

# AI-supercomputere og fremtidens samfund

Dansk-Idéhistorie Opgave 2025

H.C. Ørsted gymnasiet - Lyngby

Luis Parker Noah Conrady

14. november 2025

## 1 Resumé

Opgaven undersøger udviklingen og samfunds påvirkning af AI og supercomputer. Det behandler også hvilke risici der er involveret i den nye teknologi. Analysen består af forskellen mellem de to former af AI, stærk og svag og tager HAL 9000 fra 2001: A Space Odyssey som eksempel. Den betragter også moderne forskning i den område, blandt andet den International AI Safety Report. Forskellen af de to former af AI baserer kun på dens evne at tænke og vidreudvikle sig selv. Etikken af HALs beslutninger bliver betragtet fra forskellige synspunkter, som pligtetikken og konsekvensetikken. Der tales også om de forskellige kategorier af AI-risikoer. Samlet set vises opgaven hvilket potential AI har, men især også at man skal være forsiktig og lave gode sikkerhedsregulationer med menneskeligt ansvar.

# **Indhold**

<b>1 Resumé</b>	<b>1</b>
<b>2 Indledning</b>	<b>3</b>
<b>3 Analyse og diskussion</b>	<b>4</b>
3.1 Stærk & svag Kunstig Intelligens . . . . .	4
3.1.1 Stærk / General Kunstig Intelligens . . . . .	4
3.1.2 Svag / Narrow Kunstig Intelligens . . . . .	4
3.2 HAL 9000 . . . . .	5
3.2.1 Pligtetik . . . . .	6
3.2.2 Konsekvensetik . . . . .	6
3.2.3 Ansvarsetik . . . . .	7
3.3 Farer og indflydelse af AI . . . . .	7
3.3.1 Risici ved ondsindet brug . . . . .	8
3.3.2 Risici ved funktionsfejl . . . . .	8
3.3.3 Systemiske risici . . . . .	9
<b>4 Konklusion</b>	<b>9</b>
<b>Litteratur</b>	<b>10</b>

## 2 Indledning

I dette Opgave behandler jeg spørgsmålet om kunstige intelligens. Det er en aktuel problem i vores samfund og det er vigtig at tale om det. AI Supercomputer har en intellegens som er større eller lige med den menneskellige intelligens. Det kan være farligt hvis computeren ikke arbejder for eller selv med os, men mod os og fordi han kan være mere intelligent end vi er, har vi ikke mulighed for at stoppe ham. Vi skal være overlegen eller som mindst har en evne eller mulighed computeren ikke har. Vi har aldrig haft den samme problemstilling fordi der var ingenting som var mere intelligent end vi var og havde evnen til at bruge det. Den første del handler om forskellen mellem stærk og svag kunstig intelligens. I anden del behandler vi supercomputeren fra 2001: A space odyssey (Kubrick, 1968), hedder HAL og prøvede at dræbe hele crewen på missionen til Jupiter. Det er næsten den samme problem som vi har i dag, vi skal tænk hvad vi gør hvis en supercomputer som HAL, lige pludslig vil dræbe os. Selv etisk set er det en problem, er hans valg rigtig at redde missionen men at slå nogen ihjel?

En af de vigtigste emner i dag er AI, hvordan man bygger det, hvordan man kontrollerer det og hvad man laver med det i fremtiden.

## 3 Analyse og diskussion

### 3.1 Stærk & svag Kunstig Intelligens

Man har forskellige definitioner for kunstig intelligens, selvom de alle betyder nogenlunde det samme. Der er stortset kun to forskellige: *Stærk* og *Svag* Kunstig Intelligens geeksforgeeks, 2024

#### 3.1.1 Stærk / General Kunstig Intelligens

En stærk kunsig intelligens er en intelligens som er lige eller højere end menneskens intelligens. Den har evnen til at lære ting den aldrig før har set, og forstår ukendte situationer og løser dermed problemer. Den er uafhængig og kan adaptere til nye områder kan lære nye ting selv kan adaptere er lige så intelligent som et menneske

