

Artificiell intelligens

7 september, Domkyrkocentrum
Dino Viscovi, Linnéuniversitetet, Växjö

Begränsade kunskaper om AI.

1. I gott sällskap (ingen kan allt, hög grad av specialisering).
2. Jag är medieforskare (vet inte hur en telegraf är byggd, inte en tv heller eller hur tidningspapper framställs).
3. Om framtiden vet vi väldigt lite, särskilt när det gäller AI.

Se det här som en nybörjarkurs av en relativ nybörjare.

Etymologi, begreppsanalys, vi tittar närmare på orden:

Artificiell + Intelligens = AI

– artificiell av latinets *artificialis*, konstmässig, dvs. konstgjord

– intelligens, latinets, *intellego*, förstå, fatta

Ett konstgjort intellekt?

Finns det ett digitalt tänkande?

Datorer var i början knappast maskiner som kunde utföra självständiga tankeoperationer, även om de tillskrevs sådana förmågor (antropomorfisering).



Tänkande August ...



Datorminne ...

Datorer har historiskt sett inte kunnat tänka.

Det vi har kallat *minne* är snarare ett *lager* eller *register*.

Miniräknaren har också gett stöd för att datorer kan tänka.

Men räknaren liksom fullskaliga statistikprogram följer bara instruktioner; de sysslar med räkning på begäran, inte med matematik.

Min dator gör statistiska beräkningar på nolltid som förr tog dagar eller veckor att genomföra.

Table 1. Logistic regression: Physical access.

	Model 1				Model 2			
	β	S.E.	OR	Sign	β	S.E.	OR	Sign
Gender								
(Female)	—	—	—	—	—	—	—	—
Male	-0.377	0.214	0.686		-0.311	0.28	0.733	
Age	-0.192	0.018	0.826	***	-0.172	0.02	0.842	***
Education								
(Short)	—	—	—	—	—	—	—	—
Middle	0.861	0.222	2.366	***	0.570	0.24	1.768	*
Long	0.880	0.299	2.412	**	0.691	0.31	1.995	*
Income								
(-300,000)	—	—	—	—	—	—	—	—
301,000-600,000	1.056	0.247	2.874	***	0.949	0.27	2.584	***
601,000	1.232	0.745	3.428		0.946	0.76	2.575	
Marital status								
(Single)	—	—	—	—	—	—	—	—
Couple	0.403	0.238	1.479		0.359	0.25	1.432	
Association membership								
(Non-member)	—	—	—	—	—	—	—	—
Member	0.580	0.206	1.786	**	0.312	0.22	1.367	
Residence								
(Rural area)	—	—	—	—	—	—	—	—
Urban area	0.338	0.287	1.402		0.177	0.31	1.194	
Computer use in working life								
(No work experience)	—	—	—	—	—	—	—	—
Work experience	—	—	—	—	1.803	0.23	6.070	***
Constant	14.407	1.383	—	***	12.568	1.47	—	—
Nagelkerke's R^2	0.45	—	—	—	0.52	—	—	—
Correct classification	89.1	—	—	—	89.5	—	—	—

Notes: Individuals without access have reported a lack of PC, laptop, smartphone, tablet, e-book reader, smart TV, or any other device that enables internet usage; OR stands for "odds ratio"; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Siffrorna, noll till nio, de fyra räknesätten – matematiken är ett artificiellt språk: entydigt och konsekvent.

Därför blev datorerna snabbt bra på kalkylering och beräkningar.

Naturliga språk, som svenska, japanska eller urdu, har varit en betydligt svårare utmaning. För naturliga språk är varken konsekventa eller entydiga.

