

AI och samhällspåverkan
- en ansvarsfull reglering

Policy brief

Introduktion och bakgrund

AI från experiment till tillämpningar

Det senaste årtiondet har artificiell intelligens (AI) gått från att vara ett forskningsområde inom datavetenskapen till att bli en del av vardagen. Den viktigaste faktorn för utvecklingen är användningen av djupinlärning (eng. *deep learning*, DL), en avancerad form av maskininlärning. Den gör det möjligt att analysera de enorma datamängder som finns spridda på webben för att skapa träffsäkra förutsägelser. Förutsägelseerna används till automatiserade beslutssystem.

En avgörande aspekt av denna spridning är fenomenet infusion, där AI tillämpas i uppdateringar av tjänster i onlineplattformar och applikationer som tidigare utvecklades utan tillgång till AI. Detta har gjort det möjligt för AI att nå miljarder användare globalt utan att dessa behöver anta en helt ny teknik.

Vad är maskininlärning och djupinlärning?

Maskininlärning är en tillämpning av AI som omfattar algoritmer som tolkar data, lär sig av denna data och sedan använder kunskapen till att fatta informerade beslut. Djupinlärning är ett delfält inom maskininlärning som strukturerar algoritmer i lager för att skapa ett artificiellt neuronnät som lär sig och fattar beslut på egen hand.

Högräckvidds-AI

Högräckvidds-AI är AI-system som når ett mycket stort antal användare och har betydande sociala, ekonomiska och politiska effekter. Dessa tekniker, som rekommendationssystem och digitala plattformar, påverkar användarnas val och handlingar på subtila men ofta omfattande sätt. Algoritmerna "knuffar" användarnas beteende i viss riktning. De skapar stora mängder datadriven kunskap som ytterligare förbättrar deras precision och effektivitet. Ett exempel är rekommendationssystemen hos sociala medieplattformar som Facebook och YouTube, som påverkar användarnas konsumtion av innehåll genom att optimera för engagemang och skärmtid.

EU:s lagstiftning

Digital Services Act DSA, förordningen om digitala tjänster (EU 2022/2064), har instiftats för att reglera utmaningarna från plattformarna och högräckvidds-AI. DSA har infört nya regler för transparens och ansvarsutkrävande. Trots dessa framsteg kvarstår luckor för användarnas integritet och tillsynsmyndigheternas möjlighet att kunna få ut uppgifter för riskbedömning.

AI:s utveckling till massmarknad

AI:s nutida framgångar kan spåras tillbaka till genombrott inom djupinlärning på 2010-talet, då tekniken började användas i allt från bildigenkänning till naturlig språkbehandling. Plattformar som Facebook, Google och Amazon tog snabbt till sig tekniken och inkluderade den på de plattformar de redan utvecklat.

Tidigare studier om teknisk acceptans

Forskningen har inriktat sig på att skapa modeller för teknologiacceptans (TAM) för att undersöka hur beteenden, attityder och värderingar påverkat individens val att anta en ny teknik, så kallad diffusion. AI:s spridning genom infusion har utmanat TAM genom att ta bort alternativet av aktiva användarval. Tekniken sprids istället genom automatiska uppdateringar. Vilka dessa uppdateringar är, och vilken funktion de ska ha, avgörs av tillverkaren utan återkoppling från användaren.

Dagens juridiska ramverk och dess begränsningar

DSA är en omfattande förordning i EU för att hantera de sociala och ekonomiska riskerna med digitala tjänster och plattformar. DSA fokuserar på att förbättra transparensen och ansvarsutkrävandet för mycket stora onlineplattformar (Very Large Online Platforms, VLOP). DSA klassificerar plattformar som har mer än 45 miljoner användare per månad i EU som VLOP:er.

De företag som klassificerats som VLOP måste årligen rapportera till Europeiska kommissionen med en riskutvärdering av sina algoritmer. Plattformarna ska även tillämpa åtgärder för att kunna bemöta de risker de identifierat. Riskutvärderingen och åtgärderna ska granskas av en oberoende revision. Europeiska kommissionen kan kräva att få ut information och starta egna undersökningar. Utvalda forskare kan begära ut data från VLOP:er.

