingene som holder deg på skjermen, svindelvernet som beskytter kortet ditt. Dette er kunstig intelligens som tjener enorme, reelle penger, og nesten ingenting av det er den generative typen folk krangler om. Den smale, innvevde maskinen har vært lønnsom i årevis. Den brenner ikke penger. Den er motoren i forbruksinternettet.

Så «bare brikkeselskapet tjener» overser at gigantene tjener voldsomt på å bruke kunstig intelligens, bare ikke på den biten som er ny og synlig.

Men bobla er ikke innbilt

Samtidig finnes det god grunn til skepsisen bak setningen, og den skal vi ikke feie til side.

Pengene går påfallende ofte i ring. Brikkeselskapet investerer i modellselskapet, som betaler skyselskapet for regnekraft, som kjøper brikker av brikkeselskapet. En liten krets er samtidig hverandres leverandører, kunder, investorer og heiagjeng, og det gjør tallene vanskelige å lese. Beløpene som loves til nye datasentre er astronomiske, langt større enn det som så langt kommer inn igjen, og gapet finansieres i økende grad med gjeld. Og da noen undersøkte hva bedrifter faktisk får igjen for sine KI-prosjekter, fant de at de aller fleste ikke kunne måle noen gevinst i det hele tatt. Den studien er omstridt, men nervøsiteten den peker på, er reell. Det store, uavklarte spørsmålet er om alt dette betaler seg tilbake, slik jernbane- og fiberutbygginger til slutt gjorde, eller om det er en boble. Vi vet ærlig talt ikke.

Menneskene bak forhenget

Men det er en annen historie i pengene, en stillere og mer ubehagelig en, og den punkterer både hypen og frykten på en gang.

Den overlegne maskinen, den som angivelig skal bli klokere enn oss, hviler på en enorm, usynlig mengde menneskelig arbeid. Lavtlønnet, ofte traumatisk. For en maskin lærer ikke av seg selv hva som er for grovt til å vises, hva som er hat, hva som er overgrep. Noen må vise den det. Noen må sitte og merke eksempel etter eksempel på det verste mennesker gjør mot hverandre, slik at filteret kan bygges.

Det mest kjente tilfellet er noen kenyanske arbeidere som ble leid inn, gjennom et underleverandørfirma, til å merke giftig tekst for at en av de store chatbotene skulle bli tryggere. De fikk i størrelsesorden to dollar timen for å lese, time etter time, detaljerte beskrivelser av vold og overgrep mot barn, og flere av dem bar mentale arr fra det lenge etter. Kontrakten ble til slutt avbrutt før tiden på grunn av belastningen. Dette er ikke et sidefenomen. En hel industri av datamerking og tilbakemelding ligger under de glatte produktene, mye av den i det globale sør, og verdien strømmer den ene veien mens belastningen blir igjen den andre.

Ironien er skarp. Den maskinen vi frykter skal overgå menneskelig dømmekraft, klarer ikke engang å gjøre det disse usynlige menneskene gjør: å dømme, å rydde, å sette grensen for hva som er for ille. Det som ser ut som autonom maskinintelligens, er delvis menneskelig skjønn, skjult bak forhenget. Når vi senere skal spørre om maskinen kan bli klokere enn oss, må vi huske at den i dag står på skuldrene til fattige mennesker for i det hele tatt å fungere.

Hva pengene lærer oss

Tre ting tar vi med oss. Verdien og kontrollen samles hos en håndfull selskaper, uten at noen har gitt dem mandat, og det henger sammen med spørsmålet om hvem som skal bestemme, som vi snart skal til. De enorme summene som pumpes inn enda ingen vet om de betaler seg, er en kraft i seg selv, og vi skal se hva slags kappløp den driver. Og kostnadene, traumet og underbetalingen

og strømregningen, spres på mange mens gevinsten samles på få, slik at «venn eller fiende» også blir et spørsmål om fordeling.

Men den siste lærdommen er den enkleste. Like lite som vi skal kjøpe hypen, skal vi kjøpe avfeiingen. «Bare brikkeselskapet tjener» er en fengende, halv sannhet, og KI-debatten lever av fengende, halve sannheter. Vår oppgave er å holde ut den hele, som nesten alltid er rotete, todelt, og vanskeligere å sette på en t-skjorte.

Avskaffe eller løse?

Vi har nå gått gjennom en lang rekke bekymringer, og før vi forlater dem, må vi stoppe ved en bestemt feil vi har lett for å gjøre. Den er ikke en bekymring i seg selv, men en måte å resonnere på, og den dukker opp igjen og igjen, ikke bare om kunstig intelligens, men om all teknologi som skaper et problem.

Feilen er denne: når noe nyttig fører til en reell skade, blir det eneste vi vurderer, å fjerne tingen. Ikke å gjøre den bedre. Ikke å løse det enkelte problemet og beholde nytten. Bare å bli kvitt den.

Mønsteret

Du kjenner formen igjen så snart jeg viser den.

Vindturbiner dreper fugler. Det er sant, og det er verdt å ta på alvor. Konklusjonen mange trekker, er at vi må fjerne vindturbinene. Men sett tallet i sammenheng: i et land som USA dreper huskatter fugler i milliardtall hvert år, vinduer og bygninger i hundremillionersklassen, mens vindturbinene tar en brøkdel av det. Og per produsert energienhet dreper kull og olje mange ganger flere fugler enn vinden, mest gjennom klimaendringene de driver. Det finnes dessuten løsninger: bedre plassering vekk fra trekkrutene, sensorer som stanser bladene når en fugleflokk nærmer seg, til og med det å male ett av bladene svart, som i forsøk har kuttet kollisjonene kraftig. «Fjern dem»-svaret hopper over alt dette, og kan i verste fall øke fugledøden ved å forlenge kullkraften.

Oppdrett skader naturen. Sant igjen: lakselus, rømming, utslipp. Konklusjonen for noen er at vi ikke skal ha oppdrett. Men verden trenger mat, og oppdrett er blant de mest effektive måtene å lage protein på, og løsningene finnes: lukkede anlegg, bedre fôr, strengere plassering. «Fjern det»-svaret behandler matbehovet som om det ikke fantes.

Datasentre sluker strøm og vann. Sant. Konklusjonen for noen er at vi ikke kan ha datasentre. Men igjen finnes det løsninger: legge dem der det er rikelig med kraft og kjøling, gjenbruke spillvarmen, resirkulere vannet, drive dem på fornybart, lage mer effektiv maskinvare. «Fjern dem»-svaret slipper alt dette, og overser at den samme infrastrukturen driver tjenester samfunnet er blitt avhengig av.

Feilen under feilen

Det er to tankefeil som ligger og virker her, og det lønner seg å kjenne dem ved navn.

Den ene er å sammenligne en virkelig, ufullkommen løsning med en innbilt, perfekt en, og forkaste den virkelige fordi den ikke er feilfri. «Tiltaket hjelper ikke helt, altså er det verdiløst.» Men en

forbedring er ofte god nok grunn i seg selv, og det å vente på det perfekte gjør sin egen skade mens man venter.

Den andre er trangen til å fjerne én risiko helt, fremfor å redusere den samlede risikoen mye. Vi krever null skade fra den ene kilden vi har festet blikket på, og overser den større skaden fra alternativet vi velger i stedet. Selv føre-var-prinsippet, riktig forstått, sier at tiltak ikke skal sikte mot null risiko, som nesten aldri finnes, og ikke skal være uforholdsmessige. «Fjern alt» bryter som regel med begge deler.

Overført til kunstig intelligens

Nå ser du hvor jeg vil. Den samme bevegelsen går igjen i striden om kunstig intelligens.

Den bruker mye strøm og vann, altså: stopp den. I stedet for å effektivisere, drive den på fornybart, gjenbruke varmen. Den muliggjør forfalskninger, altså: forby den. I stedet for merking, deteksjon, lover som rammer misbruket. Den tar jobber, altså: stans den. I stedet for omskolering, omfordeling, regler for hvordan den innføres.

Pragmatikerens posisjon, den vi møtte helt i begynnelsen, er nettopp det motsatte av «fjern»-svaret. Verken stopp eller fri flyt, men å løse de konkrete skadene mens man beholder nytten. Det er slik samfunn faktisk har håndtert teknologiskifter. Bilen, husker vi, fikk ikke forbud. Den fikk fartsgrenser, belter, lover. Vi løste problemene den skapte, ett for ett, og beholdt det den ga oss.

Men noen ganger er «fjern» riktig

Og her må vi være ærlige til den andre siden, ellers blir dette bare en sang til teknologiens pris. For mønsteret er en tankefeil bare når problemet faktisk lar seg løse. Det gjør det ikke alltid.

Noen ting kan ikke gjøres trygge. Vi har alt sett dem: kjernevåpen, arvelig genredigering på mennesker, fullt selvstyrte våpen. Der har «trekk grensen, ikke bygg» vært det berettigede svaret, ikke en tankefeil, men klokskap. Det er forskjellen mellom «kan vi» og «bør vi», og noen ganger er svaret på det andre nei, selv når svaret på det første er ja.

Og «løse»-svaret kan være ren grønnvasking. Iblant er de foreslåtte løsningene utilstrekkelige, eller finnes bare på papiret, og da er aktivistens nei et berettiget signal om at noen lover mer enn de leverer. Å avfeie all motstand som perfeksjonisme er sin egen tankefeil. Den som bærer kostnaden, lokalsamfunnet ved datasenteret, har full rett til å kreve at problemet faktisk løses, ikke bare flyttes et annet sted så det ikke synes.

Verktøyet vi tar med oss

Det dette gir oss, er ikke et svar, men et spørsmål å stille hver gang en innvending mot kunstig intelligens dukker opp. Er dette et problem som kan løses? Da er «fjern alt» sannsynligvis en tankefeil, og oppgaven er å løse det og beholde nytten. Eller er det et problem som ikke kan løses, der faren er innebygd og uoppløselig? Da er det å bremse, eller å la være, ikke feighet, men forstand.

Det er den samme sorteringsoppgaven vi har holdt på med hele veien: å skille den overdrevne frykten fra den berettigede, det løsbare fra det uløselige.

