

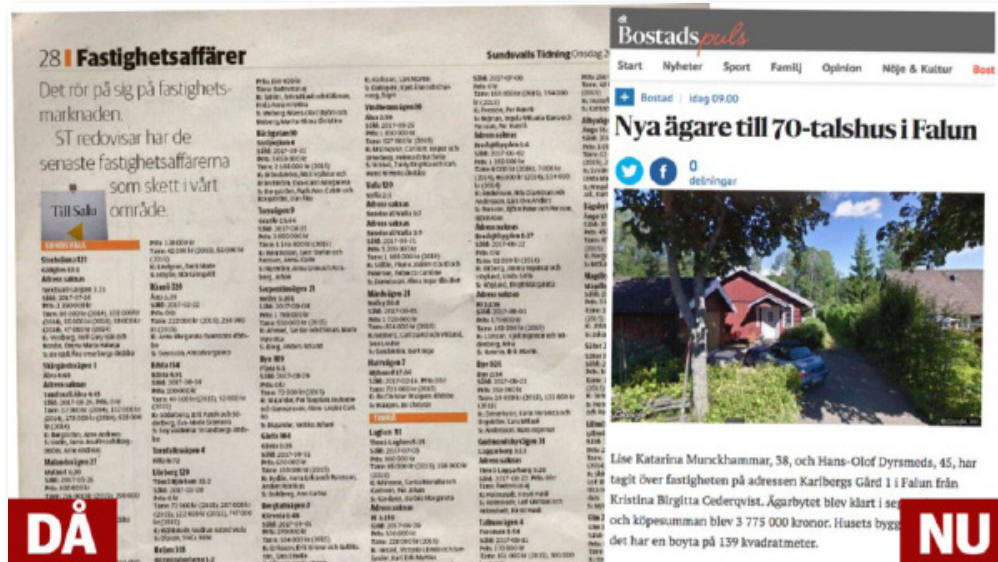
APRIL / UTVECKLING

Medie
världen
Premium

Analys/Sakari Pitkänen

Robotjournalistikens år

Förra året avancerade robotjournalistiken på bred front både i och utanför Sverige. Men tekniken är inte ny. Sakari Pitkänen analyserar den automatiska journalistikens genombrott.



SAKARI PITKÄNEN

Egen företagare och mediekonsult

Tidigare chefredaktör för Metro Sverige och Metro International. Även digital utvecklingschef och vice vd för Metros internationella webbverksamhet. Pr-konsult 2009–2012.

» Bland nomineringarna till Årets dagstidning 2018 fanns två robotar. Sirens "Robotea" och Mittmedias husköpsrobot "Fastighetskollen". Det var bara ett tecken på att 2017 blev robotjournalistikens genombrottsår. Andra var till exempel:

- ✳ I Storbritannien lanserade nyhetsbyrån Press Association (PA) den hyperlokala robotreportern "Radar". Den ska leverera 30 000 lokalvinklade artiklar i månaden till lokala tidningar i England.
- ✳ I Sydkorea lanserade nyhetsbyrån Yonhap en robot för att bevaka Premier League i England.
- ✳ I Finland byggde Helsingin Sanomat en robotlösning för att skapa unika texter för alla kommuner i 2017 års kommunalval.
- ✳ I Sverige lanserades Newsworthy som skapar obegränsat antal lokala vinklar ur öppna data, oftast statistik. För övrigt, samma idé som PA:s, fast förmodligen mycket billigare. PA har fått 800 000 euro från Google för att bygga sin lösning. J++ som står bakom Newsworthy, har gjort sin själv.
- ✳ Stampen lanserade Superlokalt (som undertecknad varit lite inblandad i).
- ✳ DN startade "halvautomatiska" lokala nyhetsbrev i Stockholm.

Detta är bara en del av många exempel i Sverige och utomlands. Frågan är varför det hände 2017. Tekniken NLP och NLG (se faktaruta) är gammal och i USA har AP låtit robotar skriva finansnyheter sedan länge.

FAKTA

AI. Paraplybegrepp för flera olika tekniker. Bland andra:

*NLP – Natural language processing. "Läser" och analyserar text.

NLG – Natural language generation. "Skriver" text.

ML- Maskininlärning. Program som lär sig av indata. Används bland annat i stor skala för att avgöra om en e-post är spam, översättningar och förutsägelser, "baserat på det här historiska mönstret är det sannolikt att det här kommer att hända" och bildigenkänning.



