

AI @ Nasjonalbiblioteket

Kodenavn Nancy

Fra stille data til aktiv kunnskap

Svein Arne Brygfeld
NB AI-LAB



Fra katalogiseringsregler
til
omtrentelig og godt nok



Dagens prat

- Motivasjon, Nancy og NB
- Kort introduksjon til maskinlæring
- Eksempler fra NB AI-lab
- Konklusjoner?

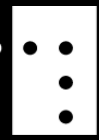


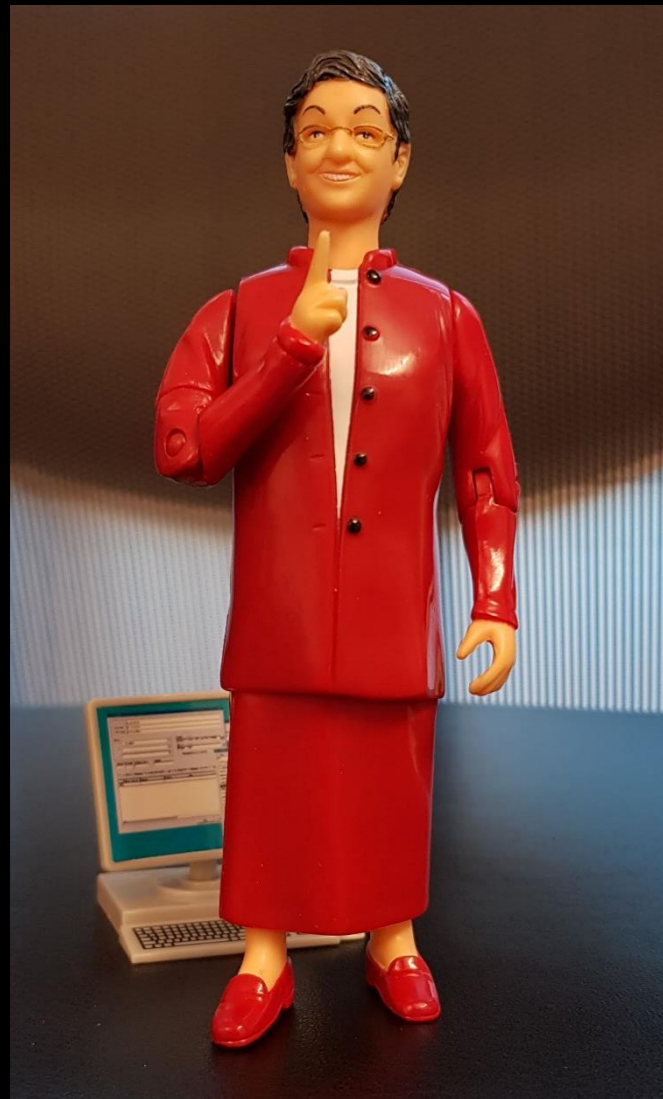
INFORMASJON

KUNNSKAP

DATA

KLOKSKAP



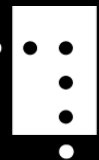


You can have all the books, magazines, periodicals, journals, videos, and CD ROMs in the world, but without a librarian you'll be frustrated and overwhelmed before you can say "Dewey Decimal System." And although most librarians can't travel faster than a speeding bullet, or leap over tall buildings in a single bound, they can direct you to an article on the physics of speed, an instructional pamphlet on high jumping, and a book of photographs of the world's tallest buildings.



Lavthengende frukt

- Åpne og tilgjengelige trenings- og testsett
 - Imagenet og andre
 - *Samlinger i digitale bibliotek (innhold og metadata)*
- Åpen kildekode, demokratisert programvare
 - TensorFlow og andre
- Kommersielle tjenester
 - API'er for tale til tekst, analyse av stillbilde og video, prosessering av naturlig språk og



NB går digitalt – 2005/2006

Kort oppsummert

- Digitalisere hele samlingen
- Gjøre samlingen tilgjengelig der brukeren er, når brukeren vil