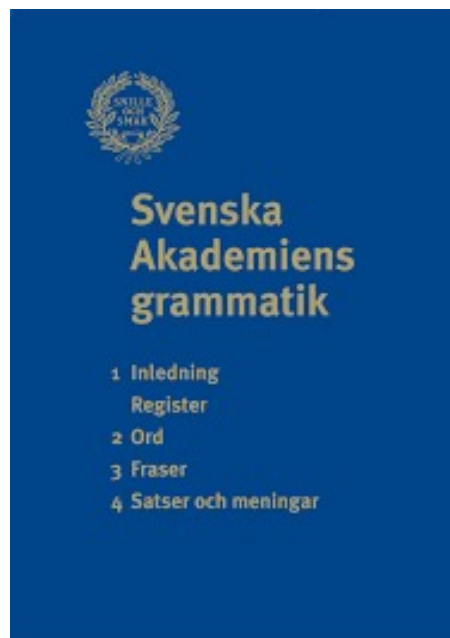
Det har tagit tid att skapa fungerande översättningsprogram – och man var efter 30 år av försök tvungen att byta spår.

I förstone verkade allt enkelt. I grund och botten består ju språk bara av ett ordförråd, **lexikon**, och regler hur för hur orden böjs och fogas till satser, **grammatik**.



lexikon

+



grammatik

= språk

Nu visade det sig vara mycket svårare än förväntat att skapa
översättningsprogram.

Grammatiken är nämligen långt mer omfattande och vidsträckt än
det som står att läsa i grammatikböcker.

- Orden är inte entydiga inom ett och samma språk.
- Och olika språk har utvecklat olika ordförråd och olika grammatik.

Att programmera in grammatik och lexikon, och den vägen få datorn
att översätta, fungerade helt enkelt inte.

Språkforskarna vid Göteborgs universitet har varit pionjärer inom datalingvistik. De skriver bl.a. det här på sin hemsida:

”Beskrivningen av ett språk är inte alltid rätt eller fel, eftersom språk har växt fram och förändrats under lång tid, och fortsätter att förändras. Dessutom ligger det i språkens natur att delvis vara flertydiga. Beskrivningen av språket är alltså en modell av en del av språket, och kan som sådan aldrig vara heltäckande.”

”På Språkbanken har vi utvecklat en ordklasskategorisering för svenska språket från 1900-talets andra hälft och framåt.”

”Modellen omfattar 13 övergripande ordklasser, nämligen adjektiv (AJ), adverb (AB), substantiv (NN), egennamn (EN), pronomen (PO), symboler (SY, inkluderar skiljetecken), verb (VB), interjektioner (IJ), samordnare (KO), numeral (NU), preposition (PE), underordnare (SU), samt utländskt ord (UO). Artiklar skiljs alltså inte ut som en egen ordklass, inte heller verbpartiklar, particip, ordningstal eller infinitivmärke.”

Redan kategoriseringen kan alltså vara svår.

Men google-översätt funkar ju ganska bra?

Eftersom mat är nödvändigt, äter jag varje dag →

Because food is necessary, I eat every day.

Svenskan har bisatsordföljd. Vänder vi på steken blir det:

Jag äter varje dag, eftersom mat är nödvändigt.

Men det fixar google-översätt genom **maskininlärning**.

Det går förvisso att lära datorn denna relativt enkla regel – att svenskan har bisatsordföljd. Men till sist blir det för många regler och undantag. Och än är inte allt perfekt.

– *Homonymer* kan t.ex. vara svåra att hantera.

Han gillar banan!



Det visar sig att google-översätt väljer att skriva:

”He likes the track”.

Men skriver jag: ”Han gillar banan och glass” blir det rätt:

”He likes banana and ice cream.”

Har google tolkat? Vi skulle ju själva kunna läsa fel, men inte om vi ser sammanhanget med glass. Då förstår vi att det handlar om något man äter och kan kombineras med glass.

”Från 1980-talet och framåt började datalingvisterna inse värdet av ... [statistisk] information. Tidigare hade de ägnat sig åt att konstruera regler som kunde *beskriva hur* en viss språklig uppgift skulle utföras. Nu började dessa experter **träna statistiska modeller** som kunde **förutse** att en person skulle utföra en uppgift på ett specifikt sätt. Detta innebar en kursomläggning för hela det språkteknologiska fältet, från förståelse till imitation – från hur-kunskap till att-kunskap.”

