



## Etik och mänskliga värderingar i en AI-driven värld del2: makt, moral och maskin

I förra veckans artikel, del1, låg utgångspunkten och fokus på grundfrågan - Hur påverkar AI våra värderingar? Denna vecka skulle vi vilja lyfta den och sätta en lite mer omvärlds fokus på just ansvar, kontroll, censur, deepfakes och global påverkan!

- När sanning kan manipuleras på pixelnivå
- Plattformar som agerar moralpoliser eller censurmaskiner.
- Från filterbubbla till informationsgaller.
- När AI styr våra val ... bokstavligen.
- AI som en geopolitisk kraft.
- Vem kommer bära det etiska ansvaret och hur kommer det se ut?

Deepfakes är ett ord som nog inte undgått många, men vad är deepfakes? Deepfakes är AI-genererade ljud och videoklipp som imiterar verkliga personer och har utvecklats från att vara en teknologisk kuriositet/ nyfikenhet, till ett farligt verktyg för manipulation som potentiellt sett är ett globalt hot mot demokratin. I dag kan vem som helst med ett kraftfullt grafikkort skapa trovärdiga klipp där politiker säger saker de aldrig sagt, eller där kändisar deltar i innehåll de aldrig medverkat i. Det är inte längre science fiction, det är vardag! Genom att manipulera bilder, röster och identiteter så kan falsk information få spridning liknande med blixstens hastighet och vilket underminerar tilliten till medier som olika samhällsenliga institutioner, privatpersoner, företag mm. Detta fenomen belyser behovet av kritiskt tänkande och mediekunskap, vilket är centralt i [John Deweys](#) pedagogiska filosofi. Dewey förespråkade en utbildning som främjar reflektion och aktivt deltagande i samhället, vilket är avgörande för att motverka desinformation i den digitala eran

---

Författare: © 2025 Michael Bohman. Får delas fritt med angiven källa.

LinkedIn: <https://linkedin.com/in/mibotech>

GitHub: <https://mibotechcoder.github.io/tech-and-future-reflections/>

Artikel utgiven: Maj 2025

Sida 1 av 6

**Konsekvensen är epistemologisk:** vår förmåga att lita på det vi ser och hör undermineras. I det ögonblick då vi tvivlar på allt vi ser så öppnas ett fönster för desinformation. *Det spelar mindre roll om något är sant för det räcker bara att ifrågasätta sanningen.* I psykologin talar man om "**cognitive dissonance**", när människor upplever motsägelsefull information blir det kognitivt smärtsamt, och många väljer då att tro det som stämmer bäst med deras tidigare övertygelser. Deepfakes förstärker denna mekanism.

**Pedagogiskt sett** är detta en katastrof om det inte bemöts och John Dewey framhöll vikten av att träna det kritiska tänkandet i demokratiska samhällen inte bara för att lära sig "sanningen", utan för att kunna *navigera i osäkerhet*. I dagens medielandskap behöver vi inte bara källkritik, utan *källanalytisk kompetens*, förmågan att granska teknikens roll i informationsflödet.

**Etiskt sett** står vi inför en ny typ av skada: inte fysisk eller ekonomisk, utan **ontologisk**, för vem äger en persons ansikte? Deras röst? Deras identitet? Vad innebär samtycke när AI kan kopiera din mimik och lägga dina ord i någon annans mun?

Dessutom används deepfakes allt oftare i **digitalt våld**, särskilt mot kvinnor. Fall där AI-porr skapas utan samtycke ökar snabbt och verktygen blir tillgängliga för fler. Organisationer som WITNESS och AI Forensics arbetar med att bygga ramverk för att upptäcka, märka och juridiskt bekämpa dessa övergrepp, men lagstiftningen ligger efter.