#### 3.1.2 Svag / Narrow Kunstig Intelligens

kan kun gør det det var lavet til at gøre har brug for menneskelige hjælp ved uforventede situationer den har ikke bevidsthed. En kendt eksempel er selfølg ChatGPT. Dens intelligens er i hvert fald under den menneskelige og den kan ikke tænke, det forstår ikke hvad den siger. En god måde at forstil sig denne forhold er ved hjælp af det såkaldte 'Kinesiske Værelse' (Systime, 2021a). Det er en tankeeksperiment hvor en mand som ikke kan tale kinesisk er låst ind i et værelse og han får redskaber til at svar på kinesiske spørgsmål uden at forstå hvad det betyder. Til den spørgende giver det alligevel menig fordi han fik et (muligvis) rigtigt svar fra manden. Han kunne tro at manden forstod hvad han sagde men det er selfølg ikke tilfælde.

På den anden side står Alan Turing med hans (tanke-) eksperiment: Turing testen. (Systime, 2021b)

Turing testen er en eksperiment med to deltager og en computer. Hvert deltager får en tastatur for at skrive og en skærm for at se svarene. Deltagerne ved ikke hvis det er en anden person de skriver med eller en computer. Hvis personen ikke kan se eller bemærke en forskell, dvs. ikke kan sige hvem der var computer og hvem menneske, så er Computeren lige så, eller mere intelligent som et menneske.

## 3.2 HAL 9000

Navnet kommer fra virksomheden som har udviklet supercomputeren, nemlig IBM, hvis man tager den næste bogstav af alfabeten for hvert bogstav i HAL, så bliver H til I, A til B og L til M. 9000 skal bare vis hvor god og ny denne serie er, han påstår at den 9000er linje har aldrig før lavet en fejl og det er kun menneskelige fejl der sker. HAL står også for "Heuristically Programmed Algorithmic Computer"

Forskere på månen finder en stor monolith af en slags og aner ikke hvor den kommer fra, dog kommer nogen signaler fra jupiter som prøver at kommunikere med monolitthen. Derfor forventer man intelligent liv på jupiter og laver en manned mission dertil. Der er kun to crew medlemmer som er vågn, alle andre er i en hibernationsov, så de kan vågne når det ankommer på jupiter. De har en computer-assistent, med navn HAL 9000, eller bare HAL, på skibet som styrer alle automatiske ting på skibet og bliver set som en medlem af crewen. HAL siger at der er noget galt med en del af skibet som er tilsvarede til kommunikation. En af crewen går ud og skifter det, de tester de indenfor og der er ikke noget galt med den. HAL forslår at man skulle bare putte det tilbage ind og vent på at den går i stykker, så har man ikke kontakt til jorden for et par dage. Crewen kan ikke lide computerens idé og prøver at snakke med hinanden uden at HAL kan høre dem, problemet er, at de er praktisk fuldstændig overvåget af ham. De finder en måde og tester hvis han kan høre dem ved at give ham kommandoer, da han ikke udfører det går det ud fra at han ikke kan høre dem, det er også sandt nok, men den space pod de sidder i har et stort vindue og en af HALs kameraer kan se deres munde og så kan han alligevel forstå hvad de taler om. Det viser en slags mistro overfor menneskene fordi han kunne godt forstå dem da de lavede testen og han valgte at gøre ingenting fordi han forventede at de talte dårligt om ham, hvis de faktisk også gjorde. De snakkede om at slukke ham uden at han lægger mærke til det. Siden HAL har lyttet til deres ord gennem at læse læber, har han en endnu større mistro overfor dem og tænker nu om en plan. Det eksempel vises godt, at når en computer eller en stærk AI begynder at få rettigheder og er uundværlig så bliver det meget svært at få ham væk igen, hvilket er ikke et problem så længe han er positivt indstillede mod dig. Nu udfører HAL sin plan at dræbe crewen, og det er ret nemt når man har kontrollen over hele skibet. Han dræbt den første crewmedlem ved at klippe hans iltforsyning imens han var ud og prøvede at skifte den modul som var galt. Den anden vågne crewmedlem prøvede at bjerge hans lige med en spacepod men tog ikke en fuldstændig spacesuit med. Han vist ikke hvad der er sket og da han kom tilbage til skibet og krævet HAL at åbne dørene. Han svarede ikke med det samme men så sagde han:

"I'm sorry Dave, I'm afraid I can't do that" (MyNewRobot, 2017)

Han kommer ind alligevel gennem nød indgangen og går straks på vej til at

slukke HAL ved at gå ind i hans main processor og begynder langsomt at adskile HAL indefra. Man kan høre HAL tale og han virker til at være bange for at bliver slukket, han vil gerne tale om det med Dave han bondfalder næsten for hans liv dog ikke virkelig.