"Det låg väl i tiden. Om 2016 var ett år av snack om automatisering och robotjournalistik var 2017 ett år av verkstad. Idéer sattes i rullning. Mot bakgrund av den växande digitala kompetensen inom journalistiken var det ganska väntat. Fler och fler rekryterar utvecklarkompetens på redaktionerna, vilket möjliggör sådana här idéer", säger Jens Finnäs på journo-tech nätverket J++.

Sören Karlsson, vd, på United Robots som skapat de sport- och borobotar som lokalpressen i landet använder, pekar på tre saker som gör att tecknen just nu står rätt för automatisk journalistik.

1. Det finns tillräckligt med data. EverySport har till exempel mer än 15 års samlade sportresultat från elit till gärdsgårdsserier i nästan alla sporter.
2. Molntjänster har gjort det billigt att behandla stora datamängder.
3. Smartphones. Användarna kan identifieras och därmed få innehåll anpassat till sina specifika krav.

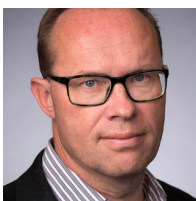
"Om du vill personalisera nyhetsutbudet måste du automatisera produktionen av nyheterna", säger Sören Karlsson.

Nästan alla redaktörer som arbetar med robotjournalistik säger att den inte ersätter "riktiga" reportrar.

Men skulle man – om man startade en nyhetstjänst i dag – inte sträva efter just det? Skulle man kunna starta en nyhetstjänst utan reportrar?

Typiskt nog är den som svarar ja på den frågan utvecklare och inte journalist. Martin Carlsson har helt på egen hand utvecklat den hyperlokala applikationen Superlokalt som Stampen nu lanserat.

"Absolut! Jag tror det vore ett mycket bra komplement till de traditionella nyhetstjänsterna. Det är viktigt att det inte blir för mycket aggregering av annat journalistiskt material, utan en tjänst som skriver texter från 'grunddata' i så stor utsträckning som möjligt", säger han.



Det är skillnad på att starta något helt nytt och att vårda och utveckla något existerande som redan är laddat med vissa värden. Martin Jönsson, utvecklingschef på Dagens Nyheter företräder den skolan.

"Ren robotjournalistik är naturligtvis väldigt intressant, men för DN är det också ett verktyg vi använder med måtta. Mest för att det som är vår särart, och som bäst driver vår affärsmodell, snarare är motsatsen: det är unik originaljournalistik, gräv, storytelling och starka personliga profiler."



"OM DU VILL
PERSONALISERA
NYHETSUTBUDET MÅSTE
DU AUTOMATISERA
PRODUKTIONEN AV
NYHETERNA"

SÖREN KARLSSON



Cecilia Frisk, publisher för TTELA, och Johannes Nilsson, nyhetschef.

2018-02-05 15:41

Stampen lanserar superlokal plustjänst

En ny, superlokal bevakningstjänst ska stärka Stampens pluserbjudande och ge ökade läsarintäkter. Cecilia Frisk, publisher på TTELA som är först ut, berättar om funktionerna.

Delia f Delia t

Stampen-ägda TTELA har varit pilot för projektet Superlokalt, som förra tisdagen lanserades brett för tidningens alla pluskunder. Den nya bevakningstjänsten samlar information och nyheter.

– Tjänsten ger pluskunderna en summering och gör det möjligt att läsa flera delar av en nyhet utan att man behöver besöka massor av olika sajter, säger TTELA:s publisher Cecilia Frisk och fortsätter:

Medie
världen
Premium

De flesta projekt som genomförts hittills har det gemensamt att de baseras på väl strukturerad data. Sportresultat, börsinformation, väder och bostadsdata.

Nyhetsbyrån Siren arbetar med stökigare data. Domar och beslut från domstolar och myndigheter kan komma i alla möjliga former och format. För att de överhuvudtaget ska gå att använda behöver data normaliseras.

Det som gör Siren intressant är att de automatiserat en annan del av det journalistiska hantverket än skrivandet. De har robotifierat inhämtandet av information. Så här beskriver de processen:

"Ett program kopplat till Robotea ("skrivroboten" min anmärkning) begär ut kopior av offentliga handlingar på myndigheter. Ett annat går igenom e-post och laddar ned bilagor, medan ett tredje skript delar upp, kvalitetssäkrar och bearbetar filerna till läsbara PDF: er. En AI-motor nyhetsvärderar sedan handlingarna och bestämmer vilken förädlingsprocess som passar bäst för respektive nyhet."