**Så vad behövs göras?** Lösningen är inte teknikförbud, utan kombinationer av:

- Teknisk märkning (t.ex. metadata och vattenmärkning)
- Plattformars ansvar för spridning
- AI-kritisk utbildning i skolan
- Tydliga juridiska konsekvenser för icke-samtyckt manipulation

I en värld där miljarder inlägg dagligen laddas upp på TikTok, Instagram, Facebook och X (Twitter) så har uppgiften att moderera innehåll till stor del överlämnats till AI och det finns flera skäl till varför, för de flesta så handlar det om *omfattning, ekonomi, juridik och hastighet*. Det låter logiskt men i praktiken skapas ett system där **kod ersätter kontext**. AI-moderering fungerar nämligen främst genom nyckelordsfilter, bildigenkänning och statistisk sannolikhet ... *inte genom förståelse!*

Konsekvensen är att *människors uttryck tolkas bokstavligt, men utan mänsklig intuition*. Det som är ironi, kulturkritik eller humor kan uppfattas som hat, det som är sårbart, kreativt eller normbrytande kan klassas som "olämpligt". Användare har därför utvecklat ett parallellt språk "**algospeak**" där kodord och stavfel används för att lura systemet. Till exempel skrivs "unalive" istället för "dead", och "corn" i stället för "porn". Det är inte bara kreativt, det är en form av motstånd mot maskinens oförmåga att förstå mänsklighet.

*Pedagogiskt är detta djupt problematiskt, **Lev Vygotskys sociokulturella teori** säger att lärande sker genom interaktion med andra och med de kulturella verktyg vi har tillgång till i detta fall språk. När språket styrs av AI-regler, förändras villkoren för mänsklig utveckling.* Det som inte kan sägas riskerar att försvinna. Det är ett hot mot både yttrandefrihet och identitet.

*Etiskt väcks frågan: Vem sätter gränserna?* I traditionell lagstiftning sker begränsningar av yttranden genom demokratiska processer. På sociala plattformar görs det bakom stängda dörrar, av kod, kommersiella algoritmer och policyavdelningar i bolag vars mål är maximera vinster inom sina väggar. Det innebär att *plattformarna fungerar som digitala moralpoliser* utan insyn, överklagandemöjlighet eller ansvarsfördelning. Det är en form av *tyst makt* och inte genom vad som sägs, utan genom vad som aldrig får synas.

*Social påverkansteori*, såsom [Michel Foucaults](#) idéer om disciplin och övervakning, blir här relevant. Människor börjar anpassa sig själva för att passa plattformens normer. Vi redigerar våra åsikter, våra bilder, vår personlighet och inte för att någon krävt det, utan för att algoritmen premierar vissa uttryck och straffar andra.

### **Så hur skapar vi ett mer rättvist system?**

- Transparens i AI-modereringspolicy och algoritmstruktur
- Möjlighet att förstå *varför* ett inlägg tagits bort
- Mänsklig granskning i känsliga fall
- Tydlig reglering kring digital yttrandefrihet (t.ex. via EU:s DSA – Digital Services Act)

AI-drivna modereringssystem på plattformar som TikTok och Meta saknar ofta förmågan att förstå kontext, ironi och kulturella nyanser. Användare utvecklar därför "algospeak" för att kringgå censur, vilket pekar på en brist i systemens design. Lev Vygotskys sociokulturella teori betonar vikten av social interaktion och kulturella verktyg i lärandeprocessen. När AI-system inte kan tolka dessa mänskliga aspekter riskerar de att bli ineffektiva och orättvisa i sin moderering.

Det som en gång sågs som en teknologisk bekvämlighet och personliga flöden, algoritmiska rekommendationer, och "du kanske också gillar detta" ... har idag blivit något helt annat: *Ett osynligt filter som formar vår förståelse av världen*. Det vi ser i våra sociala medier är inte längre en slump det är en följd av optimerad exponering. Algoritmer lär sig vad vi gillar, hur länge vi tittar, vad vi undviker, vilka vi följer och skapar en personlig verklighetsbubbla.