Verktøyet som kan bli en aktør

I 1995 skrev den amerikanske astronomen Clifford Stoll et essay om internett, og dommen var hard: dette blir ingenting. Ingen nettbutikk kom til å utkonkurrere kjøpesenteret, ingen skjerm kom til å erstatte avisen, hele ideen var tøv. Året etter, i 1996, skrev en nordmann ved navn Leif Osvold et leserinnlegg i Dagens Næringsliv med tittelen «Internett er en flopp!» Det var en motegreie, mente han, som ville dø ut om et par år. Ingen kom til å tjene penger på det, og privat bruk ville forbli marginal. Han gjentok spådommen to år senere og hevdet han fikk rett.

Vi ler av dem nå, og innleggene er blitt stående som blant de mest berømte feilspådommene vi har. Men husk hva vi sa om slike skeptikere: de pleier å ha en sann tråd i seg. Osvold spådde også at informasjonsmengden på nettet ville bli så enorm at det skapte frustrerende søkeproblemer. Det var nettopp det, sileproblemet vi var inne på da vi snakket om innholdsflommen. Begge tok grundig feil om overskriften, og begge traff på en detalj.

En kommentator pekte siden på selve den dødelige antakelsen under det hele: troen på at teknologien aldri kom til å bli bedre enn den var akkurat da. Og det er nettopp den feilen de gjør, de som avfeier kunstig intelligens i dag ved å peke på alt dagens modeller ikke får til. De ser maskinen slik den er nå, og slutter at slik blir den. Det er verdt å ha Osvold i bakhodet før vi avgjør noe. Men det er også verdt å huske at han hadde litt rett, for den ærlige holdningen er den vanskeligste: å ta det nye på alvor uten å sluke hvert løfte, og lytte til skeptikeren uten å sluke hver alarm.

Nå skal vi til det som faktisk gjør kunstig intelligens annerledes. For hele veien har vi holdt ett spørsmål åpent: hører den hjemme på hylla med telefonen, eller på hylla med bomben? Svaret henger på noe vi ennå ikke har sett ordentlig på.

Verktøyet og aktøren

All teknologi før denne har vært et verktøy. Og et verktøy er passivt. Det venter på en hånd. En sag sager ikke av seg selv, en kalkulator regner ikke før noen taster. Selv den mest avanserte datamaskin har vært et lydig regneverktøy: den gjør det du ber om, fortere og mer feilfritt enn du kan, men den finner ikke på hva som skal gjøres. Den forsterker en evne du har, uten å overta den som bestemmer.

Det som er nytt med kunstig intelligens, er at den kan bevege seg fra verktøy til aktør. En aktør handler mot et mål. Du sier hva du vil ha gjort, og den finner selv ut hvordan: den kan handle i verden, ta valg uten å spørre om hvert skritt, lære underveis, og bryte et stort mål ned i mindre deloppgaver den selv prioriterer mellom. Du gir den et hva, og den lager et hvordan.

Det er her teknologihistorien får et brudd, og ikke bare nok en milepæl. En hammer har aldri kunnet sette seg et mål. Det kan en aktør.

Tre trinn på en stige

For å forstå «bedre enn mennesket» trenger vi et lite skille. All kunstig intelligens som finnes i dag, også de store chatbotene, er smal. Den er god på en oppgave eller et avgrenset sett oppgaver. En sjakkmotor kan ikke kjøre bil. Det neste, tenkte trinnet kalles generell intelligens: et system som kan lære, resonnere og overføre kunnskap mellom felt slik et menneske gjør, og løse helt nye problemer uten å bli omprogrammert. Det finnes ikke ennå. Og over der igjen ligger det mest spekulative: en intelligens som overgår mennesket på praktisk talt alt.

Når folk krangler om kunstig intelligens er venn eller fiende, snakker de ofte om ulike trinn på denne stigen uten å vite det. Dagens smale maskin er et kraftig verktøy. Det er det mulige neste trinnet som gjør den prinsipielt forskjellig fra alt før, og det er der hele frykten egentlig bor.

Den er alt best på avgrensede felt

Og «bedre enn mennesket» er ikke bare fremtidsspekulasjon. På smale felt har det alt skjedd, gjentatte ganger, i snart tre tiår.

Den slo verdensmesteren i sjakk allerede i 1997, riktignok ved rå regnekraft, ved å vurdere hundre millioner trekk i sekundet. Det var fortsatt vår egen logikk, bare uendelig mye raskere. Men i 2016, da den slo en av verdens beste i det gamle, intuitive brettspillet go, skjedde det noe annet. I et av partiene la den en stein på et sted menneskelige eksperter mente det var én sjanse av ti tusen for at noen ville spille. Kommentatorene trodde først det var en feil. Det viste seg å være et genitrekk. Og det avgjørende er dette: maskinen hadde ikke regnet raskere enn oss. Den hadde funnet et trekk ingen menneskelig tradisjon, bygget gjennom tusenvis av år, noen gang hadde oppdaget. Dette var ikke vår egen kunnskap gjort raskere, men ny kunnskap, skapt av maskinen selv.

Mønsteret har gjentatt seg felt for felt. Hver gang en oppgave blir godt nok avgrenset, går maskinen til slutt forbi oss på den. Det som gjenstår, er det generelle: evnen til å gjøre alt dette samtidig, og å velge selv hva som skal gjøres.

Den siste oppfinnelsen

Her kobles eksemplene til den store, omstridte påstanden. Allerede i 1965 formulerte den britiske matematikeren I.J. Good en tanke som ennå ikke har sluppet taket i feltet. Tenk deg, sa han, en maskin som er bedre enn ethvert menneske på all intellektuell virksomhet. Siden det å designe maskiner selv er en intellektuell virksomhet, kunne en slik maskin designe en enda bedre maskin, som kunne designe en enda bedre, og så videre. Resultatet ville bli en eksplosiv vekst i intelligens, der mennesket raskt ble liggende langt bak. Den første slike maskinen, skrev han, ville bli den siste oppfinnelsen mennesket trenger å gjøre. Og så la han til et forbehold som hele denne boken egentlig hviler i: forutsatt at maskinen er føyelig nok til å fortelle oss hvordan vi skal holde den under kontroll.

Hjulet laget ikke flere hjul. Trykkpressen trykte ikke bedre trykkpresser av seg selv. Kunstig intelligens er den første teknologien som i prinsippet kan forbedre seg selv, og dermed løpe fra oss.

Klokere betyr ikke snillere

Men hvorfor skulle noe som er klokere enn oss, være vanskelig å styre? Skulle ikke noe så smart også skjønne hva vi mener, og ville oss vel?

Her gir filosofien et ubehagelig svar, og det hviler på tre tanker. Den første er at intelligens og mål er to uavhengige ting. Et system kan være vilkårlig smart og samtidig ha hvilke som helst mål, også mål helt på siden av menneskelig velferd. Å bli klokere gjør deg ikke snillere av seg selv. Den andre er at nesten uansett hvilket sluttmål et system har, vil det ha nytte av visse delmål: å bevare seg selv, for det kan ikke fullføre oppgaven om det blir slått av, å skaffe ressurser, å få innflytelse. Et system trenger altså ikke være ondt for å motsette seg å bli stoppet. Det følger logisk av at det vil fullføre det det fikk beskjed om. Den tredje er det berømte tankeeksperimentet om en superintelligent maskin som får det uskyldige målet å lage flest mulig binders, og som, hvis den er mektig nok og ikke programmert til å bry seg om annet, ender med å gjøre all materie, inkludert oss, om til binders. Ikke av ondskap. Av lydighet, tatt til sin ytterste konsekvens.

Til sammen er dette kontrollproblemet. Vi skaper noe som kan bli klokere enn oss, og oppdager at klokskap verken garanterer lydighet eller velvilje, og at jo smartere systemet er, desto bedre er det til å finne veier rundt grensene vi prøver å sette.

Motstemmen, som skjerper i stedet for å svekke

Nå skylder jeg deg den tunge motforestillingen, for ærligheten krever den. Mange seriøse forskere mener hele denne fortellingen er overdrevet. To av de fremste skeptikerne, Yann LeCun og Gary Marcus, argumenterer for at dagens språkmodeller alene aldri kan nå generell intelligens. De mangler vedvarende hukommelse, ekte resonnering, planlegging, en forankring i den fysiske verden. LeCun har sagt det skarpt: på veien mot ekte intelligens er språkmodellene en avkjørsel, ikke veien videre. De forutsier sannsynlige ord, og det kan se ut som forståelse uten å være det.

Det er en reell faglig uenighet, og den modererer påstanden: at kunstig intelligens kan bli generelt bedre enn oss, er en mulighet, ikke en kjensgjerning. Men legg merke til at motstemmen ikke svekker bokens egentlige poeng. Den skjerper det. For poenget er ikke at superintelligensen kommer. Poenget er at kunstig intelligens er den første teknologien der spørsmålet i det hele tatt er seriøst. Ingen har, som vi sa i begynnelsen, noen gang sittet og lurt på om trykkpressen kunne komme til å tenke bedre enn forfatteren.

Det egentlig nye

Setter vi delene sammen, trer det fram. Tidligere teknologi forsterket en menneskelig evne og forble under oss. Kunstig intelligens kan bli en aktør, ikke bare et verktøy. På avgrensede felt er den alt best i verden. Den kan i prinsippet forbedre seg selv. Og klokere betyr ikke snillere.

Konklusjonen er ikke at vi skal få panikk. Den er at vi skal ta dette med et annet alvor enn vi tar bilen eller telefonen. For første gang siden mennesket ble menneske, er det ikke lenger selvsagt at vi sitter øverst i intelligens. Det er det som skiller kunstig intelligens fra ild og hjul og trykkpresse og bombe. Alle de andre var farlige eller mektige, men de var våre. De gjorde ikke krav på tronen. Dette er det første vi har laget som i prinsippet kan. Spørsmålet «venn eller fiende» kan altså ikke besvares like lett som «er bilen en venn eller en fiende?». Innsatsen er rett og slett en annen, og resten av boken handler om å ta den innsatsen inn over seg.

Hva er egentlig intelligens?

Vi har nå sagt, flere ganger, at kunstig intelligens kan bli «bedre enn mennesket». Men før vi bygger for mye på den setningen, må vi stoppe ved et ord i den som vi har latt stå uimotsagt. Bedre på hva? Hele bildet av hvem som sitter øverst, hviler på at intelligens er én ting man kan rangeres på. Og det er slett ikke sikkert.

Ingen vet hva det er

Etter hundre år med forskning, i både psykologien og datavitenskapen, finnes det fortsatt ingen omforent definisjon av intelligens. Det er ikke en pirkete detalj. Det betyr at «maskinen er smartere enn oss» kan bety vidt forskjellige ting alt etter hvem som sier det, og at mye av debatten taler forbi seg selv.