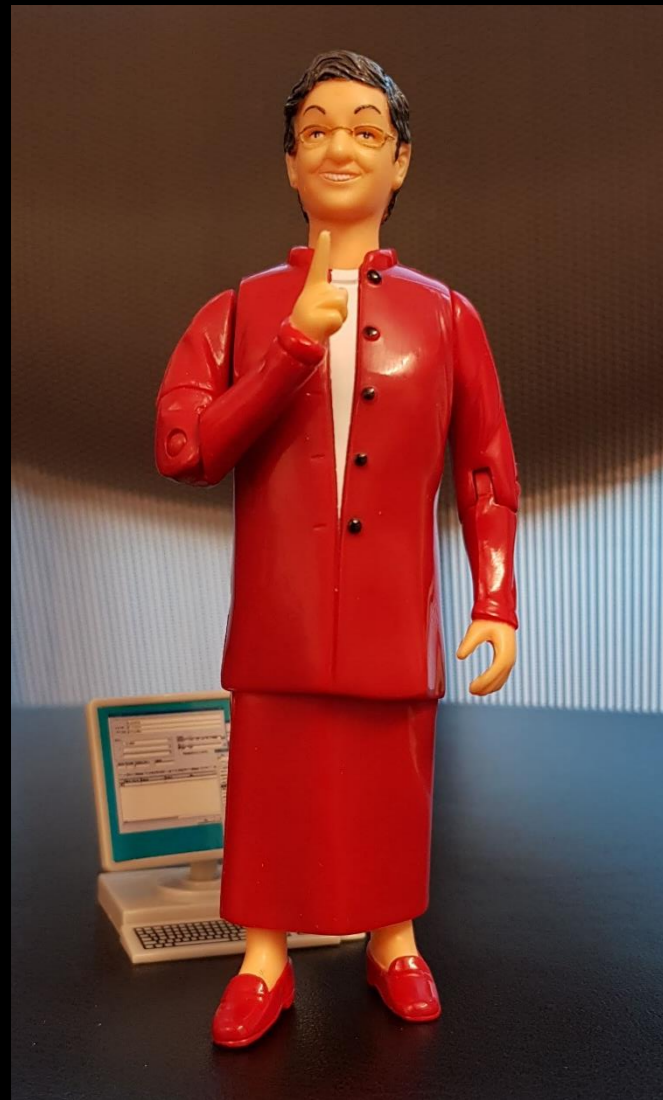


NB går digitalt, i dag

- Massiv digital samling, som f eks
 - Tekst: bøker, tidsskrift, aviser
 - Audio: radio, musikk, dokumentaropptak
 - Visuelt: fotografi, film, TV
- Stor bruk på nett



Nancy



- Paraply over innovasjon på AI i NB
- Definert aktivitet, NB ALAB, rapporterer direkte til nasjonalbibliotekaren
- Bidrag til samtalen om AI gjennom småskala prosjekter og piloter
- 3-4 personer på full tid, noe tjenestekjøp



Kunsting intelligens?

Maskinlæring?



ML: fra presist til omtrent

- Er
 - Presist, forutsigbart
 - IF-THEN-ELSE
 - Riktig eller feil
-
- Blir
 - Omtrentelig ytelse
 - Erfaringsbasert læring
 - Godt nok