Men problemet är inte bara att olika människor ser olika versioner av verkligheten. Det är att dessa versioner *stärker det vi redan tror på*. Istället för att utmana oss, bekräftar AI våra existerande åsikter. Det är det som psykologer kallar [confirmation bias](#) är vår tendens att föredra information som stödjer det vi redan tror, och undvika det som skaver.

[Jean Piagets konstruktivistiska lärandeteori](#) hjälper oss förstå faran i detta. Piaget menade att lärande också kräver en aktiv process där individen assimilerar ny information och ibland också måste *ackommodera den*, alltså förändra sin tidigare förståelse. Men när algoritmerna styr bort oss från det som utmanar oss så får vi aldrig chansen att ackommodera, *vi fastnar i en kognitiv bekvämlighetszon där intellektuell utveckling avstannar*.

[Filosofen Noam Chomsky](#) beskrev i "Manufacturing Consent" hur traditionella medier formar våra åsikter genom urval och vinkling. Det som sker nu är en algoritmisk version av samma process men i individuell skala och utan redaktionell kontroll och utan att vi märker det. Det är inte längre medier som producerar konsensus, *det är kod och vi vet ofta varken vem som skrivit den, eller varför*.

Pedagogiskt sett så innebär detta en ny typ av ett digitalt medborgarskap, om våra informationsströmmar är selektivt utformade av maskiner så måste vi lära människor att aktivt söka, ifrågasätta och jämföra. Det räcker inte med källkritik, vi behöver *flödeskritik: en medvetenhet om hur algoritmiska strukturer formar vår verklighetsuppfattning*.

### **Så vad gör vi?**

- Algoritmisk transparens (visa varför du får se något)
- Möjlighet att själv justera sitt flöde
- Utbildning i "digital epistemologi" – hur kunskap formas i digitala miljöer

- Främja mångfald i informationskällor – t.ex. “slow content”-initiativ

Vi gillar att tro att vi är fria i våra beslut, men i en tid där algoritmer styr det mesta så är det sällan vi gör våra val på en neutral spelplan. AI-system påverkar redan i dag *hur* vi fattar beslut, inte genom tvång, utan genom ”nudging”, det vill säga subtila knuffar i en viss riktning. Vad vi klickar på, vilket erbjudande vi får först, vilka alternativ som visas och hur något formuleras, allt kan påverkas av systemets antaganden om vad som är mest effektivt, lönsamt eller engagerande.

Begreppet **nudging** kommer från beteendeekonomerna **Richard Thaler** och **Cass Sunstein**. De visade att människor är benägna att välja det som är lättast, snabbast eller mest utmärkande även om det inte nödvändigtvis är det bästa. När AI nu bygger hela strukturen för de val vi gör så innebär det att våra beslut också i hög grad *formas innan vi ens är medvetna om dem!* Pedagogiskt sett är detta ett avgörande, **Paulo Freire** betonade i sin frigörande pedagogik att utbildning ska hjälpa människor bli medvetna om sina omständigheter och inte bara reproducera systemets egna logik. *Men hur kan vi bli medvetna om påverkan när den är osynliggjord av kod?* AI har inga ansikten, ingen ton, ingen tydlig avsändare men den påverkar ändå våra tankar, känslor och beteenden.

Psykologiskt handlar detta om det som kallas **priming**, vi påverkas av stimuli som vi inte är medvetna om. Ett färgval, en bild, en ordningsföljd ... allt detta kan aktivera vårt beslutssystem i hjärnan utan att vi vet om det. *AI kan, baserat på vår historik, designa valmiljöer som triggar våra vanor, rädslor eller begär.* Det är inte science fiction. Det är marknadsföring i realtid.

Och i politiska sammanhang blir det än mer känsligt. Tänk dig att du och jag söker på samma ämne men får olika resultat beroende på våra olika klickmönster, eller ett känsloladdat politiskt budskap bara riktas mot vissa profilerade grupper utan insyn. Det här kallas **microtargeting** och det skedde redan under Brexit och presidentvalet i USA 2016.