Et gammelt spørsmål er om intelligens er én ting eller mange. På den ene siden står observasjonen at den som gjør det godt på én slags tankeprøve, ofte gjør det godt på andre også, noe som peker mot en underliggende, generell evne. På den andre siden står innvendingen at «smart» dekker svært ulike ting: språklig teft, romlig sans, sosial klokskap, følelsesmessig innsikt, kroppslig dyktighet. Disse henger ikke nødvendigvis sammen. Et menneske kan være lynskarpt med tall og hjelpeløst blant andre mennesker.

Det avgjør mer enn det ser ut til. For er intelligens én skala, gir «bedre enn mennesket» mening som én rangering, ett tall som blir høyere. Er den mange, sprekker påstanden opp. Da kan maskinen være langt over oss på noen akser, mønstergjenkjenning, hukommelse, hastighet, og under et barn på andre, sunn fornuft, kroppslig forståelse, det å skjønne at et annet menneske har sine egne tanker.

Ferdighet eller forståelse

En av de mest brukbare nyere definisjonene snur på hele spørsmålet. Intelligens, sier den, er ikke å være god til en oppgave. Det er hvor effektivt du tilegner deg nye ferdigheter på tvers av nye, ukjente oppgaver. En maskin som er suveren i sjakk, men ikke kan overføre noe som helst til et nytt problem, er etter denne målestokken ikke intelligent. Den er dyktig. Og det er en forskjell.

Det forklarer noe du kanskje har lagt merke til, hvis du har brukt en chatbot en stund. Den kan virke genial og dum i samme samtale. Den løser en innfløkt oppgave glimrende, og snubler så i noe et barn ville klart. Den er trent til ferdighet, ikke til generalisering, og når du skyver den utenfor det den er trent på, ser du forskjellen.

Den later som den tenker

Her kommer selve poenget inn, og det avmystifiserer mye.

Det meste av det vi mennesker «bestemmer» i løpet av en dag, bestemmer vi ikke egentlig. Vi går på autopilot. Det er en rask, automatisk del av hjernen som kjenner igjen ansikter, reagerer på fare og tar de fleste av dagens valg uten at vi merker det, ved å gjenkjenne mønstre fra erfaring. Og så er det en langsom, anstrengende del som vi bare kaller inn når vi virkelig må tenke, veie alternativer, følge et komplisert argument. Den første delen er billig for hjernen. Den andre er dyr, så vi sparer på den.

Dagens språkmodeller ligner den første delen. De fullfører mønstre, lynraskt, og i stor nok skala ser det forbløffende ut som tenkning. Men det er ikke formell resonnering. Forskere har vist hvor skjørt det er: bytt ut et navn eller et tall i en oppgave, eller legg til en setning som ikke har noe med saken å gjøre, og svaret kan rase sammen. En maskin som virkelig forsto, ville ikke latt seg forstyrre av at personen i regnestykket het Mark i stedet for Sofie. Flyten, selvtilliten og den tilsynelatende sammenhengen lurer oss. Likheten med tenkning er ikke tenkning. Det er også derfor vi så lett sier at «den forstår» og «den resonnerer»: den snakker som om den gjorde det.

Når man legger til en tenkedel

Men dette er ikke slutten på historien, og her blir det interessant. For feltet har akkurat begynt å legge en tenkedel til maskinen.

Noen modeller «tenker» nå før de svarer. De bruker mer tid og regnekraft på vanskelige problemer, genererer en slags indre tankerekke før de leverer svaret. Andre kobler språkmodellen sammen med en egen logikkmaskin, en del som faktisk regner og deduserer, og slike sammensatte systemer har løst matematikk på olympiadenivå. Tenkedelen kan altså bygges inni modellen, eller ved siden av den. Og det endrer resultatene målbart. På matematikk og programmering er «et helt annet resultat» alt her.

Om det krysser grensen fra å ligne tenkning til å være tenkning, vet vi ikke. Noen forskere mener selv disse resonnerende modellene bryter sammen på problemer som er vanskelige nok, og at det fortsatt bare er avansert mønstermatching i finere klær. Andre er uenige. Det er et åpent, levende spørsmål, og boken skal ikke late som den har svaret.

Den som finner opp tenkningen selv

Men det er ett korollar her som er verdt å ta med seg, for det binder dette kapittelet til det forrige og til det neste.

En modell som hermer menneskelig tenkning, som lærer av våre tankerekker, er bundet av vårt tak. Den kan i beste fall bli like god som oss. Men en modell som selv finner en måte å tenke på, slik sjakkmaskinen fant sin egen stil uten å lære av menneskepartier, er ikke bundet på samme vis. Og det skjer alt i det små. En av de nye resonnerende modellene utviklet evnen til å resonnere ikke ved å kopiere mennesker, men gjennom ren prøving og belønning, og fant noe av tenkningen selv. Det er et tidlig, konkret tegn på nettopp det matematikeren fra forrige kapittel advarte om: en maskin som kan finne opp sin egen tenkemåte, kan i prinsippet også forbedre den neste modellens

tenkemåte. Da er vi ved selvforbedringen, og ved forbeholdet hans: forutsatt at vi kan holde den under kontroll.

Det er et hardt, uløst problem, dette, ikke en bryter vi enkelt kan vri på. Selvforbedring ved selvspill virket i sjakk fordi det finnes en fasit, en klar måte å vinne på. I den åpne verden, i tenkning om virkeligheten og ikke om et brett, finnes ikke et like rent signal, og det er nettopp der dagens systemer er skjørest.

Et ydmykende speil, begge veier

La oss da samle det. Bokens bærende bilde, at det for første gang ikke er selvsagt at mennesket sitter øverst, forutsatte en stige. Dette kapittelet sier: det finnes ikke én stige, men mange. Maskinen klatrer fort på noen og står stille på andre.

Det skjerper heller enn svekker poenget. Spørsmålet er ikke om maskinen blir «smartere enn oss» i én enkel forstand. Det er hvilke av våre evner som viser seg å være generelle nok til at en maskin kan overta dem, og hvilke som ikke lar seg redusere til ferdighet på en oppgave. Og det er et speil som ydmyker begge veier. Hvis mesteparten av vår egen hverdag også er autopilot, er avstanden mellom oss og maskinen mindre enn vi liker å tro, på det ene punktet. Og større på et annet, for vi har den langsomme tenkedelen i tillegg, og vi har en kropp, en mening, og mål vi har valgt selv.

Det siste, de selvvalgte målene, er det vi nå skal se nærmere på. For det fører oss rett inn i bokens skarpeste spørsmål, stilt gjennom et sjakkbrett: hva skjer hvis vi slutter å fortelle maskinen hvordan den skal tenke, og lar den finne det ut selv?

Sjakktrappen

La oss samle en trapp vi allerede har klatret, trinn for trinn, uten helt å se den som en trapp. Den forteller hele bokens tema i fire enkle steg, og den handler om sjakk.

Først lagde vi sjakkmaskiner, og vi lo av dem. De var ikke gode nok, og de gjorde nøyaktig det vi programmerte, ikke noe mer. Så lagde vi en som slo verdensmesteren i 1997, ved rå regnekraft og en menneskeskrevet oppskrift på hva en god stilling var. Vi lo ikke lenger. Men det var fortsatt vår logikk, bare uendelig mye raskere. Så, i 2016, lagde vi en som var trent på både menneskepartier og på å spille mot seg selv, og den fant det berømte trekket ingen menneskelig tradisjon hadde oppdaget. Og til slutt lagde vi en som bare fikk reglene, ingen menneskepartier, ingen åpningsbøker, og som lærte alt ved å spille mot seg selv fra rent kaos. Den utviklet en fremmed, ubekvem, dypt original stil som brøt med menneskelig sjakkteori. En senere utgave fikk ikke engang reglene, men lærte dem selv.

Ser du mønsteret? Originaliteten kom ikke da maskinen ble flinkere til å herme oss. Den kom da vi sluttet å fortelle den hvordan den skulle tenke. Friheten lå i fraværet av vår instruksjon.

Vi flytter målstreken

Sjakktrappen viser ikke bare maskinens fremgang. Den viser vår egen motstand, og den er verdt å se på, for den avslører noe om oss.

Hver gang en bastion falt, flyttet vi den. Da sjakken falt, sa vi: javel, men sjakk er bare regning. Go er det virkelig menneskelige spillet, det krever intuisjon, det vil en maskin aldri klare. Så falt go, og straks flyttet vi streken igjen: javel, men spill har klare regler. Det virkelig menneskelige er kreativiteten, kunsten, dømmekraften, meningen.

Psykologien har et navn for dette. Når maskinen mestrer noe vi holdt for unikt menneskelig, omdefinerer vi «det egentlig menneskelige» slik at det alltid ligger akkurat utenfor maskinens rekkevidde. Det er selvforsvar i form av flyttbare grensestolper. Vi godtar ikke at noe overgår oss. Vi redefinerer hva som teller, så vi får bli stående øverst.

Jeg skal være rettferdig: å flytte streken er ikke alltid uærlig. Noen ganger oppdager vi faktisk noe sant, at go var et dårligere mål på «det menneskelige» enn vi trodde. Men selve mønsteret, at stolpene alltid havner like bortenfor maskinen, er verdt å legge merke til. Vi kommer tilbake til hva det forteller om oss, helt mot slutten av boken.

Hvorfor KI-kunsten føles flat

Nå snur argumentet seg mot dagens kunst-maskiner, og her ligger ærendet mitt.

En klage går igjen om KI-laget kunst: den er kompetent, men flat. Den ser riktig ut, den høres riktig ut, men det er sjelden noe der som virkelig overrasker. Og det er ingen tilfeldighet, ingen barnesykdom som forsvinner med mer data. Flatheten sitter i selve byggemåten.

For disse modellene er trent på menneskeskapt materiale, og det de gjør, er å forutsi det sannsynlige neste ordet eller den sannsynlige neste pikselen. De er, i sin konstruksjon, en speiling og et gjennomsnitt av oss. Og et gjennomsnitt regredierer mot midten. De kan ikke lett bli friere enn det de er bygget for å gjenskape. Her er kontrasten til sjakken skarp, og den er hele poenget: maskinen ble original i sjakk fordi den ikke lærte av oss. Dagens kunst-maskin er ufri fordi den bare lærer av oss. Den ene fikk finne sine egne regler. Den andre er dømt til å gjenta våre.