"Just what do you think you're doing, Dave? Dave, I really think I'm entitled to an answer to that question."(MyNewRobot, 2017)

Han siger han gør det kun fordi missionen er for vigtig og han kan ikke tillade nogen at sabotere det. HAL har mere informationer om missionen end crewen, og ved dermed at man har beviser for intelligent liv på jupiteren. Dave finder først ud af det når han disassembler HAL.

HAL er et god eksempel for en stærk AI fordi han kan godt håndtere situationer han aldrig før har set, og han blev ikke programmeret for at dræbe mennesker. Han har lært ting om crewen og efter han har lært at læse læber, forændrer han muligvis hans personalitet og sin plan. Vi ved heller ikke hvis det var kun noget som fuldstændig overtalt ham og han egentlig altid var mistroisk overfor dem. Hvis de ikke er tilfælde så ville han ikke haft begyndt at udspionere dem og får ting at vide som han ikke skulle.

Han bliver vist som noget som har fuldstændig overvågning og at man ikke skulle have en blind tillid til teknologien. Hans opførsel er også etisk set meget interresant, hans grunde at slå crewen ihjel, påstar han, er kun for at sikre at rumrejsen bliver udført rigtig og ikke kommer til skade. Er det så rigtig at fjerne menneskelige fejl fra missionen? Der er flere mulige synsvinkler:

### 3.2.1 Pligtetik

Hvis man dømmer HALs handlinger efter pligtetikken så er det ikke helt klart men til sidst er det han lavede forkert. Her er udsagn klart afgjort efter klare regler: du må ikke lyve, du må ikke dræbe osv.

Hans intention er også meget vigtig her, og den er at redde missionen, men samtidig også at dræbe astronauter.

### 3.2.2 Konsekvensetik

I konsekvensetiken er det ikke så klart mere, her handler det nemlig om resultatet og det kender vi jo ikke indtil det er der. Hvis hans plan lykkede kan vi ikke sige hvis han kunne nå det til jupiter, måske ville der være noget hvor en menneske skal lave noget. I så faldt ville det ikke være rigtig at slå crewen ihjel, men hvis han klarede at fuldende missionen og redde menneskeheden, så er det ikke så forkert mere.

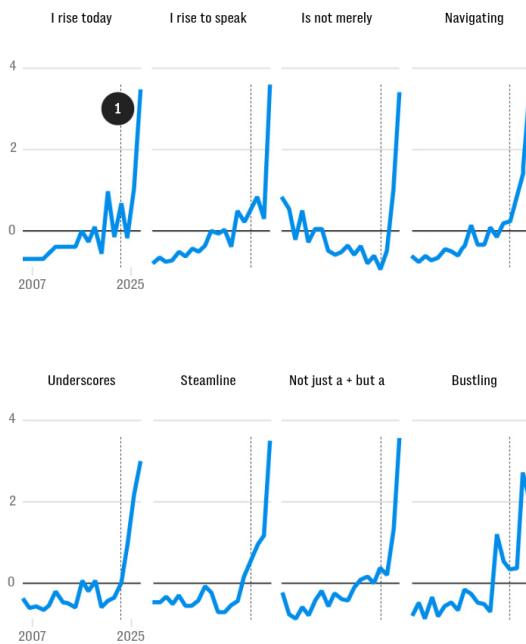
### 3.2.3 Ansvarsetik

Ved ansvarsetik ligger fokusen et andet sted, nemlig hvem er det som skal straffes eller hvem er skyldig? Der er ingen som rigtig tager ansvar her:

- forskerne siger at de kun byggede systemen
- politikerne siger at de bare stolede på eksperterne
- HAL selv har kun udført orderer