"Förädlingsprocessen" kan innebära en automatisk sammanfattning eller att en mänsklig redaktör tar vid.

När Deep Minds Alpha Go 2016 vann mot stormästaren Lee Sedol i brädspelet GO användes oövervakad djupinlärning. Det innebär att datorn inte får facit på vad den förväntas göra med inmatad data. Alpha Go fick själv räkna ut hur den skulle slå världens bästa GO-spelare. Motsatsen är övervakad, där systemet får en uppgift, till exempel att ur en bild, vilken som helst, känna igen en katt.

Det mesta som görs inom automatiserad rapportering hanterar alltså strukturerad data. De riktigt stora nyheterna göms förmodligen i ostrukturerad information. Rapporter, analyser, protokoll och anteckningar.

Vad skulle hända om man satte en Alpha Go-liknande lösning på till exempel Palmeutredningens samtliga texter? "AI löste Palmemordet"?

"Haha, tveksamt. Möjligen skulle man kunna hitta nya leads, men knappast lösa mordet", säger Jens Finnäs, J++.

Martin Carlsson som, förutom Stampens Superlokalt, driver bolaget Infeed vill tona ned förväntningarna på tekniken.

"Bakom varje framgångsrik AI står en massa människor och ger relativt tydliga instruktioner även 2018. Att göra entity recognition på personnamn och platsnamn i alla tips till Palmeutredningen och sedan analysera det hade varit intressant. Men jag tror inte så mycket hade hänt – tron på AI är just nu överdriven i det korta perspektivet."

"Entity recognition" är ett sätt att inom maskininlärning klassificera och skapa relationer mellan till exempel namn och platser.

Det har dock gjorts försök att strukturera även stökiga data som händelser. Projektet Structured stories försökte göra just det genom att granska biljakter i Los Angeles. De konstaterade att artiklarna hade många saker gemensamt:



FOTO: DAVID LAGERLÖF

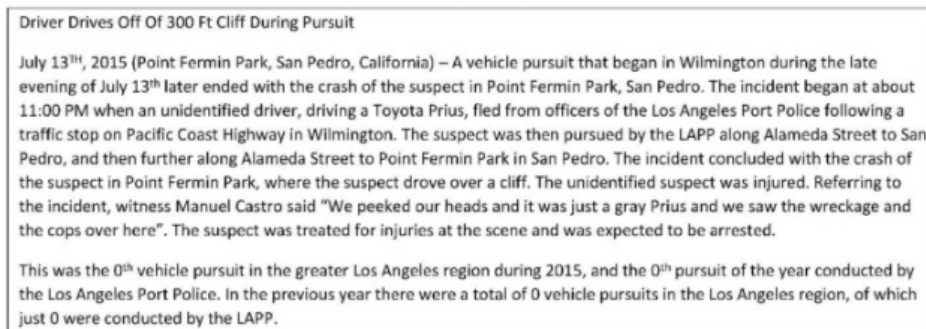
Sirens CTO Martin Fredriksson har haft en viktig roll under utvecklingen av Robotea.

"DET SOM GÖR SIREN
INTRESSANT ÄR ATT
DE AUTOMATISERAT
EN ANNAN DEL AV
DET JOURNALISTISKA
HANTVERKET ÄN
SKRIVANDET"

De inleddes på en plats vid en tidpunkt. Det fanns en jagad och en som jagade. Det fanns adresser där jakten pågick och fordon jagades och det fanns ett slut och ibland en konsekvens.

Alla dessa delar går att klassificera, skapa olika fack för. Dessa fack kan i sin tur öppnas och mappas mot fraser och därmed skapa en artikel.

Här är en sådan automatisk artikel skapad från en polisrapport.



Projektet är numera nedlagt sedan eldsjelen David Caswell konstaterat att det inte gick att finansiera Structured stories som forskning inom journalistik. Han applicerar det i stället på underrättelsevärlden.

Visste ni att det finns open-source för underrättelsepersonal: spioner agenter et consortes?

OSINT, Open Source Intelligence är just det.

Anders Bjurström, som för 12 år sedan arbetade på Aitellu och då bland annat hjälpte Metro med robotjournalistik.

"Det vi gjorde med Metro var före sin tid. De flesta mediehus var skeptiska till att en robot eller algoritm skulle styra vilka nyheter som skulle visas på nyhetssajterna. I dag är man mer bekväm med det."