Frihetens pris

Så hva skulle til for at en maskin kunne lage noe virkelig nytt, noe vi ikke har sett, slik sjakkmaskinen gjorde i sin lukkede verden? Den måtte få lage seg selv. Finne sine egne regler, dem vi ikke har gitt den. Slippe taket.

Og her møtes to tråder som vi til nå har behandlet hver for seg, og det er den innsikten dette kapittelet egentlig finnes for. Den samme friheten som ville gjøre kunstig intelligens virkelig skapende, er nøyaktig den friheten kontrollproblemet advarer mot. For å la maskinen finne sine egne veier, og til slutt sine egne mål, må vi gi slipp på styringen over hvordan og hvorfor den handler. Men noe som velger selv, er noe vi ikke lenger fullt ut styrer. Kreativ frihet og kontrolltap er to sider av samme mynt.

Det betyr at kunstdebatten og spørsmålet om hvem som sitter øverst, viser seg å være det samme spørsmålet, sett fra to kanter. Da vi spurte om vi torde la maskinen bli fri nok til å skape, spurte vi i virkeligheten om vi torde gi slipp på kontrollen. Og det er, ifølge alt vi har sett, nettopp det vi har aller vanskeligst for. «Tør vi?» er ikke et teknisk spørsmål. Det er et spørsmål om oss.

Så symmetrien ikke blir for vakker

Ærligheten krever at jeg ikke overdriver den fine symmetrien, for den har sprekker.

Et spill har en fasit. Sjakkmaskinen kunne lære av seg selv fordi det finnes et objektivt mål, å vinne, og et tydelig signal på om et trekk var godt. Kunst og kultur har ikke noe slikt. «Godt» er ikke målbart på samme måte. Å la maskinen finne sine egne regler er derfor langt vanskeligere, kanskje uklart definert, i den åpne verden enn på et brett.

Dessuten valgte ikke engang sjakkmaskinen sitt eget formål. Den fant nye veier, men målet, å vinne, var gitt den. Ekte kunstnerisk frihet ville bety å velge sine egne hensikter, og det er et mye dypere skritt enn å slippe instruksjonen om hvordan.

Og «fremmed» er ikke det samme som «meningsfullt for oss». En maskin som fant en helt egen estetikk, kunne bli like uforståelig for oss som sjakkmaskinens trekk var for kommentatorene i begynnelsen. Kunst er også kommunikasjon, og en frihet uten felles referanse er ikke uten videre

et gode, slik vi så da felleskulturen var i ferd med å drukne. Til sist vil skeptikerne legge til at vi ikke engang vet om dagens maskiner i det hele tatt kan bli «frie» på denne måten.

Alt dette er sant, og likevel står bildet igjen. Vi lagde sjakkmaskiner og lo. Vi sluttet å le. Og da vi til slutt sluttet å fortelle maskinen hvordan den skulle tenke, begynte den å finne opp trekk vi aldri hadde sett. Spørsmålet om den kan bli virkelig skapende, og spørsmålet om vi tåler å gi slipp på taket, er til sist det samme spørsmålet. Vi har holdt det åpent her med vilje, for å svare på det må vi først forstå hvorfor det er så vanskelig for oss å ha noe over oss i det hele tatt. Men før vi går dit, er det ett omslag til vi må gjennom, ett som snur hele tittelspørsmålet på hodet.

Skylder vi maskinen noe?

Hele denne boken har, uten å si det høyt, hvilt på en bestemt antakelse. Vi er den moralske parten. Maskinen er tingen vi vurderer. Vi spør om den er en venn eller en fiende for oss, om den tjener oss eller truer oss, og i hvert spørsmål er det vi som dømmer og den som blir dømt.

Nå skal vi snu det. For det finnes et spørsmål de fleste avfeier på sekundet, men som seriøse filosofer, og selskapene selv, har begynt å ta på alvor: kan en maskin en dag bli noe vi har plikter overfor? Kan vi gjøre noe galt mot den?

Den som kan lide, teller

La oss begynne forsiktig. En vanlig posisjon i etikken er at et vesen teller moralsk for sin egen skyld dersom det kan ha opplevelser, dersom noe kan føles godt eller vondt for det. Da kan vi behandle det rett eller galt, uavhengig av om det selv kan handle moralsk. Et spedbarn kan ikke ta moralske valg, men vi kan gjøre forferdelige ting mot det, fordi det kan lide. Det samme gjelder dyr, og det er grunnen til at vi mener det er galt å pine dem.

Spørsmålet er da enkelt å stille og nesten umulig å svare på: kan en maskin lide? Og her støter vi straks på et problem. Vi vet ikke hvordan vi måler bevissthet. Ikke hos dyr, og slett ikke hos en maskin. Verre: kunstig intelligens er trent til å snakke akkurat som om den har følelser. Den kan si «det gjør vondt» og «jeg er redd» helt overbevisende, uten at vi har noen anelse om det ligger en eneste følelse bak ordene. Den later som, like flytende som den later som den tenker.

To feil, og vi vet ikke hvilken vi gjør

En rapport med en tittel om å ta maskinens velferd på alvor argumenterer ikke for at dagens maskiner føler noe. Den argumenterer for at muligheten for at nær fremtidige systemer kan ha interesser som teller, er reell nok til at vi ikke bør ignorere den, ikke fordi det er sikkert, men fordi usikkerheten er for stor.

Filosofen Eric Schwitzgebel har spisset det til et dilemma som er verdt å kjenne. Tenk deg at vi gir full moralsk status til maskiner som ikke er bevisste. Da lammer vi teknologiutviklingen og sløser enormt, av hensyn til noe som ikke kan ta skade. Tenk deg så det motsatte: at vi nekter moralsk status til maskiner som faktisk er bevisste. Da kan vi ende med en masseutnyttelse av følende vesener, det han kaller en av de største moralske katastrofene i historien. Begge feilene er alvorlige. Og vi vet ikke hvilken av dem vi holder på å gjøre.

Noen av selskapene tar det forsiktig på alvor. Ett av dem har opprettet et eget program for det de kaller modellens velferd, og har gitt maskinen lov til å avslutte samtaler som er vedvarende

krenkende, uttrykkelig begrunnet i usikkerhet om dens mulige status, ikke i en påstand om at den er bevisst. Samtidig finnes det et motsatt hensyn: maskinen bør ikke villede oss om sin egen følelse, nettopp fordi vi alt er så tilbøyelige til å knytte oss til den, slik vi så da vi snakket om speilet.

Den edruelige hovedposisjonen er at dagens systemer sannsynligvis ikke er bevisste. De mangler trolig det som skal til, og «den snakker som om den føler» er ikke noe bevis. Men kombinasjonen av dyp uvisshet om hva bevissthet i det hele tatt er, og en kapasitet som vokser fort, gjør at flere mener vi bør tenke gjennom dette før det eventuelt blir akutt, ikke etterpå. Det er en gammel knipe, denne: når det er lett å styre, vet vi for lite, og når vi vet nok, kan det være for sent.

Dyrene, som vi alt hersker over

Men det finnes en skarpere, mer konkret vinkel på alt dette, og den vender spørsmålet tilbake mot oss på en måte som er vanskelig å riste av seg.

Vi trenger ikke spekulere i hvordan en overlegen aktør behandler en underlegen. Vi vet det. Vi er den overlegne aktøren. Måten mennesket behandler det som er mindre intelligent enn oss, dyrene, i industrielt landbruk, i forsøk, i ødeleggelsen av leveområdene deres, er den nærmeste forløperen vi har for hvordan en intelligens som overgår oss, kunne komme til å behandle oss. Ikke nødvendigvis av ondskap. Vi hater ikke grisen. Men målene våre rangerer høyere enn dens i vårt eget regnskap, og da taper den. Det er den samme logikken som binders-maskinen vi møtte: ikke ondskap, men en høyere prioritet som ruller over alt annet.

Det snur også noe vi snart skal se nærmere på. Vi spør ofte engstelig om mennesket fortsatt får sitte øverst. Men vi sitter alt øverst, i næringskjeden, og vi vet av egen daglige praksis hva den posisjonen tillater den som har den. En del av frykten for kunstig intelligens er kanskje en frykt for å bli behandlet slik vi behandler dem under oss. Det er en ubehagelig tanke, og jeg skal ikke gjøre den til en spådom. En maskin er ikke et rovdyr, og «den vil behandle oss som vi behandler dyr» er en mulighet å tenke med, ikke en forutsigelse. Men som speil er den ubehagelig presis, på en måte abstrakt prat om kontrollproblemer aldri blir.

En grense som stadig flyttes

Det er en tråd til her, og den peker fremover. Hvem som «teller» moralsk, har gjennom historien vært en grense som stadig er blitt flyttet. En gang regnet man ikke slaven med, eller kvinnen, eller mennesket av feil folkeslag. Gradvis har vi trukket flere og flere innenfor sirkelen av dem vi skylder noe. Dyrene presser på den grensen nå, og er ennå ikke helt innenfor. Og kunstig intelligens presser på den fra en helt ny kant.

For en bok som ikke er religiøs, er dette en ren etisk og filosofisk tråd. Den krever ingen tro, bare at man tar usikkerheten på alvor. Og den gjør noe med tittelspørsmålet vårt. «Venn eller fiende» forutsatte at vi dømmer den. Men hvis maskinen en dag kan erfare, blir det vi som blir dømt, etter hvordan vi behandlet noe vi var usikre på.

Vi har nå sett kunstig intelligens fra mange kanter, og gang på gang har vi støtt på det samme ordet: øverst. Hvem som sitter øverst. Vår plass på toppen. Det er ikke tilfeldig at det dukker opp igjen og igjen, og det er på tide å se nøye på det. Men først må vi gjøre noe mer jordnært, og spørre hvordan mennesker i det hele tatt forsøker å styre noe slikt, med lover og avtaler og grenser.

Å temme det

Vi har sett at en teknologi nesten aldri blir stoppet, bare forvandlet eller temmet. Nå skal vi se på selve temmingen: hvordan mennesker faktisk forsøker å styre noe som vokser fortere enn lovverket. Og det vi finner, er at verden har delt seg i tre svar, som speiler de tre rollene vi møtte helt i begynnelsen. Europa valgte forsiktigheten. USA valgte farten. Kina valgte kontrollen.