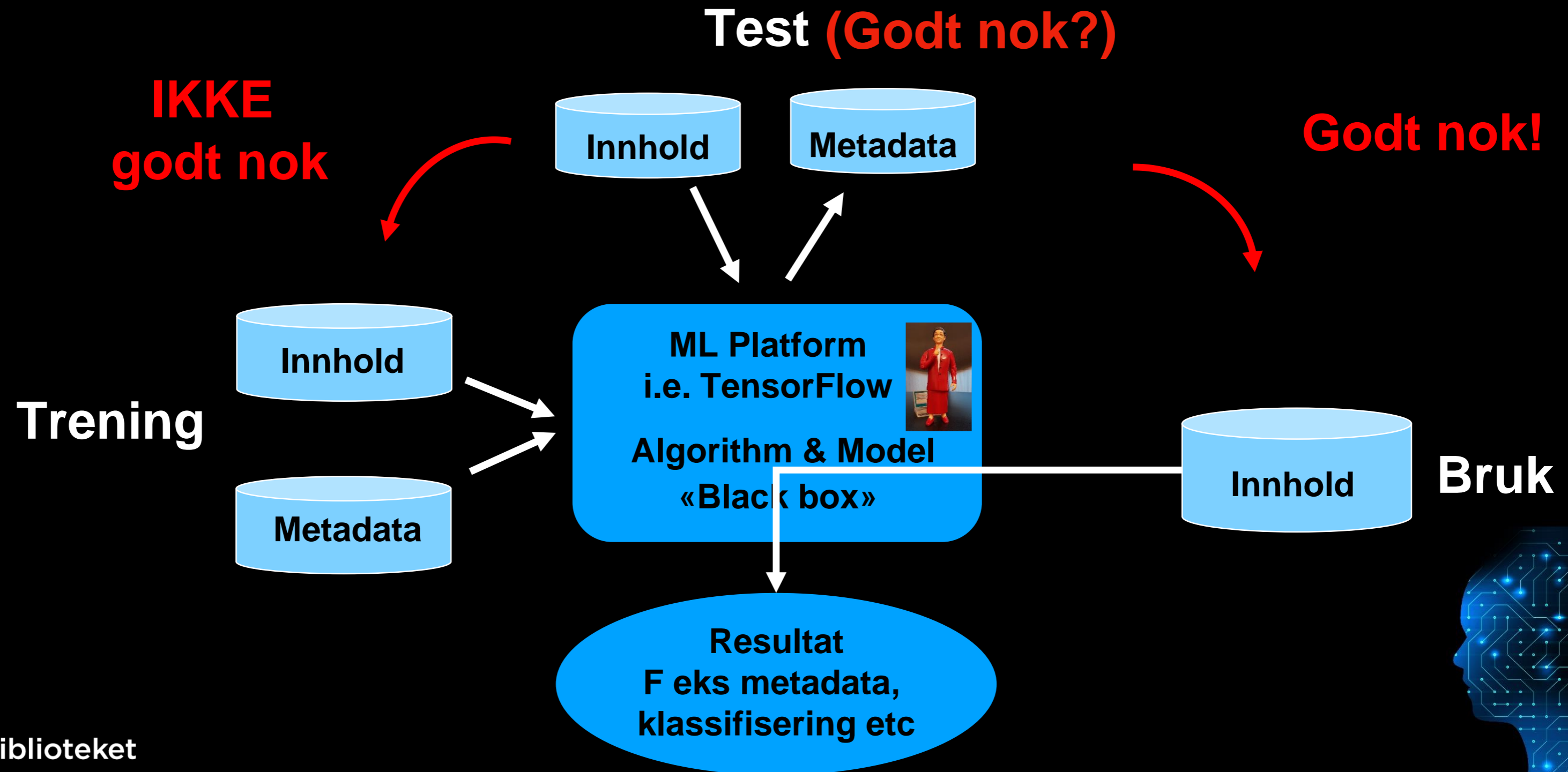


Maskinlæring, grovt sett

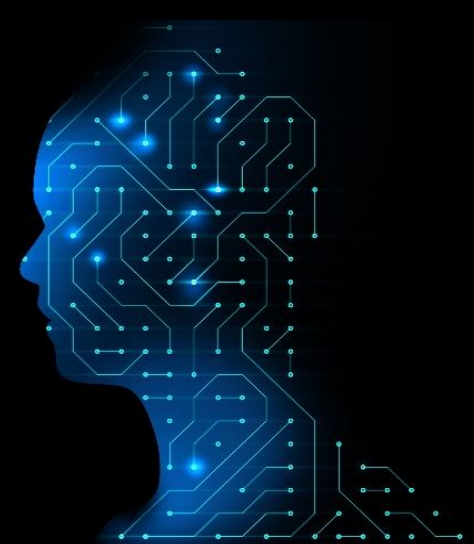