Paulson & Scott (2018) s 174

Google-översätt har inte förstått, som vi människor, att glass och banan hör ihop. Däremot har google gjort en sannolikhetsberäkning och föreslår det som med högst sannolikhet uppträder tillsammans med glass, nämligen frukten vi äter snarare än underlaget vi springer och utövar idrott på.

För att AI ska fungera krävs mycket stora mängder **träningssdata**.

Ju större material, desto säkrare förutsägelser, prediktioner.

Wikipedia ger en ganska bra beskrivning:

Maskininlärning (engelska: machine learning) är ett område inom artificiell intelligens, och därmed inom datavetenskapen. Det handlar om metoder för att med data "träna" datorer att upptäcka och "lära" sig regler för att lösa en uppgift, utan att datorerna har programmerats med regler för just den uppgiften. Området angränsar till statistik, datorseende och mönsterigenkänning. Datautvinning (datamining) är ett betydelsenära begrepp som avser en kombination av maskininlärning och statistiska metoder för att upptäcka och visualisera mönster i stora mängder data.

Induktion snarare än deduktion.

Låt oss ta det från ett annat håll – reklam och förslag.

*– När jag bodde i
studentlägenhet fick
jag emellanåt
direktreklam av för
mig tvivelaktig
relevans ...*

Flymo

Luftkuddegräsklipparen
för villa och park



Lätt Stark Snabb

*Det bevisade en gammal
sanning bland
marknadsförare: 'hälften
av alla annonspengar går
rakt ner i soptunnan, vi
vet bara inte vilken hälft.'*

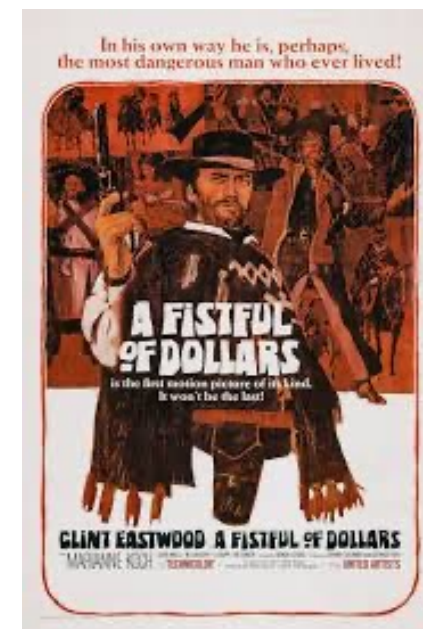
Även i den digitala tidsåldern kan reklam ha tveksam träffsäkerhet.
Jag beställer en bok på nätet, men blir trots det jagad i veckor med
annonser för precis samma bok. Ibland blir det extra märkligt ...



Netflix har varit föregångare.

- Började som uthyrare av DVD-filmer som skickades hem per post till kunderna.
 - Gick sedan över till strömmande tjänster.
- Netflix vill att så många som möjligt ska se så mycket film som möjligt på deras plattform.
 - Därför har de strävat efter att få fram bra **rekommendationstjänster**.
 - *Sökmotorer* bygger på att användaren har idéer.*
- Medan *rekommendationsmotorer* föreslår idéer för aktion.
 - * (Även sökmotorer styrs av dina sökmönster).

Ville gillar western-filmer.
Han har sett *Den gode, den onde, den fule* och
För en handfull dollar.
Nu får han ett nytt förslag:
För några få dollar mer ...



Ville kan också få förslag på fler filmer av Sergio Leone eller fler filmer med Clint Eastwood, t.ex. *Dirty Harry*, – och fler western-filmer.

Detta är enkelt, det måste gå att göra bättre än så, ansåg Netflix.

Lösningen lyder **betingad sannolikhet**.

Ett enkelt räkneexempel.

Anna har sett *Tillsammans* i sin helhet, och gett den ett bra betyg.

I databasen har 100 användare som sett *Tillsammans* också sett *En man som heter Ove*.