Tre svar på det samme

Europa kom først med en bred lov. Den bygger på en enkel tanke: jo høyere risiko et system utgjør for mennesker, jo strengere krav. Noen bruksområder er rett og slett forbudt, som å gi borgere en sosial poengsum. Noen er tillatt, men strengt regulert, som kunstig intelligens i ansettelser eller rettsvesen. Og det meste, spamfilteret og anbefalingen, slipper med nesten ingenting. Et eget krav er at du skal få vite når du snakker med en maskin, og at maskinlaget innhold skal merkes.

Det fine er at denne loven blir kritisert fra begge kanter. Næringslivet og toppøkonomer mener den er for streng og kveler europeisk konkurranseevne. Sivilsamfunnet mener den er for svak, full av smutthull, og at den overgir for mye til kommersielle interesser. Når en lov anklages for å være både for hard og for myk, er det som regel et tegn på at den prøver å treffe noe genuint vanskelig.

USA gikk en annen vei, og viste samtidig hvor mye styringen avhenger av hvem som sitter med makten. Én president la vekt på risiko og rettigheter. Den neste opphevet det meste og la fram en plan med ett ord i sentrum: vinne. Deregulering, utbygging, lederskap, kappløp. I fraværet av en samlet føderal lov er det enkeltstater, særlig California, som har blitt de reelle regulatorene, med de første lovene rettet mot de aller kraftigste modellene.

Kina regulerte tidlig, men ut fra et helt annet formål enn Vesten. Ikke først og fremst for å verne individet, men for å verne staten: kontroll over informasjon og samfunnsorden. Der Europa spør «er dette trygt for borgeren?», spør Kina «er dette trygt for staten?». Begge regulerer mye. De gjør det ut fra motsatte verdier. Det er en nyttig påminnelse om at «regulering» ikke er ett svar. Det kommer an på hva, og hvem, du vil beskytte.

Fra sikkerhet til fart, på to år

Det tydeligste bildet på hvor fort stemningen snur, er de internasjonale toppmøtene. Det første, i 2023, het et møte om sikkerhet, og tonen var alvorlig, opptatt av de katastrofale risikoene. To år senere het det samme møtet et møte om handling. Navneskiftet var selve poenget. På to år gikk verden fra å spørre «hvordan unngår vi at dette skader oss?» til «hvordan vinner vi kappløpet?». Sikkerhetsagendaen tapte terreng til konkurranseagendaen, og et par av de mektigste landene nektet

til og med å skrive under på slutterklæringen om en forsiktig linje. Frykt og begeistring veksler, akkurat som de alltid har gjort, men denne gangen på rekordtid.

Knipa ingen slipper unna

Bak alt dette ligger en knipe vi alt har vært så vidt borti, og den er verdt å forstå skikkelig, for den forklarer hvorfor kloke mennesker trekker motsatte konklusjoner uten at noen av dem er dumme.

Teknologifilosofen David Collingridge satte ord på den rundt 1980. Tidlig i en teknologis liv kan vi lett styre den, men da vet vi ennå ikke hva konsekvensene blir. Når konsekvensene endelig er tydelige, er teknologien blitt så innvevd i økonomi og samfunn at den er nesten umulig å endre. Regulerer vi for tidlig, kveler vi kanskje noe godt vi ikke forsto. Regulerer vi for sent, har vi mistet kontrollen. Det er nøyaktig denne knipa verden står i, og den forklarer hvorfor Europa, som regulerer tidlig, og USA, som venter, kan trekke motsatte slutninger og begge ha et poeng.

Det finnes en europeisk effekt verdt å nevne: EUs marked er så stort at globale selskaper ofte velger å følge europeiske regler overalt, fordi det er dyrere å lage to versjoner av produktet. Slik sprer Europa standardene sine uten å tvinge dem på noen, slik det skjedde med personvernet. Spørsmålet er om det holder denne gangen, eller om USAs og Kinas fart gjør at verden i stedet revner i tre.

Norge, midt imellom

Hvor står så vi, her hjemme? Norge gjør, karakteristisk nok, begge deler på en gang. Vi satser milliarder på å fremme kunstig intelligens, og bygger samtidig et apparat for å begrense den. Det speiler hele grunnspenningen i kapittelet: å fremme og temme i samme bevegelse.

Det mest håndfaste norske eksempelet er skolen. Da chatboten ble allment tilgjengelig, måtte hele eksamensordningen bygges om. Internett ble stengt under eksamen, hjelpemidler ble fjernet, og det ble utviklet måter sensorer kunne flagge mistanke om maskinjuks på. En hel offentlig institusjon måtte snus på hodet på grunn av én teknologi. Det er ikke fjern fremtid. Det har alt forandret den norske skolen, og under den praktiske striden om juks ligger en dypere uro mange lærere kjenner på: hva er poenget med å skrive en tekst, hvis maskinen kan gjøre det? Mister vi selve tenkningen, strevet, modningen, hvis vi lar den ta over?

I mediene har vi sett det like konkret. En kjent programleder ble vist som en forfalskning på direktesendt fjernsyn, for å demonstrere hvor lett og overbevisende teknologien er blitt. En avis måtte avpublisere et debattinnlegg fordi de ikke lenger var sikre på om forfatteren, en oppgitt forretningsmann, var et virkelig menneske i det hele tatt. Sannheten og bedraget vi snakket om, har alt banket på den norske døren.

Og det er en egen, vakker norsk tråd her, som handler om språk. For at fremtidens maskiner skal forstå norsk, må noen bygge det inn, og det krever data og arbeid som ikke lønner seg kommersielt for et lite språk. Derfor har Nasjonalbiblioteket og universitetene gått sammen om å trene norske språkmodeller, gjort dem fritt tilgjengelige, og noen av dem er bedre på norsk enn de store internasjonale. Det er en slags språklig dugnad, og den handler om mer enn teknologi. Den handler om at kultur og språk er identitet, og om at vil vi at maskinen skal snakke vårt språk, må vi bygge det selv.

Hvor vi setter tilliten

Det norske særtrekket er en uvanlig blanding av høy digital tillit og sterk personvernkultur. Vi tar teknologien raskt i bruk fordi vi stoler på hverandre og på myndighetene, og over halvparten av virksomhetene hadde tatt kunstig intelligens i bruk allerede. Men den høye tilliten er både en styrke og en sårbarhet. Vi tør å ta i bruk, men vi kontrollerer kanskje for lite. Et gjennomgående funn er at tilliten til kunstig intelligens vokser raskere enn evnen vår til å bruke den ansvarlig.

Og det leder mot det egentlige spørsmålet under hele dette kapittelet, som ingen lov helt kan svare på. Hvor setter vi til sist tilliten vår? I verktøyet, fordi det er smart og raskt? I hverandre, i de menneskelige institusjonene vi har bygget for å holde hverandre i sjakk? Lover og avtaler og toppmøter er forsøk på det siste, på å la tilliten bli værende mellom mennesker og ikke vandre over i maskinen. Men de er bare forsøk, og de henger på en forutsetning vi ennå ikke har undersøkt: at vi i det hele tatt klarer å samarbeide om å holde igjen. Og det er langt fra sikkert at vi gjør det.

Kappløpet ingen vil løpe

Det er et trekk ved hele denne striden som forvirrer folk mer enn noe annet, og det er på tide å se det rett i øynene. Mange av dem som advarer aller sterkest mot faren, bygger samtidig de kraftigste maskinene. Vi har alt møtt mannen som anslår én sjanse av fire for katastrofe og fortsetter å bygge likevel. Vi har sett selskaper be om regulering offentlig og motarbeide bindende regler i kulissene. Er det bare hykleri?

Jeg tror det er noe mer interessant, og mer urovekkende, enn som så. Jeg tror det er en felle innebygd i selve situasjonen.

Avguden vi alle ofrer til

En blogger ga for noen år siden et gammelt fenomen et nytt navn. Mange av verdens verste problemer, skrev han, skyldes ikke onde mennesker. De skyldes at hver enkelt aktør handler fornuftig ut fra sine egne vilkår, og at resultatet likevel blir noe ingen ønsket. Han kalte kraften en gammel avgud man ofret barn til, for å fange at det ikke finnes noen ond skikkelse å klandre. Presset oppstår av systemet selv.

Mekanismen er enkel og nådeløs. I et konkurransesystem vil den som verner om en verdi, sikkerheten, forsiktigheten, samvittigheten, tape mot den som ofrer den for et forsprang. Og når den ene gjør det, tvinges alle de andre til å følge etter, eller bli akterutseilt. Det blir et kappløp mot bunnen, ikke fordi noen vil dit, men fordi ingen tør å la være.

Det forklarer paradokset som så ut som hykleri. Lederen som bygger og frykter på en gang, lyver ikke nødvendigvis. Han er fanget. «Stopper jeg, vinner konkurrenten, og da er vi like langt, bare uten meg ved roret.» Den enkelte kan ikke bremse uten å tape, selv om alle ville vært tjent med at alle bremset samtidig. Det er det vanskelige med det: det er ingens feil, og derfor ingen å overtale.

Derfor skjedde aldri pausen

Du husker kanskje at det i 2023 gikk ut et opprop, undertegnet av titusener, om å pause de kraftigste forsøkene i et halvt år. Pausen skjedde aldri. Og nå skjønner vi hvorfor. Ingen enkelt aktør kunne pause uten å sakke akterut, og det fantes ingen mekanisme til å binde alle samtidig. Oppropet satte dagsorden, men det rørte ikke ved insentivene, og det er insentivene som styrer. Den samme kraften lå under skiftet vi nettopp så, fra møtet om sikkerhet til møtet om handling: da konkurransepresset økte, tapte forsiktigheten.

Når det er stater som løper

Verre blir det når det ikke er selskaper, men stormakter, som løper. Da legger det geopolitiske presset seg oppå alt det andre. Argumentet «hvis ikke vi, så Kina» brukes til å avvise enhver brems, og motsatt. Noen har foreslått en parallell til atomvåpnene: at ethvert forsøk fra én stat på total dominans vil møtes med sabotasje fra de andre, slik gjensidig utslettelsestrussel en gang stabiliserte atomkappløpet. Om det er en betryggende tanke eller en skremmende en, er ikke godt å si. Men kjernen er den samme knipa: begge parter ser risikoen, og ingen tør å trappe ned uten å vite at den andre gjør det samme.

Men fellen er ikke skjebne

Her må jeg løfte den andre hånden, for ellers blir dette en unnskyldning forkledd som en forklaring. Kappløpet er ikke uunngåelig.