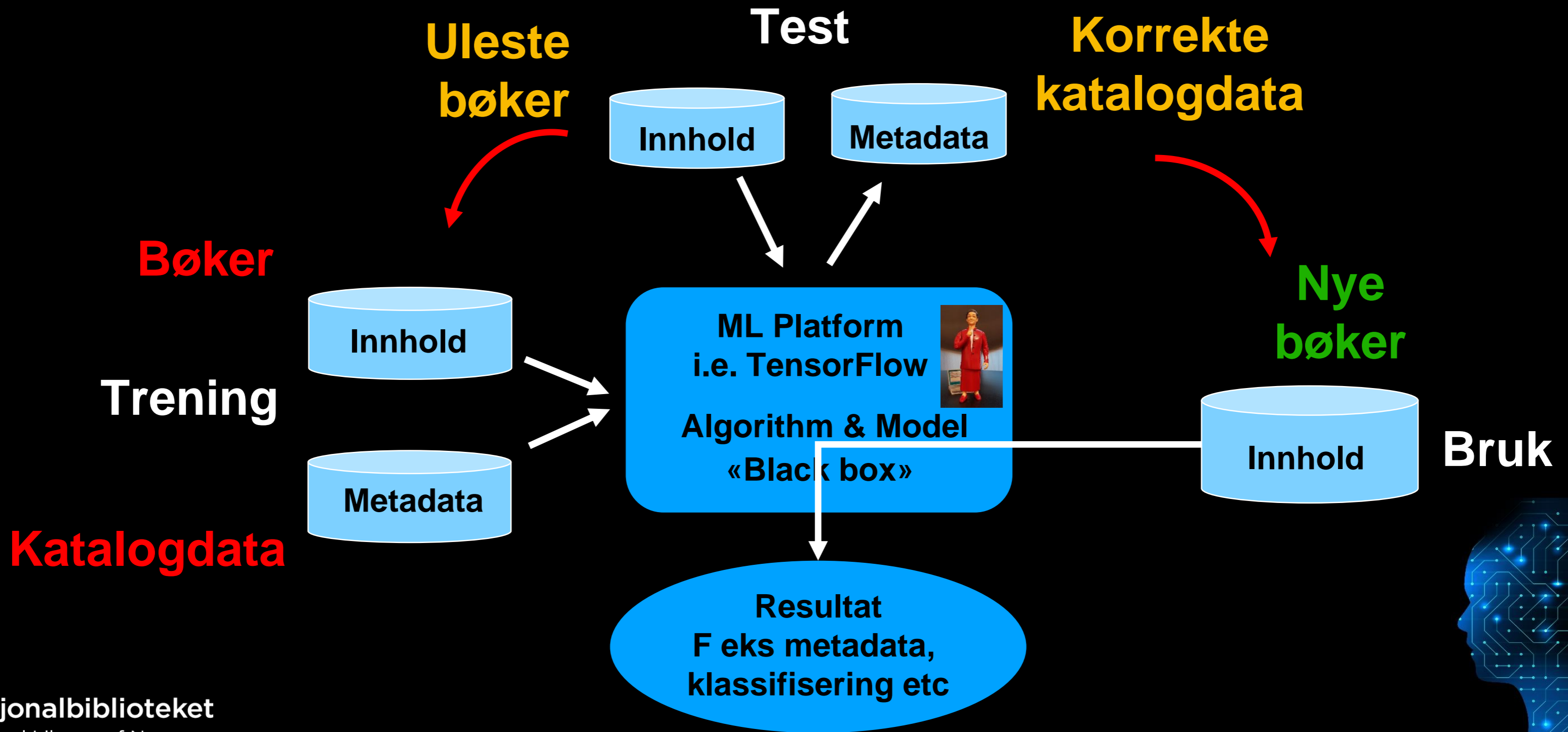
- Unsupervised learning
- Supervised learning
- Reinforced learning