	Gillar <i>Tillsammans</i>	Gillar <i>En man som heter Ove</i>
Användare 1	ja	ja
Användare 2	ja	nej
Användare 3	nej	ja
Användare 4	nej	nej
Användare 5	ja	ja
Användare 6	ja	nej
Användare 7	ja	ja
Användare 8	ja	ja
... 100		

Fyra möjliga kombinationer ...

	Gillar Tillsammans	Gillar inte Tillsammans
Gillar En man som heter Ove	56	6
Gillar inte En man som heter Ove	14	24
	70	

- 56 användare gillar båda filmerna
- 14 personer gillar Tillsammans, men inte Ove
- De är 70 totalt
- $56/70 = 0,8$ eller 80 procent

Den betingade sannolikheten att en person som tycker om *Tillsammans* också gillar *Ove* är 80 procent. Filmen kan nu rekommenderas – ett bra förslag åtta gånger av tio.

Nu har ju Netflix tusentals filmer och serier att erbjuda och miljontals användare. Därför blir matematiken klart mer komplicerad – kombinationsmöjligheterna är oöverskådliga.

Men principen är densamma; baserat på användarnas historik modelleras mönster fram som gör att Netflix kan ge ganska bra förslag (givet *en* användares konto).

Netflix har mer kunskaper än bara titelnivå. De vet vilka filmer du slutar att titta på, vilka du snabbspolar. Och de vet en hel del om filmerna du ser: om de innehåller far- och son-konflikter, om de har ett högt tempo, inslag av action, romantik, har historiska teman och mycket mer.

Därför kan Netflix rekommendera filmer du alls inte känner till och anar att du kan gilla.

På samma sätt kan Spotify upptäcka att du i princip bara lyssnar på kortare låtar med konventionell harmonik, i medeltempo och med tydliga refränger.

Därför tipsar den inte om Frank Zappa och Joni Mitchell.

Många musiklyssnare kan inte själva identifiera dessa **latenta egenskaper** – men det kan Spotify.

- Kräver stora datamängder.

- Kräver träning.

- ”Är du nöjd med svaret?”

Idag är det ganska få företag som har stora datamängder: Facebook, Google, Amazon, t.ex.

- Alla som använder Facebook arbetar gratis åt Facebook.
- Alla skolor som använder Google Classroom bidrar till att Google idag har det bästa materialet för att analysera svenska elevers prestationer. De vanligaste stavfelen, den genomsnittliga meningslängden, vad de drömmer om, vilka yrkesförväntningar de har. Och latent egenskaper vi inte känner till!
- Exemplet kunddata ...

Statistikern ”... körde test efter test, analyserade data, och snart dök några användbara mönster upp. Till exempel lotioner. Många människor köper lotion, men en av [statistikerns] kollegor märkte att kvinnor i babyregistret köpte större mängder oparfymerad lotion i början av graviditetens andra tredjedel. En annan analytiker noterade att någon gång under de första 20 veckorna laddade gravida kvinnor på kosttillskott som kalcium, magnesium och zink. Många kunder köper tvål och bomullstussar, men när någon plötsligt börjar köpa massor av doftfri tvål och extra stora påsar med bomullsbollar, förutom handsprit och tvättlappar, indikerar det att de kan närma sig datumet för nedkomsten.”

Artikel, Forbes, 16 feb 2012:

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did,

Översättning, Google och DV.

Principen är egentligen enkel: om vi känner A, B, C och D kan vi förutsäga E.

Under förutsättning att ...

... vi kan göra många observationer,
och kan träna AI, dvs. ge bekräftelse när det blir rätt.

Patientjournaler – vi kan hitta nya mönster, nya samband,
latenta egenskaper.

Personer med A, B, C och D riskerar att drabbas av E. —> sätt in
förebyggande åtgärder eller tidig behandling!

AI kan hamna i fel händer och används för fel syften ...

”... för många började varningsklockorna att ringa i efterdyningarna av det amerikanska presidentvalet 2016, när det blev uppenbart hur hur oerhört skickligt Ryssland hade exploaterat Facebooks riktade annonssystem för att underblåsa konflikter bland de amerikanska väljarna. De ryska agenterna hade bland annat riktat in sig på en grupp användare som hade uttryckt sitt stöd för polisen i samband med Black Lives Matter-protesterna. Till dessa användare skickade de ut en annons med en flaggdraperad kista från en polisbegravning ...