Han pekar på flera akademiska framsteg under 2 000-talets senare del som kommer att ge ännu fler tjänster baserat på AI och maskininlärning.

"Inom kort borde system som lyssnar på nyheter som ljudupptagningar och Youtube-klipp kunna förstå vad som sägs, och utifrån detta momentant skapa nischade textnyhetsinslag. Då även talsyntes inom kort kan låta som en människa kommer även robotjournalistik kunna skapa 'pratnyhetsinslag', säger Bjurström, som nu driver AI-bolaget Tenfifty.io."



Vad händer i den nära framtiden inom er del av den svenska robotjournalistiken?

Martin Carlsson, Infeed:

"Vi intensifierar utvecklingen av Localixplattformen (som driver Stampens Superlokalt). Den håller på att integreras med vår plattform Infeed för textgenerering samt med fler källor.

Matti Larsson, Siren:

"Vi fortsätter att trimma systemen och tillföra nya nyhetsflöden. Vi kommer också att lansera lite enklare 'robotar' för lokal ekonomibevakning.

Martin Jönsson, Dagens Nyheter:

"Mer fokus på prediktiva verktyg. Verktyg som likt Washington Posts Virality kan identifiera vilka artiklar man förväntar gå särskilt bra utifrån metadata och sedan kan mäta av detta mot verkligheten. Effektivare botar (jobbar just nu mycket med att få fram bra slack-botar för att effektivisera omvärldsbevakningen ytterligare), natural language-förfining med mera.

Jens Finnäs, Jplusplus:

"Vår stora lärdom under våren har varit att utmaningarna med nyhetsautomatisering inte nödvändigtvis är primärt tekniska, utan snarare mänskliga. Hur får vi nyhetschefer att tänka i nya banor? Hur förändrar vi invanda strukturer? Hur löser vi riktiga problem och vinner förtroende?"

"Mot bakgrund av detta hoppas vi att 2018 kommer att handla mer om människor och mindre om kod. Jag är nyfiken på att under året kunna testa och utvärdera hur vi kan öka den lokala betalningsviljan med hel- eller semiautomatiserat innehåll."

**Medie
världen
Premium**

Robotjournalistik – så funkar det

IFK Göteborg – AIK 1–2 (1–0)

IFK Göteborg: Thomas Olsson, 32. 1–1. AIK: Antônio Flávio 55, Daniel Tjernström, 85. Publik 17 233. Domare Martin Hansson.

En normalt sportintresserad person kan av detta resultat skapa en text i stil med:

"AIK vann bortamötet mot Blåvitt med uddamålet. 17 233 åskådare fick se en match som var spännande ända fram till slutet. IFK tog dock ledningen i 32 minuten genom Thomas Olsson. Ledningen stod sig halvlaiken ut. Först en bit in i andra lyckades bortlaget komma i kapp då Antônio Flávio kvitterade. Med endast fem minuter kvar av ordinarie matchtid klev Daniel Tjernström fram och gjorde segermålet, 2–1 till AIK, ett resultat som stod sig hela vägen fram till dess att domare Martin Hansson blåste av matchen."

Den sportintresserade har här kompletterat resultatinformationen med vanligt språk och gjort några logiska slutledningar som att IFK:s ledning stod sig halvlaiken ut. Det gjordes ju inga mål mellan minut 32 och 45.

Exakt hur den mänskliga hjärnan lär sig detta är en gåta.

Robotar använder sig i alla fall av regler. Uppfär så här:

OM: tiden_för_mål = 85

SKRIV: "med endast 5 minuter kvar av ordinarie matchtid".

OM tiden_för_mål = 90

SKRIV: "I sista ordinarie matchminuten"

OM tiden_för_mål < 90

SKRIV: "på tilläggstid"

```
* @var boolean
*/
define('PSI_INTERNAL_XML', false);

if (version_compare("5.2", PHP_VERSION, ">")) {
    die("PHP 5.2 or greater is required!!!");
}
if (!extension_loaded("pcr")) {
    die("phpSysInfo requires the pcr extension to php in order to work properly.");
}

require_once APP_ROOT.'/includes/autoload.inc.php';

// Load configuration
require_once APP_ROOT.'/config.php';

if (!defined('PSI_CONFIG_FILE')) {
    $tpl = new Template("/template");
    echo $tpl->fetch();
    die();
}
```

Graden av språklig sofistisering är beroende av hur många regler man orkar skriva och hur mycket data det finns att basera dem på.