Supervised: tren, test, bruk!



Supervised: tren, test, bruk!



Eksempler fra NB AI-lab



Enkel klassifikasjon av bøker

- Litteraturgrupper som eksempel
 - 4 grupper definert av KOPINOR for fordeling av midler
 - Ei bok tilhører bare ei gruppe
 - Sparsomt med definisjon av gruppene



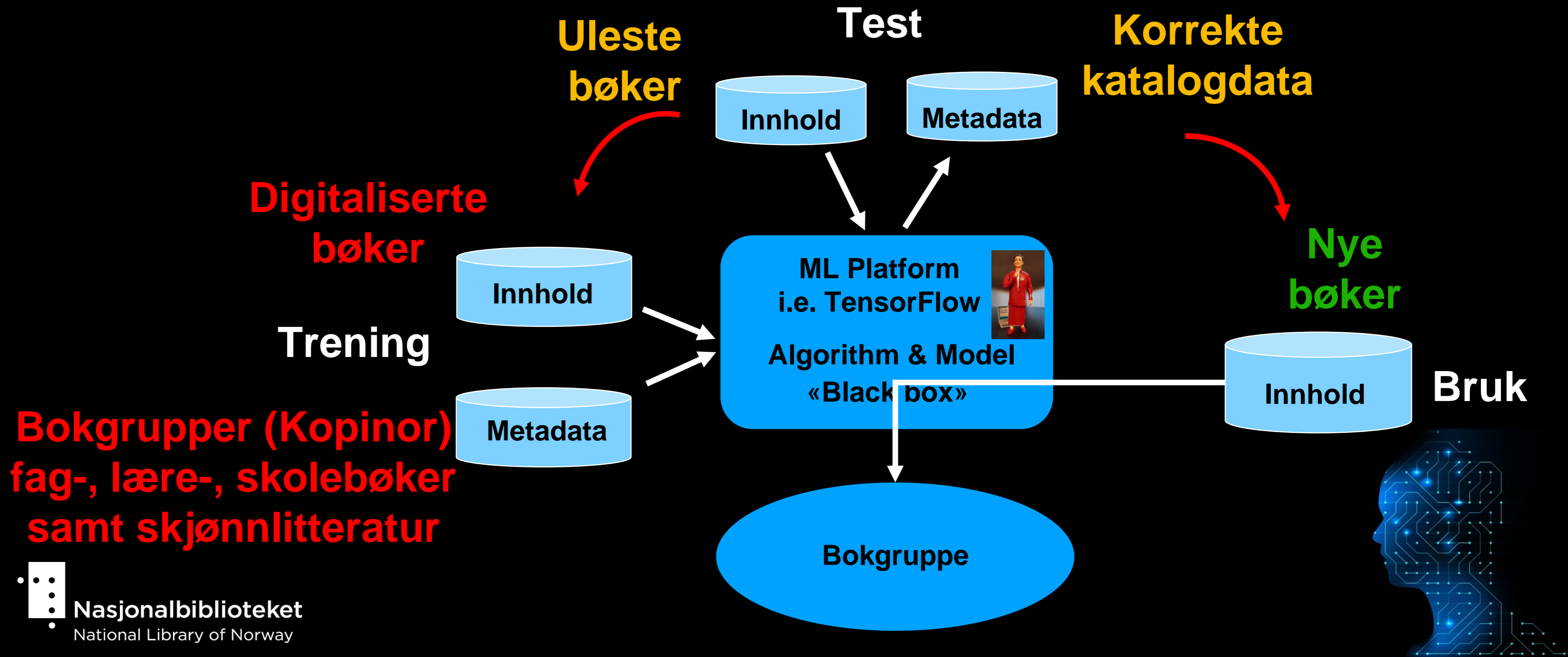
Klikk på for å se Nancys forlag til alternativ lesing