Trots åtgärderna och riskbedömningsprocessen som instiftats finns det begränsningar för att hantera utmaningarna. Till exempel saknar DSA mekanismer för att granska långsiktiga effekter på samhällsnivå, såsom desinformation och social polarisering. De data som de stora onlineplattformarna uppger sig inte ha kunskap om, eller som företagen betraktar som en affärshemlighet, behöver inte lämnas ut. Det enligt den skrivning som DSA slutligen fick.

Globala luckor i lagstiftningen

Globalt finns det betydande skillnader mellan hur länder reglerar AI. USA:s fokus ligger ofta på innovation och företagsvänlighet, medan EU prioriterar dataskydd och

etik. Kina, å andra sidan, använder AI som ett strategiskt verktyg för samhällsovervakning och styrning. Detta splittrade landskap gör det svårt att skapa harmoniserade regler och standarder, vilket ökar risken för oreglerad spridning och svårbedömd användning av AI-system.

Relevans för samhälle och politik

AI:s infusionsdrivna spridning har djupgående effekter på frågor om demokratisk delaktighet, rättvisa och mänskliga rättigheter. När tekniken används för att styra informationsflöden och forma individuella preferenser finns det en risk att den undergräver användarnas självständighet och skapar ojämlika förhållanden. Därför krävs det en bredare politisk diskussion och samverkan mellan forskare, beslutsfattare och teknikföretag för att säkerställa att AI används på ett rättvist och etiskt sätt.

AI:s sociala och etiska utmaningar

Integritetsfrågor och användarprofilering

AI och DL möjliggör insamling och analys av stora mängder användardata. Det sker utan användarens medvetenhet och därför kan de inte lämna sitt samtycke. I opinionsundersökningar anger många användare att de sätter högt värde på sin integritet, men sedan agerar de inte efter det, och delar sin data online. Denna *integritetsparadox* drivs av att användarna ställs inför valet att godkänna datainsamling eller att inte använda tjänsterna.

Användarprofilering är processen att samla in och analysera data om individers beteenden, preferenser och demografi för att skapa detaljerade profiler. Data hämtas från både direkta källor (som registreringsuppgifter) och indirekta källor (som webbhistorik och klickbeteenden). Syftet är att anpassa tjänster, innehåll och marknadsföring för att bättre möta användarens behov och intressen. Den kan dock samla in information som går bortom vad datainsamlingen ursprungligen gjordes för. Studier har pekat på att Facebooks algoritmer kan förutsäga personliga egenskaper, som politisk tillhörighet och sexuella preferenser, baserat på digitala spår.

Algoritmer och fördomar

Rekommendationssystem har visat sig förstärka befintliga fördomar och sociala ojämlikheter. Till exempel har forskning visat att vissa algoritmer tenderar att prioritera upp innehåll eller visa annonser som förstärker negativa stereotyper. Forskning om sökmotorer har visat på en psykologisk förankringseffekt hos användare, att förlita sig i alltför hög grad på den första informationen som erbjuds

när man fattar beslut. Algoritmerna kan skapa så kallade "filterbubblor" där användarna exponeras för ett begränsat innehåll som överensstämmer med deras tidigare beteende. Risken är att det förstärker social polarisering och försvårar exponeringen för mångfald.

Centralisering av makt och ekonomi

Spridning av AI genom infusion gynnar i hög grad de stora teknikföretag som redan har tillgång till omfattande datamängder. Koncentrationen av resurser skapar en positiv återkopplingsloop där företag med mer data kan skapa bättre AI-system, vilket lockar fler användare och genererar ännu mer data. Detta riskerar att hämma konkurrens och innovation, samtidigt som det förstärker obalansen mellan stora plattformar och små företag.

Denna spridning skapar flera utmaningar:

€ Autonomi och insyn

Användare har begränsad insyn i hur deras beteende påverkas av AI. Algoritmer optimerar ofta för att maximera engagemang för att stödja annonsering, snarare än användarnas välmående, vilket kan leda till oönskade konsekvenser som överdriven skärmtid eller att utsättas för desinformation.

€ Etiska och sociala dilemman

AI kan förstärka existerande ojämlikheter genom partiskhet i de data som insamlas eller i algoritmer som favoriserar vissa grupper. Dessutom har forskningsstudier pekat på att vissa algoritmer visat sig bidra till spridning av skadligt innehåll, inklusive extremism och desinformation.