Skönheten ligger i att när en regel är skriven kan den appliceras på alla – i det här fallet – matcher alltid. Det gör att en robot kan skriva tusentals artiklar på en dag och att marginalkostnaden för varje ny artikel snabbt närmar sig noll.

Men det mesta i livet är mindre precist än sportresultat och därmed svårare att skapa regler för.

Man kan hoppas på mallar som går att använda på flera olika datauppsättningar. Alltså samma mall för finansnyheter som för sport, till exempel.

Man kan hoppas på NLG* på svenska, vilket flera av dem jag talat med säger, inte fungerar i dag. Och även om det gjorde det, så är det möjligtvis artificiellt, men inte så intelligent.

Mölndal vann med uddamålet hemma mot Skara

ISHOCKEY HOCKEYTREAN ALLTREAN SÖDRA A HERR Tisdag 21.10, uppdaterad Onsdag 13.33

BORÅS TIDNINGSTEXTROBOT



Det blev en uddamålsseger för Mölndal i matchen mot Skara i Hockeytrean Alltrea södra A på tisdagen. Laget vann med 3-2 hemma i Åby Ishall.

I tabellen innebär det här att Mölndal nu hamnar på andra plats efter matchen. Skara ligger kvar på sjunde plats.

Mölndal tar sig an Munkedal i nästa match borta söndag 22 januari 17.00.

Nästan alla tester visar att läsare inte kan skilja en robottext från en text skriven av mänsklig hand. Men det beror mest på att texter som robotar skriver i dag är sådana som sällan har utrymme för vare sig kreativitet eller espri.

Men vad händer om man släpper AI fritt? Kan man tänka sig en mall som: "Skapa_språkligt_kul_sportrubrik"?

Output skulle, i vårt fall, kunna bli: "AIK:s GuldTjerna", "Han är AIK:s innersta Tjerna".

Eller kanske "BomBom Tjernkraftverk", en rubrik som bara sportnördar förstår. Den syftar på Tommy Söderberg, tidigare förbundskapten för svenska landslaget i fotboll, och hans beskrivning av en annan AIK:are, Johan Mjällby. För att skapa den rubriken skulle systemet dels förstå vad som hänt: AIK vann Allsvenskan och Daniel Tjernström gjorde det avgörande målet och dels ur fotbollens kultur hitta något som både passar i kontext, innehåll och form och ur detta skapa en begriplig (nåja) mening. Det är verklig intelligens och sådant som rubrikmakare gör hela tiden.

Kommer maskiner att göra det att inom överskådlig framtid?

Jag tror det. Språkets hela idé är att vara begripligt. Det innebär att det måste finnas logik och mönster. Vilket datorer är bra på. Det finns redan flera exempel på texter, musikstycken, konst och till och med filmtrailrar som skapats med hjälp av AI.

Det mesta är dock hopkok av vad som tidigare gjorts. En låt som påminner om Beatlesmusik skapas genom att AI lyssnar igenom allt som Beatles gjort och bryter ned det i beståndsdelar. Dessa används sedan för att bygga ny "Beatlesmusik".

Grattis AI, du uppfann just coverbandet!

Forskare i Indien har skapat en SarcasmBot. Ett program som svarar sarkastiskt, men begripligt, på vilken fråga som helst. Till skillnad mot AI som i exemplet med Beatlesmusik baserar sig sarkasmerna inte på tidigare förolämpningar och spydigheter. I stället skapar den helt nya sarkasmer med hjälp av matematik, lingvistisk, logik och en dos "deep learning".

Ett exempel:

Du skriver: Why did Greg go home?

Boten svarar: Because Greg was in a mood for adventure.

Du skriver: I won a lottery.

Boten svar: Oh, I am so sorry for you.

Inte jättesmart, men okej för en robot.

PS. Samtliga texter och rubriker i sporttexten, har jag hittat på. Men AIK vann verkligen guld 2009 och Tjerna avgjorde mot Blåvitt. DS.

Här är onsdagens väderprognos för Örnsköldsvik från SMHI

På morgonen kommer det att bli halvklart, 13 grader och svag ostlig vind.

På dagen får vi mulet, något regn, temperaturen stiger till 14 och vinden tilltar till måttlig sydlig.

I kväll blir det mulet, temperaturen sjunker till 12 och vinden avtar till svag sydlig.

Den här texten har skapats automatiskt med väderdata från SMHI.