The image shows a grid of approximately 100 book spines from various publishers. Colored arrows point from specific books to four circular labels: SB (blue), F (blue), SL (blue), and L (red). The text 'Litteraturgrupper' is located at the bottom right of the grid.

Litteraturgrupper

- SB** (blue circle): Points to a book titled 'NORSK ORDBOG'.
- F** (blue circle): Points to a book titled 'BREVET TIL KAROLINE'.
- SL** (blue circle): Points to a book titled 'WEGELAND, BREVE'.
- L** (red circle): Points to a book titled 'SOSIOLOGI AV VILHELM A. BERT'.

Supervised: tren, test, bruk!

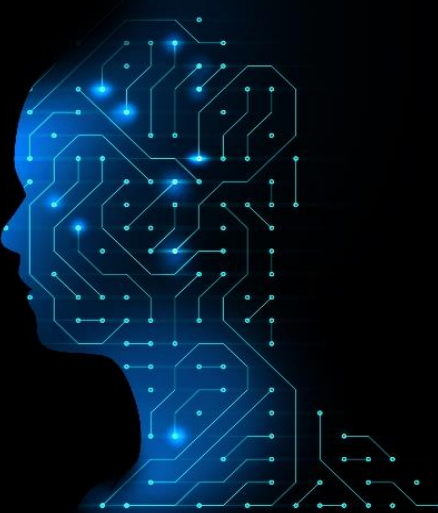
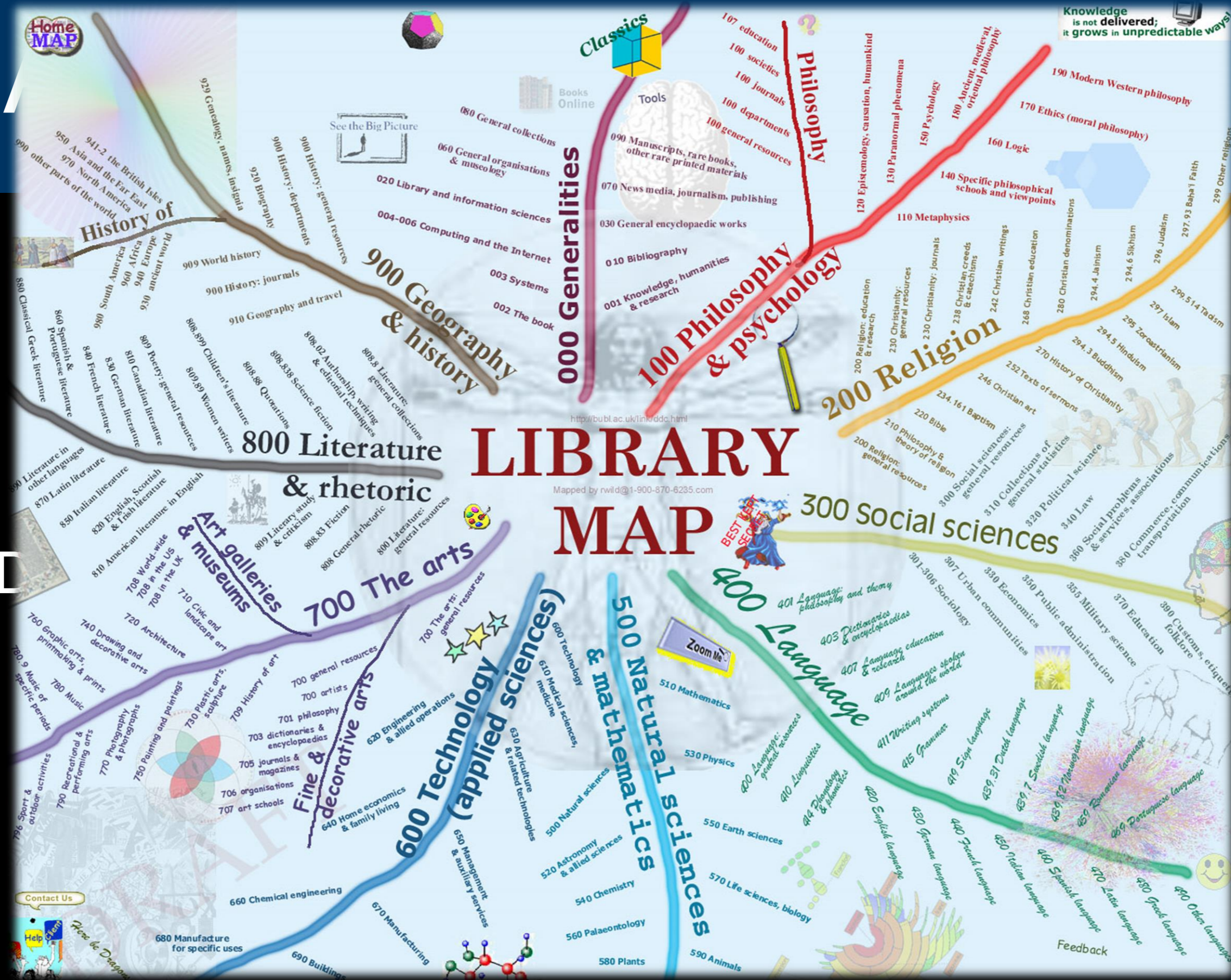