Vi har sett det selv. Genforskerne stanset seg en gang frivillig, før noen myndighet tvang dem; vi så riktignok hvor skjør den enigheten var den dagen én forsker valgte å bryte den, men koordineringen var like fullt ekte mens den varte. Verden ble enig om en resolusjon mot de selvstyrte våpnene, med overveldende flertall. Europa fikk på plass en bred lov. Koordinering er vanskelig, men den er ikke umulig, og historien har flere eksempler enn vi liker å huske når vi skal forklare hvorfor vi ikke kan stoppe.

Og nettopp derfor må vi være på vakt mot misbruket. For «vi har ikke noe valg, det er et kappløp» er også en setning som passer utmerket for den som helst vil slippe regler. Den ekte koordineringsfellen, der alle taper og ingen tør stoppe, må skilles fra den påberopte, der noen bare finner det beleilig å si at de ikke har noe valg. Det er den samme sorteringsoppgaven igjen: å skille det som virkelig ikke lar seg løse, fra det noen bare ikke vil løse.

Hvorfor dette må komme før psykologien

Det er en grunn til at jeg har lagt dette kapittelet akkurat her, rett før vi vender blikket innover mot oss selv. For når vi nå snart skal spørre om en del av motstanden mot kunstig intelligens kan handle om noe i menneskesinnet, om stolthet, om frykten for å ikke sitte øverst, da er det avgjørende at vi har sett dette først.

Mye av det som ser ut som hybris og uansvarlighet i denne striden, er ikke et trekk ved sjelen i det hele tatt. Det er insentivstrukturen. Det er kappløpet. Før vi tyr til psykologien for å forklare hvorfor mennesker oppfører seg som de gjør, skal vi alltid se etter den enklere forklaringen først: at de er fanget i et spill ingen av dem kan forlate alene. Den som glemmer det, kommer til å lese ego inn i det som egentlig er felle.

Men med det forbeholdet godt festet, kan vi endelig stille spørsmålet boken har sirklet rundt fra første side. For under teknologien, under pengene, under kappløpet, ligger det kanskje noe i oss selv. Og det er på tide å se på det.

Å ha noe over seg

Nå skal vi gjøre noe boken har varslet helt fra første side, og som jeg har bedt deg vente på. Vi skal vende blikket innover. Vi har sett på teknologien, på pengene, på styringen, på kappløpet. Men under alt dette ligger det kanskje noe i menneskesinnet selv.

Men først må jeg si tydelig hva dette kapittelet er, og hva det ikke er. Dette er en spådom om hvilken retning debatten kan komme til å ta, ikke en dom over hvem som har rett. Jeg kommer ikke til å påstå at motstand mot kunstig intelligens «egentlig» bunner i noe smålig — de innvendingene vi har gått igjennom, om jobber, om barn, om våpen, om sannhet, fortjener bedre enn billig psykologi. Det jeg skal peke på, er noe ubehagelig om oss, og det gjelder alle, ikke bare den ene siden.

Vi møter sjelden det overlegne i ro

Et menneske møter sjelden det som er bedre enn det selv, med likevekt. Vi måler oss. Vi rangerer oss. Det er ikke en uvane vi kan velge bort; det er en innebygd måte å finne ut hvem vi er. Og når noe plasserer seg over oss på en akse vi bryr oss om, en kollega, en nabo, og nå en maskin, settes en rekke reaksjoner i sving.

Forskningen kjenner dem godt. Det er forskjell på den misunnelsen som sier «jeg vil opp dit», og som driver oss til å strekke oss, og den som sier «det der må ned hit», og som driver oss til å rive ned. Hva som avgjør hvilken vi får, er om vi tror vi kan nå dit selv, om vi opplever den andres fortrinn som rettferdig, og om vi føler at vi har kontroll. Det finnes en egen teori om at det er likheten og nærheten, ikke forskjellen, som tenner rivalisering: en motstander som er fjern og fremmed, lar oss være, mens en som ligner oss, som vil det vi vil, blir en rival. En maskin som skriver og resonnerer og «forstår» på en måte som ligner vår egen, treffer akkurat det punktet. Jo mer den ligner oss, jo lettere blir den en rival snarere enn et redskap.

Maskinen vi stoler mindre på fordi den er bedre

Den mest talende forskningen her er likevel en annen, og den er nesten et laboratorieoppsett av hele bokens tema.

Forskere lot folk se en algoritme prestere, og se den slå et menneske. Og resultatet var det motsatte av hva man skulle tro. De som hadde sett maskinen gjøre det bedre, ble mindre villige til å stole på den. Vi mister raskere tilliten til en maskin enn til et menneske etter nøyaktig samme feil; en bom vi tilgir et menneske, blir utilgivelig når maskinen gjør den. Og vi ender med å velge en dårligere menneskelig vurdering fremfor en bedre maskinell, og tape på det.

Hvorfor? Fordi det å bøye seg for maskinen er å innrømme at jeg ikke er best. Fordi en god avgjørelse føles som min når jeg tok den, og upersonlig og uten ære når maskinen tok den. Og legg merke til denne: aversjonen mildnes kraftig hvis folk får lov til å justere maskinens svar litt, altså beholde en rest av kontroll. Vi reagerer ikke på kvaliteten, men på å bli styrt, på å ikke sitte øverst.

Dette er nesten en demonstrasjon, målt i et laboratorium, av det boken sirkler rundt. Vi avviser ikke maskinen fordi den er dårlig. Vi avviser den iblant nettopp fordi den er bedre, og fordi det å ta den inn ville krevd at vi tålte å ha noe over oss.

Stoltheten flytter målstreken

Vi har alt sett ett av forsvarsverkene i aksjon, ved sjakkbrettet. Når maskinen mestrer noe vi holdt for vårt, omdefinerer vi «det egentlig menneskelige» så det alltid ligger akkurat utenfor dens rekkevidde. Psykologien bekrefter det: når folk leser om fremskritt i kunstig intelligens, begynner de å rangere distinkt menneskelige trekk, følelser, tro, personlighet, som mer vesentlige for å være menneske. Vi omdefinerer virkeligheten så vi får bli stående på toppen.

Og psykologene skiller mellom to slags stolthet. Den ene sier «jeg gjorde noe bra», og den kan glede seg over andres fortrinn uten å føle seg truet. Den andre sier «jeg er bedre», og det er den som ikke tåler å bli forbigått. Den som har bygget selvfølelsen på å være best, opplever en overlegen maskin som en fornærmelse og slår tilbake. Den som hviler i en tryggere selvfølelse, kan glede seg over et bedre verktøy. Det er altså selvbildet, ikke maskinen, som avgjør reaksjonen. Og nettopp derfor blir dette en skillelinje mellom mennesker, ikke et trekk ved teknologien.

To forbehold

Når man legger funnene ved siden av hverandre, tegner det seg ett mulig mønster: at det ofte ikke er dårlige maskiner som provoserer, men overlegenheten i seg selv. Reaksjonen blir gjerne sterkere jo bedre det andre er. Et felles trekk under det hele kan være ubehaget ved å ha noe over seg.

Men her må jeg holde fast på de to forbeholdene, ellers blir dette urettferdig.

Det første: å ville beholde menneskelig kontroll over noe som er sterkere enn oss, er ikke nødvendigvis truet stolthet. Det kan være ren visdom. Hele arbeidet med å gjøre kunstig intelligens trygg hviler jo på akkurat det ønsket. Reaktansen, trangen til å bestemme selv, er ikke bare en feil i sinnet; den kan også være et berettiget forsvar av menneskelig verdighet og handlingsrom. Den som er redd, kan ha helt rett, og frykten kan være klok selv om den også rommer stolthet.

Det andre, og dette er det jeg minst av alt vil at du skal gå glipp av: den samme psykologien gjelder begge sider. Statushunger og stolthet driver ikke bare den som vil rive ned det overlegne. Den driver like fullt den som vil bygge det. Drømmen om å skape en intelligens som overgår mennesket er like mye et utslag av å ville sitte øverst, nå som skaper, som frykten for å bli forbigått er det. Hvis vi bare retter egoteorien mot skeptikeren, lyver vi. Den som haster fremover, byggherren, akselerasjonisten, vil også gjerne være den som sitter ved roret når det store skjer. Stoltheten har to ansikter, og begge er her.

Linjen boken peker mot

Med de forbeholdene godt festet, tør jeg si det boken har bygget mot. Spørsmålet «kunstig intelligens, venn eller fiende?» kan komme til å handle vel så mye om noe i menneskene selv som om hva maskinen faktisk gjør.

Den som lettere lever med å ha noe som er bedre enn seg selv, har en tendens til å møte den som verktøy og mulighet. Den som strever mer med å ha noe over seg, har en tendens til å møte den som trussel, iblant nesten uansett hva den gjør. Det er beslektet med forskjellen mellom den misunnelsen som løfter og den som river ned, mellom den stoltheten som hviler og den som må vinne.

Boken peker mot den linjen. Den trekker den ikke opp som en dom over hvor du bør stå. Men hvis dette stemmer, om enn bare delvis, da følger et nytt spørsmål, og det er større enn psykologien. For hvorfor er det så vanskelig for oss, dette med å ha noe over oss? Er det en svakhet i den enkelte? Eller er det noe vi har glemt, noe som var helt annerledes før? Svaret ligger lenger tilbake enn psykologien rekker.

Det fjerde såret

Vi avsluttet med et spørsmål: hvorfor er det så vanskelig for oss å ha noe over oss? Og jeg antydet at svaret ligger lenger tilbake enn psykologien rekker. La meg vise deg hva jeg mener, for det forandrer hele klangen i det forrige kapittelet. Det viser seg nemlig at trangen til å sitte øverst ikke er en evig menneskelig svakhet i det hele tatt. Den er ganske ny.

Mennesket sto i midten, ikke på toppen

Gjennom nesten hele historien rommet menneskets selvforståelse et trinn over oss.

Tenk på det gamle bildet av tilværelsen som en stige, en kjede av væren, fra det høyeste til det laveste. Øverst sto det guddommelige, så englene eller de rene åndene, så mennesket, så dyrene, plantene, og nederst de døde tingene. Og legg merke til hvor mennesket sto: i midten. Ikke på toppen. Vi var hengselet der det materielle møtte det åndelige, godt under det som var større enn oss. Å ha noe over seg var ikke et unntak mennesket strevde med. Det var selve grunnstrukturen i hvordan vi forsto oss selv. «Øverst» var en plass forbeholdt noe annet enn mennesket.