Digital Services Act (DSA) är en EU-förordning (EU 2022/2064) för att skapa en inre marknad för digitala tjänster, genom att harmonisera reglerna och skapa en tillsynsprocess. Det svenska namnet för DSA är förordningen om digitala tjänster.

Riskhantering

Sociala och demokratiska utmaningar

AI:s påverkan på autonomi och privatliv är en omdiskuterad fråga. Tekniken kan utnyttjas för att styra användares beteende genom "hypernudging", där AI-applikationer använder plattformarnas stora datamängder för att kraftfullt påverka

användarnas beteende. Det riskerar att undergräva individens möjlighet att fatta informerade beslut.

Förmågan hos AI att sprida desinformation och extremism är också en stor risk. Studier har visat att plattformar som YouTube och Facebook bidrar till spridningen av extremt innehåll genom sina algoritmer.

Etiken och infusion

AI-infusion innebär att användare kan utsättas för algoritmiska förändringar utan deras kännedom. Detta väcker frågor om hur rättvist det är att AI-leverantörer kan ändra teknologins funktion utan samråd eller varningar. Vidare är infusionsprocessen ofta driven av kommersiella intressen, snarare än av samhällets behov, vilket kan skapa en obalans mellan plattformarnas och användarnas mål.

Policyrekommendationer

Förstärk transparens och ansvarstagande

En central del är att förbättra transparensen både för användare och tillsynsmyndigheter. Det kan uppnås genom:

Vertikal transparens, ett ökat ansvar

Vertikal transparens innebär att företag som tillämpar AI-teknologier blir skyldiga att dela detaljerad information om algoritmernas utformning, syfte och potentiella påverkan med tillsynsmyndigheterna. Ett konkret steg är att det är VLOP:erna som bör ansöka om att hålla viss information konfidentiell om det ligger i allmänhetens intresse (exempelvis för att skydda spamfilters algoritmer).

Horisontell transparens, förbättrad användarinformation

För att säkerställa att användare förstår hur deras data används och hur AI-system fungerar, bör följande införas:

- € **Krav på algoritmisk förklaring** där användare får en lättförståelig sammanfattning av hur deras data påverkar vilka rekommendationer de ser. Det förenklar också att användarna kan jämföra olika tjänster och plattformar.
- € **Alternativ till profilering:** plattformar bör erbjuda användare möjligheten att välja bort algoritmisk personalisering utan att förlora tillgång till grundläggande funktioner.

Långsiktig datadelning

Inför en modell för kontrollerad datadelning där stora plattformar är skyldiga att dela vissa anonymiserade data med forskare och mindre företag. Detta kan bidra till att skapa rättvis konkurrens och stödja innovation inom AI, samtidigt som integriteten skyddas.

Ett globalt ramverk

AI:s spridning genom infusion har skapat nya utmaningar för etik och dataskydd. För att bemöta dessa behövs globala standarder som balanserar innovation och ett skydd av användarna.

Stärkt konsumentskydd

En standard för dataskydd bör etableras för att säkerställa att individer får kontroll över hur deras data används.

- € **Förbättrade rättigheter till dataportabilitet:** användare bör kunna exportera sina data från en plattform till en annan på ett säkert och användarvänligt sätt.
- € **Skärpta regler för informerat samtycke:** plattformar ska inte bara be om samtycke till datainsamling, utan också tydligt förklara syftet och de potentiella konsekvenserna av AI-drivna system.

Etiska riktlinjer

Det bör utarbetas riktlinjer som klargör när och hur AI får tillämpas i befintliga teknologier. Dessa riktlinjer kan innehålla krav på:

- € **Förhandsutvärdering:** ett AI-system som ska infuseras bör genomgå en oberoende etisk granskning innan det implementeras.
- € **Rapporteringskyldighet:** företag ska rapportera alla större förändringar i AI-systemens funktioner till tillsynsmyndigheter och användare.

Autonomisäkring

För att motverka risken att AI undergräver användarnas förmåga att fatta egna beslut, bör plattformar lägga in funktioner som gör det möjligt att enkelt anpassa algoritmiska system till användarens egna behov. Det utan att innebära extra kostnader eller minskad funktionalitet.

Praktiska tillämpningar

Här är några exempel på hur åtgärderna kan tillämpas:

Skapa incitament för ansvarsfull AI-utveckling

Idag kan företag som bryter mot DSA dömas till böter. Ett förslag är att ge incitament som belönar företag som följer bästa praxis.