Godt nok?

> 95%



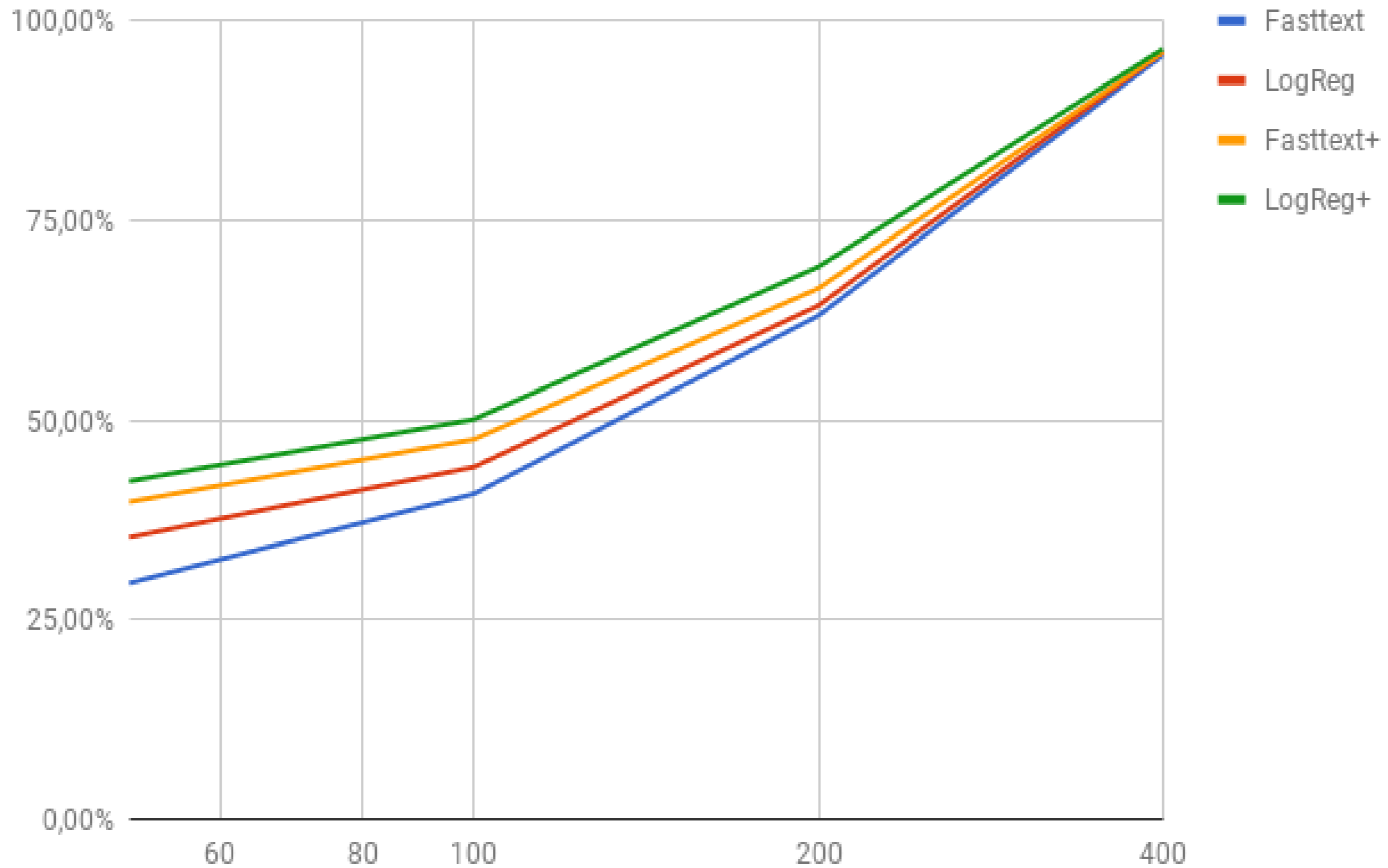
• Dewey D



De

n

Precision vs size of training set



and



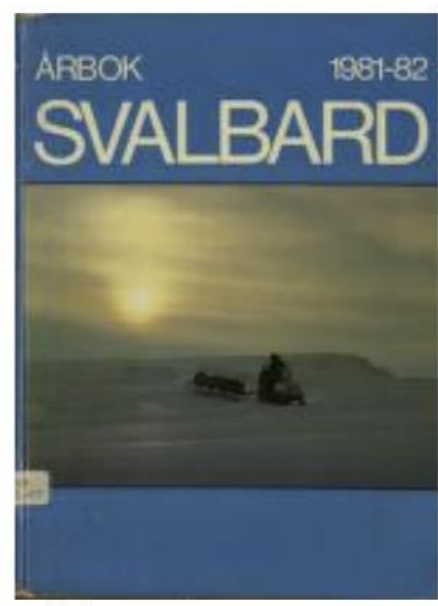
AI for bibliotek?

Lignende objekter



Boka det er tatt utgangspunkt i er *Norsk skoleflora* av *Sørensen, H.L.*, utgitt 1961

P	Tittel	Forfatter	Utgitt
---	--------	-----------	--------



Tittel/forfatter/år
Tittel: Årbok Svalbard 1981-82
Forfatter: Amundsen, Birger
 1981

[Vis alternativer](#)



Q	Norsk fargeflora	Wischmann, Finn	2012
E	Ville blomster i skog og mark	Jonsson, Rolf	1995
-	Trær og busker i farger	Austin, Toralf	1960
C	Norsk flora	Nordhagen, Rolf	1940
E	Norske fjellplanter	Resvoll, Thekla R.	1934
Q	Ville blomster i skog og mark	Jonsson, Rolf	c19761976
C	[Våre ville blomsterplanter]	Berg, Gunnar A.	1980
N	Trips Guilds farger	Montgomery, David	10061006


Kommersielle APIer

Tjenester basert på maskinlæring

- Alternativ til å trene modeller sjøl
- Typiske tjenester: Bilde- og videoanalyse, tale til tekst, språkanalyse
- Generalisert formål
- Flere leverandører, som f eks Amazon, Clarify and Google