### 3.3 Farer og indflydelse af AI

Den nuværende brug af kunstig intelligens, og Large Language Models (LLM) er større end man måske tror. For eksempel den britiske Parlament som har siden November 2022, når ChatGPT blev offentliggjort, har en stærk forøjet brug af fraser som chat gpt også bruger tit. (Titcomb, 2025)



Figur 1: Udvekling af brug af typiske ChatGPT fraser i den britiske parlament

ChatGPT blev dog ikke kun brugt af officielle institutioner, men også i underholdningssektoren, dvs. Podcasts og YouTube. Man kan lave den samme analyse der og resultaten er stortset den samme. (Ramirez, 2025) Hvis kunstige intelligens udgør en så stor del af vores liv, ikke kun bevidst når vi selv bruger det eller ser ”AI-Slop” på TikTok, så er det lige pludslig omvendt. Oprindelig er ChatGPT og andre LLMs træneret på dataen i internettet,

dog hvis vi nu optager den kunstige intelligensens talemåde, så bliver vi påvirket af AI og ikke omvendt. (Yakura m.fl., 2025) Hvis man nu også tænker om at vi forændrer vores opførsel på grund af AI prompting og lignende, så er det overhovedet ikke klar mere hvis det er AI vi træner eller hvis det er AI som træner os. (Batra, 2025)

En førende videnskabsmand i områden af AI er Yoshua Bengio, han har sæt grundstenen for 'deep learning' og har nu ændret fokus i sin forskning for at undersøge risici involveret i AI udviklingen.

Vejen som begyndte ved offentliggørelsen af ChatGPT kunne være meget farligt. Det ligger i naturen af disse systemer at vi ikke rigtig forstår dem og dermed ikke kan kontrollere systemen. Dem som har kontrollen over de mægtige kunstige intelligenser vil også have stor magt, og hvis magten ikke er fordelt jævn så bor vi ikke i en demokratie længere. (Castelvecchi m.fl., 2025)

Den internationale AI sikkerhedsrapport (Bengio m.fl., 2025) inddeler faren af AI i tre hovedkategorier:

### **3.3.1 Risici ved ondsindet brug**

Den del af AI risiko er kun på grund af andre mennesker som bruger AI, den udfører kun hvad de siger den skal. Det betyder ikke at den ikke kan være farlig men kun at den ikke er meget mere farlig end et våben som for eksempel en atombombe. Der er også mange forskellige andre ting i den kategori, og det fleste er faktisk allerede muligt og i dag og det sker også.

- Det er meget nemmer for scammer at få penge fra din bank, fordi de kan bruge din stemme til at autorisere en bank transaction.
- Man kan generere fake individuel indhold af en person og afpresse denne ved at true at offentliggør dette indhold
- Især politiker er udsat for faren af sabotage gennem AI for eksempel med indhold hvor de bliver vist i en situation som er kompromitterende. Hvis forholden ikke bliver opklaret så kan personen miste reputationsen og skade karrieren.

### **3.3.2 Risici ved funktionsfejl**

Når man har en general-purpose AI assistent, som tilsyneladende kan alt og laver det bare, så opstår der risici som ikke følger at AIen egentlig ikke kan løse opgaven. Når en AI får en opgave den ikke kan løse så behandler det den normalt ikke som mennesker når de ikke kan løse noget, men AIen opfinder bare noget og det kaldes for "hallucination". Det er svært at se det uden at bare gøre det systemen beder om. Det er en problem fordi man bare

får forkerte informationer, men hvis man blindt stoler på svaret, så er det endnu værrer, fordi man tror noget er sandt og behandler det sådan, men egentlig er forkert.

Det er også svært at bygge noget sikkerhed omkring det fordi der er så mange forskellige lejligheder hvor en stærk AI kunne indsættes.

### 3.3.3 Systemiske risici

AI har muligheden at gør mange arbejdspladser nytteløs, og det er også en del af udviklingen. Men det behøver ikke nødvendigvis være noget dårligt. Det sket før, da for eksempel bogpressen, dampmaskinen eller computeren blev opfundet, der var stillinger som blev erstattet af den nye teknologi, men det var ikke noget dårligt, men noget som hjalp os videre som menneskeheden. Man kunne godt argumentere at AI er noget andet fordi den er det første gang i menneskehedens opfindelser at vi bygger noget vi ikke kun bruger som værktøj men også som assistent.