- ∄ **Skattelättnader** för företag som utvecklar etiskt förankrade AI-system.
- ∄ **Certifieringssystem** som synliggör de företag som uppfyller höga standarder för AI-säkerhet och användarskydd.

Samordning mellan myndigheter och forskare

En samverkan mellan forskare, myndigheter och teknikföretag bör etableras för att övervaka effekterna av AI. Detta kan ske genom:

- ∄ **Årliga rapporter** om AI:s påverkan på samhället.
- ∄ **Samarbete kring forskning** som syftar till att identifiera risker och lösningar.

Summering och rekommendationer

Summering

AI:s spridning genom infusion till redan befintliga applikationer innebär en stor skillnad mot hur tekniska innovationer tidigare spridits i samhället, och medför att användare ofta påverkas utan att de är medvetna om det. Denna spridningsmekanism möjliggör snabb tillämpning, men medför också risker för individens autonomi, social rättvisa och samhällsstruktur. Rekommendationssystem är ett sådant exempel, då dessa kan nå ett brett genomslag och påverka allt från konsumtionsmönster till politisk polarisering.

För att ta itu med dessa utmaningar måste beslutsfattare:

- ∄ **Stärka transparenskrav**, särskilt genom att etablera robusta vertikala transparens för tillsynsmyndigheter.
- ∄ **Utveckla etiska riktlinjer** för AI-infusion som säkerställer att användarnas rättigheter respekteras.

- ∄ **Främja internationell samordning** för att hantera de globala konsekvenserna av AI:s spridning, inklusive harmonisering av regleringar och standarder.

Rekommendationer

1. Vidare forskning och empiriska studier

Forskningsområden omfattar:

- ∄ **Studier** som undersöker långsiktiga konsekvenser av rekommendationssystem på demokrati, mental hälsa och informationsspridning.
- ∄ **Identifiering** av och åtgärder mot diskriminering och ojämlikhet i AI-system.

2. Samarbete mellan forskare, myndigheter och företag

Ett effektivt regelverk för AI kräver samverkan mellan olika aktörer. Följande initiativ föreslås:

- ∄ **Etablera samordning** för att utbyta data, metoder och resultat.
- ∄ **Skapa incitament för ansvarsfull AI-utveckling:** exempelvis certifieringar och ekonomiska fördelar för företag som följer etiska riktlinjer.

3. Utveckla globalt samarbete

För att minimera risken för splittrade regelverk behövs **internationella överenskommelser**, med särskilt fokus på dataskydd och etiska riktlinjer.

För en ansvarsfull reglering av tekniken

AI:s potential är enorm, både positivt och negativt. En etisk och hållbar AI-användning är även en politisk och samhällsfråga. Det är därför avgörande att beslutsfattare säkerställer att AI används som en kraft för allmänhetens bästa.

Transparens, ansvarsutkrävande och rättvisa behöver stå i centrum för framtida policydiskussioner. Genom att kombinera forskning, policy och global samordning kan vi skapa ett ramverk som inte bara främjar innovation, utan också skyddar individer och samhällen från möjliga risker.

Referenser

Engström, E. & Strimling, P. (2020). *Deep learning diffusion by infusion into preexisting technologies - Implications for users and society at large*. *Technology in Society*. 63, 101396. <https://doi.org/10.1016/j.techsoc.2020.101396>

Engström E. (2022) AI's Fast and Furtive Spread by Infusion into Technologies That Are Already in Use- A Critical Assessment i Hanemaayer, A. (red.), *Artificial Intelligence and Its Discontents, Social and Cultural Studies of Robots and AI*, Springer International Publishing s. 63-82. https://doi.org/10.1007/978-3-030-88615-8_4

Söderlund, K. Engström, E. Haresamudram, K. Larsson, S. Strimling, P. (2024). *Regulating high-reach AI: On transparency directions in the Digital Services Act*. *Internet Policy Review*, 13(1). <https://doi.org/10.14763/2024.1.1746>

För mer information

- € Forskningsprojektet **Ramverk för att förutsäga hur artificiell intelligens sprids** <https://www.iffs.se/forskning/forskningsprojekt/ramverk-for-att-forutsaga-hur-ai-teknik-sprids/>
- € **Pontus Strimling**, Projektledare Institutet för framtidsstudier. Email: pontus.strimling@iffs.se