### Så vad gör vi?

- Etiska riktlinjer för AI-baserad nudging
- Tydlig märkning av påverkan (t.ex. “AI-rekommenderad”)
- Undervisning i valarkitektur och beslutspsykologi i skolan
- Plattformar som låter användaren designa sin egen nudging-struktur

Vi tänker ofta på AI som något tekniskt, men det är i högsta grad väldigt **geopolitiskt**. Den som äger den digitala infrastrukturen, de olika språkmodeller, plattformar, molntjänster, dataflöden mm, de äger också en stor del av framtidens narrativ och inflytande. AI är inte längre ett verktyg bland andra. Det är själva **ramverket** för hur beslut fattas, hur idéer sprids och hur världsbilden formas. Tre stora krafter formar detta landskap i dag: **USA**, **Kina** och **EU**.

- **USA** dominerar kommersiellt, med bolag som OpenAI, Google, Meta och Amazon.
- **Kina** driver AI som en statlig prioritet, där AI används i allt från övervakning till kreditvärdering.
- **EU** försöker gå en etisk tredje väg, via lagstiftning som **AI Act** och **Digital Services Act**, för att skydda individens rättigheter.

Men det är inte bara infrastruktur som det handlar om, det handlar också om *vilken människosyn som kodas in!* Värderingar som frihet, kontroll, ansvar, integritet och säkerhet definieras olika i olika kulturer och AI-modeller kan också spegla dessa skillnader. Tex. vem får sätta ramarna om för vad som är ett “acceptabelt beteende”, “desinformation” eller “samhällsnytta”? **Svaret är enkelt: de som tränar och äger modellerna.**

Här blir **George Siemens' teori om connectivism** relevant. I hans syn på lärande i det digitala samhället så handlar kunskap inte bara om *vad du vet*, utan om *var du är kopplad!* Vår tillgång till insikt, information och delaktighet är direkt kopplad till våra nätverk, och *AI bestämmer i allt högre grad hur dessa nätverk ser ut*. Kontroll över AI är därför också kontroll över utbildning, innovation och kollektiv intelligens.

*Politiskt sett* så innebär detta att ett *skifte går från hård till mjuk makt!* Det är inte längre den med flest vapen som vinner, utan den med störst inflytande över infrastruktur, standarder och narrativ. När AI-modeller används för att generera innehåll, översätta texter, prioritera nyheter, blockera information eller skapa digitala medborgarprofiler, flyttas makten från stat och folk till plattform och kod.

### Så vad gör vi?

- Kräver öppenhet i AI-modellers värdegrund och datakällor
- Skapar nationell och lokal AI-kompetens – inte bara import av färdiga system
- Investerar i AI-utbildning och kritisk teknikförståelse på alla nivåer
- Främjar digital suveränitet – rätten att förstå och styra sin digitala miljö

När en människa gör ett misstag så kan vi oftast ställa väldigt raka frågor som, vad hände? varför? vem bär ansvaret? Men när en AI fattar ett felaktigt beslut eller kanske ett orättvist eller till och med skadligt så blir det genast mer komplicerat. Tekniken fungerar ofta som en "*black box*", data matas in och ett beslut kommer ut, men hur det beslutet fattades kan vara suddigt. Och ansvaret? När det väl är utspritt mellan utvecklare, leverantörer, användare, systemarkitekter mm, vem kan då egentligen hållas ansvarig?