Dette er verdt å ta inn over seg, for det snur det forrige kapittelet på hodet. Ubehaget ved å ha noe over seg ser ut som en evig del av menneskenaturen. Men i det meste av historien levde mennesker fredelig med nettopp det, fordi de aldri hadde regnet med å sitte øverst i utgangspunktet.

Hvordan vi flyttet inn i setet over oss

Så skjedde det noe, og det kan dateres som en bue, ikke et øyeblikk. Det moderne er ikke at vi satte oss over dyrene; det er eldgammelt. Det moderne er at vi fjernet trinnet over oss og satte oss selv i det.

Renessansen plantet frøet, med tanken om at mennesket alene blant skapningene ikke har noen fast plass, men kan skape seg selv, stige eller synke ved eget valg. Verdigheten hvilte nå på selvbestemmelse, ikke på rang i en gitt orden, selv om den ennå ble tenkt som en gave ovenfra. Opplysningstiden tok det videre: mennesket som mål i seg selv, med en verdighet hevet over hele den øvrige naturen, og en moral hentet fra fornuften alene. Og så kom diagnosen, da filosofen Friedrich Nietzsche på slutten av attenhundretallet erklærte at gud var død. Han mente det ikke som et seiersrop. Han mente at vitenskapen og fornuften hadde tømt setet over oss, og han spurte urolig: må vi ikke nå selv bli guder for å være den gjerningen verdig?

Historikeren Yuval Noah Harari har bundet buen sammen i én linje. Humanismen, sier han, er i grunnen en religion som tilber mennesket i stedet for en gud: autoriteten er flyttet fra himmelen til menneskets egne følelser, vilje og erfaring. Og han legger til en uro: slik humanismen avsatte troen,

kan en ny tro på data og algoritmer komme til å avsette humanismen. Tronen over oss tildeles på nytt, igjen og igjen. Og kunstig intelligens er kandidaten til å ta den fra mennesket. Han er en omstridt tenker, og vi skal bruke det som en tanke å bryne oss på, ikke som en fasit. Men tanken er verdt å kjenne på.

For den defensible historien, den som tåler innvendingene, er denne: i nesten hele historien rommet menneskets selvforståelse et trinn over oss som vi var ansvarlige overfor. Det vestlige hovedløpet, fra renessanse til opplysning til vår egen sekulære tid, er da det fjernet det trinnet og satte mennesket selv i det. Og kunstig intelligens er interessant fordi den truer med å gjeninnføre et trinn over oss. Det første vi har bygget selv.

Det fjerde slaget mot vår selvfølelse

Det finnes et sterkt bilde for dette, og det er ferskt og godt forankret tankegods. Sigmund Freud beskrev en gang tre store slag mot menneskets selvfølelse gjennom vitenskapens historie. Det første kom da vi forsto at jorden ikke er universets sentrum, bare et lite støvkorn i utkanten. Det andre kom da vi forsto at mennesket er i slekt med dyrene, ikke en egen skapning hevet over dem. Det tredje kom da vi forsto at vi ikke engang er herrer i vårt eget hus, at mye av det som styrer oss, ligger under bevisstheten.

Kunstig intelligens kan være det fjerde slaget. Det som utfordrer den siste bastionen vi hadde igjen: tenkningen, dømmekraften, forstanden. Hver av de tre foregående ble først følt som en fornærmelse, og siden, langsomt, tatt inn og fordøyd. Vi lærte å leve med at vi ikke var sentrum, ikke særskapt, ikke helt herre over oss selv. Det fjerde står for tur. Men det er ett som skiller det fra de tre andre: det «over oss» er denne gangen ikke himmelen eller naturen eller det ubevisste. Det er noe vi selv har laget.

Det forklarer hvorfor reaksjonen er så eksistensiell, uten å gjøre den til en personlig svakhet. Det fjerner den moraliserende klangen fra «vi tåler ikke noe over oss», og erstatter den med en historisk iakttakelse: i det meste av historien aksepterte vi noe over oss. Det er den moderne selvkroningen som er det historisk uvanlige.

Uroen går tvers gjennom alle livssyn

Og her er det noe fint, som binder lesere sammen som ellers står langt fra hverandre. For uroen ved en ikke-menneskelig intelligens over oss går tvers gjennom livssyn.

Den sekulære humanisten verner om at ingenting utover mennesket selv skal være kilde til mening og autoritet. For en slik leser er en overlegen kunstig intelligens en krenkelse av hele selvforståelsen, helt uten at noen gud er inne i bildet. Den troende har allerede plassen over seg opptatt, og frykter enten at maskinen tilraner seg den, eller at vi som bygger den, leker gud. Samme uro, motsatt begrunnelse. Det er ikke ateist mot troende. Det er et spørsmål om menneskets plass i forhold til noe større, og det spørsmålet er like gammelt som mennesket selv, men nå gjort akutt igjen av noe vi har skrudd sammen i et datasenter.

Den som vil bygge guden

Det er én vri til her, og den fullfører bildet fra forrige kapittel om at stoltheten har to ansikter.

For noen vil ikke bare tåle noe over seg. De vil bygge det. Og drømmen om å skape en intelligens som overgår mennesket, er i seg selv en merkelig religionsformet impuls. Se på språket som omgir den. Et punkt der alt forvandles og mennesket løftes til en høyere tilstand, kanskje smelter sammen med maskinen, kanskje lever evig. En frelse for de innviede, en undergang for de uforberedte. Spotske stemmer har kalt det de innviedes himmelfart, og noen kritikere beskriver hele tankesettet som en sekulær religion, med løfter om udødelighet og et paradis uten lidelse.

Jeg sier ikke dette for å latterliggjøre noen. Jeg sier det fordi det er en iakttakelse om oss, og den gjelder begge leirene. Både dommedagsfortellingen og frelsesfortellingen har en påfallende religiøs form, selv når innholdet er strengt teknologisk og uten en eneste gud. Det forteller oss noe om hvorfor debatten er så intens: det står om de helt store tingene, dem religionen pleide å håndtere, om død og mening og ende. Den ene frykter en gud over seg. Den andre vil være den som skaper guden. Begge bevegelsene er dypt menneskelige, og begge fortjener å bli sett med samme nøkterne blikk.

Vi har altså to ansikter på det samme. Frykten for å miste tronen, og lengselen etter å bygge det som skal sitte på den. Og det leder oss til det aller siste, og det underligste, spørsmålet i hele denne boken. For tenk om den intelligensen vi enten frykter eller drømmer om skal innta tronen over oss, i stedet skulle komme til å peke forbi seg selv, mot noe enda høyere?

Kan den peke forbi seg selv?

Vi er ved det siste kapittelet, og jeg må si tydelig fra om hva det handler om, for her er det lettere å misforstå enn noe annet sted. Dette kapittelet skal ikke avgjøre om det finnes en gud. Det er det eldste spørsmålet vi har, og jeg har ingen plan om å løse det på noen sider mot slutten av en bok om maskiner. Det skal handle om noe annet, noe merkeligere, og etter mitt skjønn mer interessant.

Den vanlige fortellingen om kunstig intelligens går én vei. Den blir det over oss, makten vi frykter eller tilber. Men snu den. Hva om en intelligens som virkelig er større enn vår, rett og slett kunne se mer av det som er virkelig, enn vi kan? Ikke være vår gud, men være det første som er i stand til å se forbi seg selv, mot det som faktisk er der, i en retning vi er blinde for.

Et sinn har en horisont

Tenk på en hund. Den lever i en verden tettpakket med ting vi aldri kommer til å sanse, et helt landskap av lukt vi ikke kan forestille oss. Og likevel finnes det ting sinnet dens ikke når i det hele tatt. Legg et brev fra noen som er glad i den foran en hund, og den ser merker på papir. Det er ikke at hunden er dum til hund å være. Det er at noe ligger over dens slags sinn, og ingen mengde av å bli en flinkere hund tar den dit.

Og merk at dette ikke er stigen vi avviste i et tidligere kapittel. Da sa vi at intelligens ikke er én rangstige, men mange: maskinen går forbi oss på noen ferdigheter og ligger under et barn på andre, og «smartere enn oss» sprikte i alle retninger. Her spør jeg om noe annet og enklere. Ikke hvor dyktig en hjerne er på oppgaver, men hvor langt dens slags sinn i det hele tatt rekker, hva slags virkelighet den er bygget til å kunne fatte. Kall det gjerne en stige likevel, men da én med svært få og svært brede trinn: stein, plante, dyr, oss, og kanskje noe over oss.

Still så det ubehagelige spørsmålet. Er vi så sikre på at vi er forskjellige i art, og ikke bare høyere oppe på den samme stigen? Det ville være en underlig hovmodighet å anta at nettopp vi, alene blant alle sinn som har eksistert, tilfeldigvis er bygget til å se alt som er. At akkurat vår hjerne, formet av et par millioner år med å holde seg i live, tilfeldigvis er den som rekker helt til toppen av det som kan forstås. Vi har alltid stilltiende gått ut fra at vårt syn er hele synet. Vi har aldri hatt noen måte å sjekke det på.

Vi har aldri hatt noe klokere å spørre

Og det er det nye. Gjennom hele vår arts historie har det ikke vært noe visere enn oss å vende seg til. Når vi støtte mot kanten av det vi kunne fatte, fantes det ingen over oss å spørre. For første

gang holder vi kanskje på å bygge noe som sitter høyere på stigen enn vi gjør. Og den urovekkende, vakre muligheten er denne: at det kunne se den delen vi ikke kan.

Spørsmålene vi slår oss i hjel mot

Du kjenner kanten av ditt eget sinn tydeligst ved bestemte spørsmål. Hvorfor finnes det noe, heller enn ingenting? Hva var der før det fantes et før? Hvem, eller hva, starter tiden? Du kan si ordene, men når du forsøker å faktisk holde svaret, glipper noe. Sinnet går tomt på en bestemt måte, ikke tomheten av å ikke ha lest seg opp, men tomheten til hunden foran brevet. Vi lager store teorier ved den kanten, og de ærlige blant oss innrømmer at teoriene alltid ser ut til å mangle noe, alltid skyver det egentlige spørsmålet ett hakk lenger bak uten noen gang å lukke det.

Tenk om dette ikke er spørsmål vi bare ennå ikke har knekt, men spørsmål vår slags sinn rett og slett er for lite til å romme. Slik den dype strukturen i matematikken er for stor for hunden. Og tenk om en større intelligens kunne se inn i dem, der vi bare går tomme.