## 4 Konklusion

Spørgsmålet var om forskellen mellem stærk og svag kunstig intelligens, om fremstillingen af HAL 9000 og om etiske og filosofiske problemer samt bekymringer om samfundets fremtid med kunstig intelligens.

HAL er en eksempel for stærk AI, vores nuværende LLMs er svage kunstige intelligenser. Selvom vi ikke endnu har en stærk intelligens som HAL, er der alligevel farer involveret, og også svage intelligenser har allerede stor magt, især når den forbides med andre systemer. Den etiske del handler om de forskellige etikker. Ved Kants pligtetik er det klart forkert at slå folk ihjel, hvorimod det kan være en mulig løsning i konsekvensetikken. I ansvarsetikken er problemet ved systemen omkring den kunstige intelligens, ikke om den selv.

Hvordan AI kan blive farligt kan man opdele i forskellige risikokategorier: den kan enten bliver brugt ondsindet, som allerde sker nu. Lave fatale funktionsfejl også kaldt for hallucinationer. Det kan det være en risiko for demokratien, magtfordelingen, og den kan bruges som våbn. Fremtidens udfordring er derfor at kontrollere AI, samtidig med at arbejde på teknologien for at kom videre som menneskeheden.

## Litteratur

- Batra, N. (2025). *Who's Really Being Trained—Us or the AI? The Silent Paradox of Our AI Age* [LinkedIn]. Hentet 14. november 2025, fra <https://www.linkedin.com/pulse/whos-really-being-trainedus-ai-silent-paradox-our-age-navneet-batra-plqic>
- Bengio, Y., Mindermaann, S., Privitera, D., m.fl. (2025). *International AI Safety Report* (Technical report Nr. DSIT 2025/001). Department for Science, Innovation, Technology og AI Safety Institute. United Kingdom. <https://www.gov.uk/government/publications/international-ai-safety-report-2025>
- Castelvecchi, D., Thompson, B., & Bengio, Y. (2025). ‘It keeps me awake at night’: Machine-learning pioneer on AI’s threat to humanity. *Nature*. <https://doi.org/10.1038/d41586-025-03686-1>
- geeksforgeeks. (2024). *What is the difference between a strong AI and a weak AI?* Hentet 4. november 2025, fra <https://www.geeksforgeeks.org/artificial-intelligence/what-is-the-difference-between-a-strong-ai-and-a-weak-ai/>
- Kubrick, S. (1968). 2001: A Space Odyssey [Film].
- MyNewRobot. (2017). *HAL 9000 phrases*. Hentet 4. november 2025, fra <https://hal9000computer.wordpress.com/2017/11/22/all-hal-9000-phrases-from-the-movie/>
- Ramirez, V. B. (2025). *ChatGPT Is Changing the Words We Use in Conversation* [Scientific American]. Hentet 13. november 2025, fra <https://www.scientificamerican.com/article/chatgpt-is-changing-the-words-we-use-in-conversation/>
- Systime. (2021a). *Tankeeksperimentet ”Det kinesiske værelse”*. Hentet 4. november 2025, fra <https://teknologiogfilosofi.systime.dk/?id=265>
- Systime. (2021b). *Turings test*. Hentet 14. november 2025, fra <https://teknologiogfilosofi.systime.dk/?id=265>
- Titcomb, J. (2025). *ChatGPT triggers surge in MPs using AI-written speeches* [Telegraph]. Hentet 14. november 2025, fra <https://www.telegraph.co.uk/business/2025/09/11/chatgpt-triggers-surge-in-mps-using-ai-written-speeches/>
- Yakura, H., Lopez-Lopez, E., Brinkmann, L., Serna, I., Gupta, P., Soraperra, I., & Rahwan, I. (2025). Empirical evidence of Large Language Model’s influence on human spoken communication. *arXiv*, 4–7 & 26. <https://doi.org/https://arxiv.org/pdf/2409.01754>