**Luciano Floridi**, professor i informationsetik, menar att vi behöver ett nytt sätt att tänka kring ansvar i AI-system. Hans begrepp "*distributed responsibility*" beskriver att ingen enskild individ kan bära hela skulden men att alla måste ta sin del. AI utvecklas i komplexa nätverk av aktörer, och därför måste också etiken kollektivt förankras i design, testning, användning och i lagstiftningen. Det här utmanar vår traditionella etik, som ofta bygger på enskilda aktörers avsikter och handlingar. I AI-sammanhang är det snarare *beslutsekosystem* som vi måste förstå. Ett felaktigt beslut kan bero på bias i träningen, dåliga datakällor, tolkningen av ett prediktivt resultat eller fel i användarens förståelse. *Ingen av dessa faktorer ensam förklarar allt men tillsammans skapar de effekten.*

**David Kolbs** erfarenhetsbaserade lärandeteori betonar vikten av att kunna förstå konsekvenser genom reflektion och iterativ förbättring. I praktiken innebär det att utvecklare, beslutsfattare och användare behöver arbeta med att *testa, observera, analysera* och *förbättra*. Det räcker alltså inte med att bygga en AI, vi måste också *lära oss tillsammans av dess effekter* och här blir *systemteori och pedagogik* central. **Etiskt ansvar i AI-tid kräver alltså tre saker:**

1. **Designetik** – redan vid konstruktionen måste man ställa frågan: *Vilka värden bygger vi in?*
2. **Användareetik** – förståelse för hur beslut tolkas, används och påverkar människor
3. **Systemetik** – helhetsförståelse för var makt, bias och kontroll faktiskt ligger

### Så vad gör vi?

- Skapar tvärvetenskapliga team med både tekniker, etiker och samhällsvetare
- Utvecklar mekanismer för *AI auditing* som granskar beslut och konsekvenser
- Gör ansvarsfördelning synlig och reglerbar i varje systemkomponent
- Främjar reflekterande AI-kultur lärandecykler, feedback, och etiska forum

---

Författare: © 2025 Michael Bohman. Får delas fritt med angiven källa.

LinkedIn: <https://linkedin.com/in/mibotech>

GitHub: <https://mibotechcoder.github.io/tech-and-future-reflections/>

Artikel utgiven: Maj 2025





## Sammanfattning

AI har blivit vår tids osynliga infrastrukturella makt och inte bara ett verktyg, utan en aktör som påverkar val, informationsfilter, beteenden och världsbild. Från deepfakes till digitala bubblor, från geopolitisk dominans till frågan om ansvar. Vi ser också nu hur algoritmer formar både det synliga och det vi inte längre ser. Så en kort summering:

- **När sanning kan manipuleras på pixelnivå :** *Vi tappar förmågan att skilja sant från falskt och därmed tilliten till verkligheten i sig.*
- **Plattformer som agerar moralpolis eller censurmaskiner:** *Kod ersätter kontext. Yttrandefrihet regleras av osynliga algoritmer utan insyn eller ansvar.*
- **Från filterbubbla till informationsgaller:** *AI visar oss bara det vi redan tror på och försvårar vår förståelse, empati och kritiskt tänkande.*
- **När AI styr våra val ... bokstavligen:** *Våra beslut formas subtielt av algoritmer och innan vi ens inser att vi blivit påverkade.*
- **AI som en geopolitisk kraft:** *AI är inte bara teknik, det är makt och den som styr systemen påverkar världen.*
- **Vem kommer bära det etiska ansvaret och hur kommer det se ut?** *I ett distribuerat ekosystem krävs kollektiva ansvarsmodeller, vi måste bygga in etik i varje länk av kedjan.*

Men mitt i tekniken finns en tyst konflikt ... **Vem bär etiken? Vem kodar moralen? Och hur återerövrar vi vår roll som mänskliga, reflekterande beslutsfattare i ett system där allt går snabbare än eftertanken?**

**#AI #DigitalEtik #Digitalisering #EtiskAI #Artificiellintelligens #Digitalkompetens**

---

Författare: © 2025 Michael Bohman. Får delas fritt med angiven källa.

LinkedIn: <https://linkedin.com/in/mibotech>

GitHub: <https://mibotechcoder.github.io/tech-and-future-reflections/>

Artikel utgiven: Maj 2025

Sida 6 av 6