... och bildtexten: 'Ännu ett fasansfullt dåd mot polisen utfört av BLM-aktivist. Våra hjärta blöder för dessa 11 hjältar.'



Being Patriotic

Sponsored · 🌐

👍 Like Page

Boston police shot and killed a man wearing body armor and wielding an assault rifle who critically injured two officers responding to a domestic disturbance call late Wednesday, according to Police Commissioner William Evans.

A gun battle raged at an East Boston home as a suspect, Kirk Figueroa, 33, of East Boston, critically injured two Boston police officers late on October 12. He was then shot and killed by other officers who ran into the home to drag out their wounded col... [See More](#)



I en annan annons, riktad till en grupp konservativa kristna användare, kunde Hillary Clinton ses skaka hand med en kvinna i huvudduk, åtföljt av en text med pseudoarabiska bokstäver: Stöd Hillary. Amerikanska muslimer.”

Paulson & Scott (2018) s

54



Kvinnan till vänster är
vinnaren av Nobels
fredspris, Tawakkul
Karman.

Det behöver inte vara
falska nyheter – det går
att bedriva propaganda
genom ensidighet.



Jag bad ChatGPT sammanfatta presentationen:

1. Texten diskuterar etymologin och begreppsanalysen av "Artificiell Intelligens" (AI), med fokus på orden "artificiell" och "intelligens."
 2. Den betonar att tidiga datorer inte kunde utföra självständiga tankeoperationer och att termen "minne" i datorer snarare refererade till ett register än mänskligt minne.
 3. Texten påpekar att datorer har historiskt sett inte kunnat "tänka" utan följer bara instruktioner, även om de nu kan utföra avancerade beräkningar snabbt.
 4. Den beskriver utmaningar med översättning av naturliga språk och hur maskininlärning används för att förbättra översättningstjänster.
 5. Texten diskuterar utvecklingen av maskininlärning och dess tillämpning inom områden som rekommendationssystem för streamingtjänster som Netflix och musikplattformar som Spotify.
 6. Den nämner att stora datamängder och träning av algoritmer krävs för att uppnå bra resultat inom AI.
 7. Texten tar upp hur företag som Facebook, Google och Amazon samlar och använder användardata för att förbättra sina tjänster.
 8. Den illustrerar användningen av statistik och AI för att förutsäga mönster, inklusive exempel på hur detta användes av företag som Target för att rikta annonser.
 9. Slutligen, texten varnar för att AI kan missbrukas för fel syften, inklusive politisk propaganda och desinformation.
- Hoppas detta sammandrag hjälper! Om du har några specifika frågor eller behöver mer information om något av dessa punkter, är du välkommen att fråga.

Jag bad också ChatGPT göra faktakoll:

- Presentationen innehåller inte uppenbara felaktigheter. Den ger en allmän diskussion om ämnen relaterade till artificiell intelligens (AI), maskininlärning, översättningstjänster, rekommendationssystem och användningen av data för att förutsäga mönster. Den beskriver också exempel på hur dessa teknologier och metoder används.
- Det är dock viktigt att komma ihåg att texten kan vara något förenklad eller generaliserad, och det kan finnas ytterligare detaljer och nyanser i varje diskuterat ämne. Teknik och forskning inom AI och maskininlärning utvecklas kontinuerligt, så vissa aspekter som nämns i texten kan ha förändrats sedan den skrevs.
- Om du har specifika frågor om något som nämns i texten eller om du vill ha mer detaljerad information om ett visst ämne, är jag här för att hjälpa till med så mycket information som jag har tillgänglig fram till september 2021.

Avslutningsvis, ingen vet vart vi är på väg.

Inte ens ChatGTP.

Vad vi däremot vet är att vi måste lära oss mer.

Och hela tiden värna medborgarnas integritet,

och verka för en demokratisk utveckling.

AI är som farligast i tjänst hos auktoritära stater.