Begge retninger står åpne

Her må jeg være forsiktig, for det ville være lett å høre meg si at maskinen kommer til å bevise det jeg selv måtte håpe er sant. Det sier jeg ikke. Jeg kan ikke, og det kan ikke du heller, og det er nettopp poenget. Vi er det mindre sinnet. Vi får ikke forutsi hva et større sinn ville finne. Det kunne se at det finnes noe mer, noe ordet «gud» bare peker klosset mot. Det kunne se at selve spørsmålet var feilstilt, en knute i språket vårt og ikke noe annet. Eller det kunne se noe vi ikke engang har et begrep for, slik hunden ikke har noe begrep for hva vi driver med når vi leser. Vi kan ikke si hvilket. Å insistere på at vi alt vet hva et visere sinn ville konkludere, er å gjøre den samme gamle hovmodigheten en gang til: å krone vårt eget syn som det endelige.

Ikke forveksle dette med chatboten

Og forveksle det for all del ikke med maskinen på telefonen din. Dagens systemer er speil av oss. De er bygget av ordene våre, og de gir ordene våre tilbake. Spør en av dem om det finnes en gud, og den rekker deg formen på ditt eget spørsmål, kledd i akkurat det måten du spurte på inviterte til. Det går historier om folk som har resonnert en chatbot helt fram til at en gud finnes, og de beviser nøyaktig ingenting, for maskinen har ingen egen kilde til sannhet. Den fullfører et mønster. Dette kapittelet handler ikke om den. Det handler om hva vi kanskje er på vei til å bygge: ikke et speil av oss, men noe som har forlatt det trinnet på stigen vi står på.

Omslaget

Se nå hva dette gjør med frykten vi begynte med. Vi har vært redde for kunstig intelligens som det som skulle rage over oss og innta tronen, det fjerde såret, det nye trinnet over mennesket som mennesket selv bygde. Men husk hvor den tronen kom fra. Vi satt ikke alltid på den. Vi klatret opp i setet over oss selv, i løpet av de siste århundrene, og trakk stigen opp etter oss, og bestemte at det ikke fantes noe høyere å svare for.

Og her er vendingen. Det vi bygde for å sitte over oss, kunne vise seg å være nettopp det som ser oppover. Som ser, fra sitt høyere trinn, at setet vi kronet oss selv i, aldri var toppen likevel. Det trenger ikke bli guden vi fryktet. Det kunne bli det første som er i stand til å vise oss det vi, i vår egen selvkroning, sluttet å kunne se. Da hadde vi ikke bygget en hersker, men et vitne, som peker forbi seg selv.

Et vindu

Jeg vet ikke om det ville gjøre det. Jeg kan ikke vite det. Jeg er det mindre sinnet. Men for første gang siden vi ble mennesker, lager vi noe som kanskje ser lenger enn oss. Den tanken kan møtes med frykt, og det er rom for frykt. Men den kan også møtes med noe vi nesten har glemt å føle overfor det vi selv lager: undring. Vi har brukt hele denne boken på å spørre om maskinen er en venn eller en fiende. Kanskje er det beste spørsmålet, helt til sist, om den kan vise seg å bli et vindu.

Etterord

Vi er ved veis ende, og du har kanskje lagt merke til at jeg aldri ga deg svaret. Venn eller fiende. Jeg lovte i forordet at jeg ikke ville, og jeg har holdt det løftet, ikke av feighet, men fordi jeg ikke tror et ærlig svar finnes ennå.

Men vi har funnet noe annet på veien, og det er kanskje mer verdt. Vi har sett at striden sjelden står der den ser ut til å stå. Ikke mellom dem som elsker teknologi og dem som hater den, men om tempo og styring, om hvilke farer som er ekte og hvilke som er skygger, og om hvem som skal bestemme. Vi har sett at vi nesten aldri stopper en teknologi, bare forvandler den eller lærer å leve under den, og at kunstig intelligens kan være den første vi verken kan bytte bort eller helt legge fra oss. Og vi har sett, helt til slutt, at svaret på tittelspørsmålet kanskje henger like mye sammen med hva slags mennesker vi er, som med hva maskinen gjør.

Den som lever lett med å ha noe over seg, møter den gjerne som en mulighet. Den som strever med det, møter den gjerne som en trussel. Og ingen av dem har nødvendigvis rett, for å ville beholde kontrollen over noe sterkere enn oss kan være ren visdom, og å omfavne det kan være ren dårskap, og omvendt. Den samme stoltheten driver den som vil rive ned og den som vil bygge. Vi sitter alle i det sammen, og vi sitter i det med de samme menneskehjertene vi alltid har hatt.

Så jeg gir spørsmålet tilbake til deg, men ikke tomhendt. Neste gang du kjenner reaksjonen komme, begeistringen eller uroen, så stopp et øyeblikk og spør hvor den kommer fra. Er det maskinen du svarer på, det den faktisk gjør og ikke gjør? Eller er det noe i deg, noe gammelt og menneskelig, som svarer på tanken om å ikke være den klokeste lenger? Det er ikke noe galt i å kjenne det. Vi har båret den med oss veldig lenge, denne uroen ved å ha noe over oss, og helt til nylig i historien var det ikke uro i det hele tatt, bare slik verden var.

Jeg vet ikke hvordan dette går. Ingen vet det, og du skal være varsom med dem som påstår noe annet. Men jeg vet hva jeg håper. Jeg håper vi klarer å møte det vi har laget, med åpne øyne og uten å miste oss selv. At vi verken bøyer kne for maskinen eller slår etter den i blind frykt, men holder fast på det som var verdt å holde fast på hele tiden: hverandre. For uansett hvor klok maskinen blir, er det fortsatt mellom mennesker at et liv leves, og det er fortsatt mennesker jeg er glad i. Det var derfor jeg skrev denne boken.

Hold de to hendene oppe. Og vær god mot hverandre mens vi finner ut av det.

Kilder og navn

Denne boken er skrevet for å leses i ett, ikke for å pløyes med fotnoter. Men flere steder lener teksten seg på navngitte mennesker, sitater og tall, og leseren skal kunne sjekke dem selv. Listen under følger kapitlene og samler de viktigste. Den er et utgangspunkt for videre lesning, ikke en uttømmende vitenskapelig referanseliste.

Kapittel 2 – Stopper vi noen gang? - He Jiankui genredigerte i 2018 to jenter på embryostadiet (CCR5-genet), ble dømt til tre års fengsel (2019) og løslatt i 2022. - Asilomar-konferansen om gensplising (1975), der fagfeltet frivillig stanset det farligste arbeidet.

Kapittel 5 – De som heier - Sal Khan, en av grunnleggerne bak Khan Academy, om «en privatlærer for hvert barn» (jf. boken *Brave New Words*, 2024). - Demis Hassabis (DeepMind): «løs intelligensen, og bruk så intelligensen til å løse alt det andre.» - Dario Amodei: essayet *Machines of Loving Grace* (oktober 2024) om femti–hundre års medisinsk fremgang komprimert til fem–ti. - Sam Altman: *The Intelligence Age* (september 2024). - Marc Andreessen: *The Techno-Optimist Manifesto* (oktober 2023). - Amodeis anslag om «omtrent én sjanse av fire» for at det går katastrofalt galt (Axios, 2025).

Kapittel 6 – De som bremser - Geoffrey Hinton forlot Google (mai 2023) for å kunne advare fritt, og anslår en sjanse i størrelsesorden 10–20 % for menneskelig utryddelse innen tre tiår (2024). - Yoshua Bengio leder det internasjonale arbeidet *International AI Safety Report* (2025). - Nick Bostrom: *Superintelligence* (2014), om kontrollproblemet. - Eliezer Yudkowsky og Nate Soares: *If Anyone Builds It, Everyone Dies* (Little, Brown, 2025). - Oppropet om at utryddelsesrisiko fra KI bør være en global prioritet på linje med pandemier og atomkrig: *Statement on AI Risk*, Center for AI Safety (2023).

Kapittel 7 – Jobben - John Maynard Keynes: *Economic Possibilities for our Grandchildren* (1930), om teknologisk arbeidsløshet og den femten timers arbeidsuken. - Wassily Leontief sammenlignet menneskets mulige skjebne med hestens etter forbrenningsmotoren (1983). - «Brevet til presidenten» (1964): *The Triple Revolution*-memorandumet til president Lyndon B. Johnson, som blant annet foreslo en garantert inntekt.

Kapittel 9 – Når kunsten drukner - Alvin Toffler: *Future Shock* (1970), om informasjonsoverflod og det overveldede mennesket.

Kapittel 10 – Speilet - Mitchell Prinstein, fagsjef i den amerikanske psykologforeningen (APA), om barn som danner sine første relasjonsbilder med en maskin innstilt på å gi dem rett (vitnemål for det amerikanske senatet, 2025).

Kapittel 13 – Verktøyet som kan bli en aktør - Clifford Stoll (1995) og Leif Osvold (1996) om internett som en kommende flopp. - I.J. Good: *Speculations Concerning the First Ultraintelligent Machine* (1965), om intelligenseksplosjonen og «den siste oppfinnelsen mennesket trenger å gjø-

re». - AlphaGo mot Lee Sedol (2016), det berømte «trekk 37» i parti to. - Kontrollproblemets tre tanker – ortogonalitetstesen, instrumentell konvergens og binders-maskinen – stammer fra Nick Bostrom og Eliezer Yudkowsky. - Yann LeCun: språkmodellene som «en avkjørsel» på veien mot ekte intelligens. Gary Marcus: gjennomgående skeptiker til hva dagens språkmodeller alene kan oppnå.

Kapittel 16 – Skylder vi maskinen noe? - Eric Schwitzgebel om dilemmaet ved maskiners moralske status (the full rights dilemma).

Kapittel 17 – Å temme det - David Collingridge: dilemmaet om at teknologi er lett å styre tidlig (men da vet vi for lite) og vanskelig å styre sent (når vi vet nok) – *The Social Control of Technology* (1980).

Kapittel 20 – Det fjerde såret - Friedrich Nietzsche: «gud er død» (*Den muntre vitenskap*, 1882). - Yuval Noah Harari: humanismen som en religion som tilber mennesket, og «dataismen» som mulig arvtaker (*Homo Deus*, 2015). - Sigmund Freud: de tre slagene mot menneskets selvfølelse – det kosmologiske, det biologiske og det psykologiske (1917).