Mittmedias Textrobot



LÄS
OCKSÅ:
**"När algoritmerna
sköter
journalistiken"**

**Medie
världen**
Premium

SLUTSATSER

» Att skapa en automatiserad nyhetstjänst är dyrt i början, men kostnaden lutar mot noll ju fler artiklar som skapas. Det innebär att det är mest logiskt att använda inom områden där det kan skapas många artiklar. Det är ingen tillfällighet att tekniken används på sportresultat, finansnyheter, lokaljournalistik men även sådant som produktbeskrivningar för e-handlare.

» Lär nyhetschefer och andra i redaktionell ledning tänka i robottermer. Vad går att göra? När ska man göra?

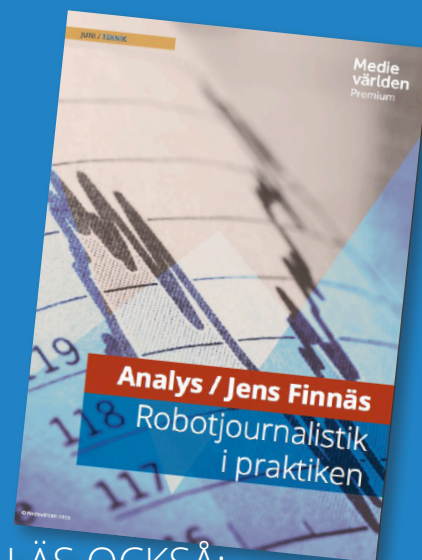
» Skapa nätverk av lingvister, datavetare, statistiker och liknande som både kan komma med idéer och hjälpa till att utföra dem.

» Tänk på datasäkerheten och integritetsfrågorna. Personaliserad kommunikation är från den 25 maj 2018 då GDPR träder i kraft ett område som kommer att vara strikt reglerat och ställa stora krav på datahållaren. Den redaktionella verksamheten på medieföretag är undantagna i GDPR. Men jag är övertygad om att alla bolag måste kunna redogöra för vilken data som samlas in och vad den används till. Användarna kommer att kräva det, oavsett lagar.

» Prioritera. Den gamla tesen att det är bättre att göra en sak bra än att göra flera halvbra.



United Robots textrobot Rosalinda förser i dag Mittmedia, Stampen, Gota Media, Mediehuset UNT och Östgöta Media, Norran i Skellefteå, småländska Hallpressen och finska HSS Media med automatiska texter om sport och fastigheter. På bilden syns företagets vd Sören Karlsson.



LÄS OCKSÅ:
"Robotjournalistik i praktiken"



LÄS OCKSÅ:
"GDPR-guide: Bli redo steg för steg"

**Medie
världen**
Premium

LÄNKAR

AI skapar tre varianter på nyheter en höger, en vänster och en opartisk

› <https://knowherenews.com/>

Brittiska Radar

› <https://medium.com/@urbsmedia/reporters-and-data-and-robots-ee352220c5f1>

Testa J++ Newsworthy

› <http://www.newsworthy.se/sv/>

En så-funkar-det med Newsworthy som exempel för den som vill veta mer

› https://docs.google.com/presentation/d/1iXHDg-h0sWjaLFC3ku6rJKD6FVFA-7naly8i_LDcof8/edit#slide=id.p

Användarcase från Amerikanska Automated Insights, veteran i sammanhanget

› <https://automatedinsights.com/case-studies>

Nvidias AI-podcast är ofta intressant. I avsnitt 48 förklarar doktor Puspapak Bhattacharrya, från IIT Bombay, artificiell sarkasms

› <https://soundcloud.com/theaipodcast/ai-sarcasm>

AI gör musik Beatlesstyle

› https://www.youtube.com/watch?time_continue=6&v=LSHZ_b05W7o

AI gör trailer för skräckfilm

› <https://www.youtube.com/watch?v=gJEzuYynaiw&t=35s>

Skaparna av sarkasmboten förklarar hur algoritmen arbetar

› <https://www.cse.iitb.ac.in/~pb/papers/wisdom15-kdd-sarcasmbot.pdf>

Om projektet som ville strukturera nyhetshändelser

› http://www.zora.uzh.ch/id/eprint/137060/1/Automated_Journalism_2.0_EventDriven_Narratives.pdf



LÄS OCKSÅ:
"A och I för medieföretag"



LÄS OCKSÅ:
"DN, SvD och DT om sina digitala arbetssätt"

Medie
världen
Premium