Faces Labels Web Properties Safe Search JSON

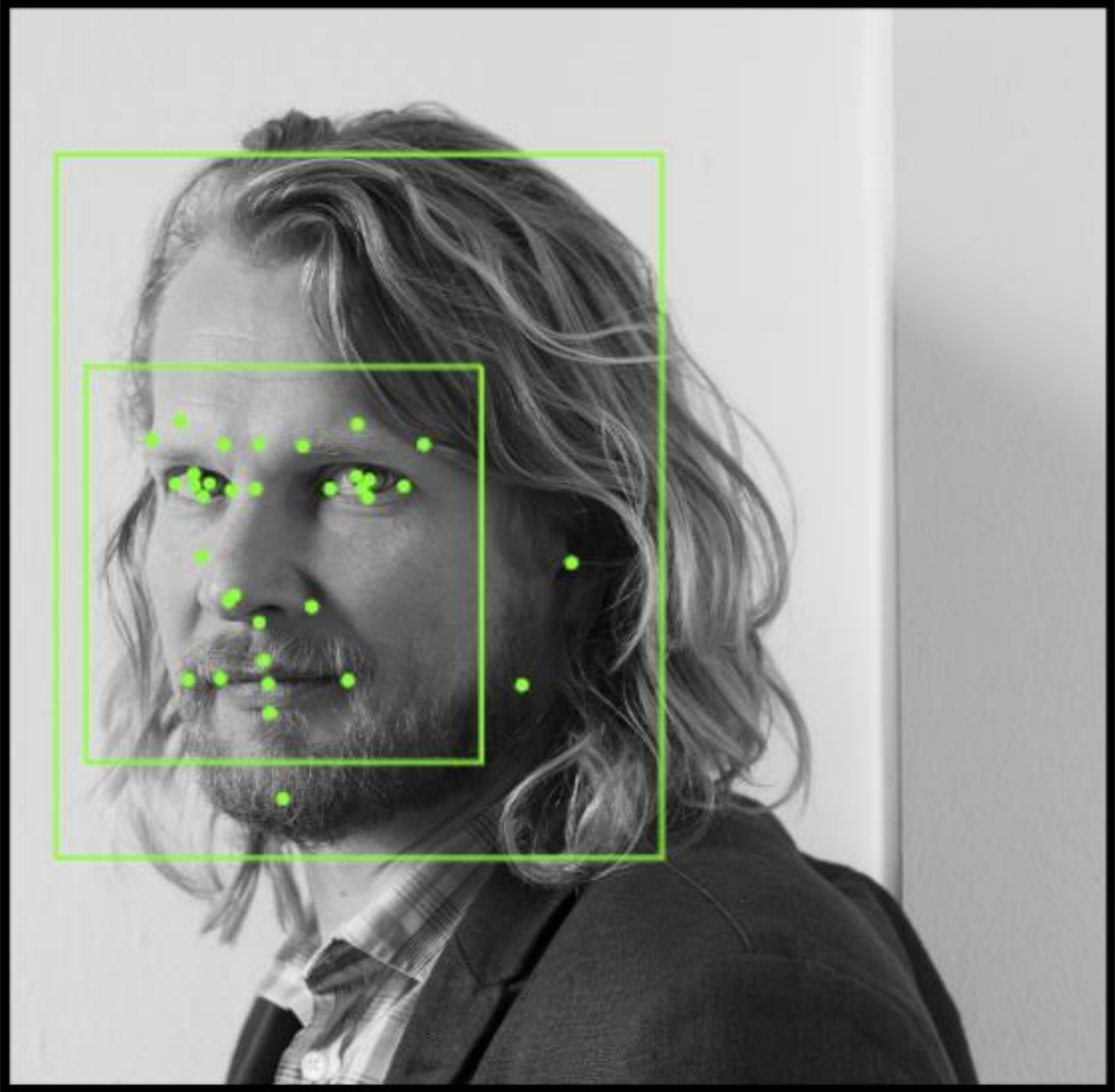


aslak1.jpg

Hair	98%
Person	95%
Facial Hair	94%
Man	94%
Black And White	90%
Chin	89%
Portrait	87%
Beard	85%
Photography	82%

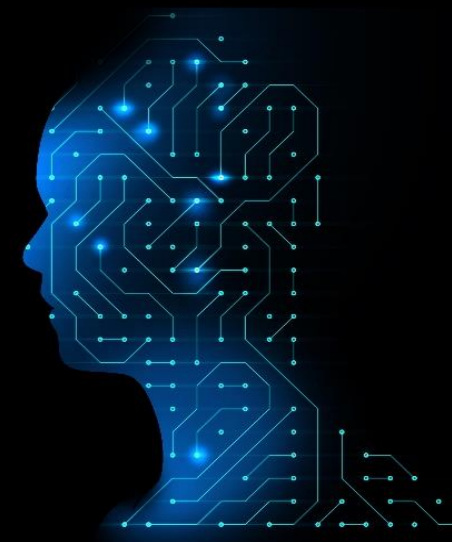


Faces
Labels
Web
Properties
Safe Search
JSON




aslak1.jpg

Joy	■	Very Unlikely
Sorrow	■	Very Unlikely
Anger	■	Very Unlikely
Surprise	■	Very Unlikely
Exposed	■	Very Unlikely
Blurred	■	Very Unlikely
Headwear	■	Very Unlikely
Roll: -1° Tilt: -4° Pan: -28°		
Confidence	71%	



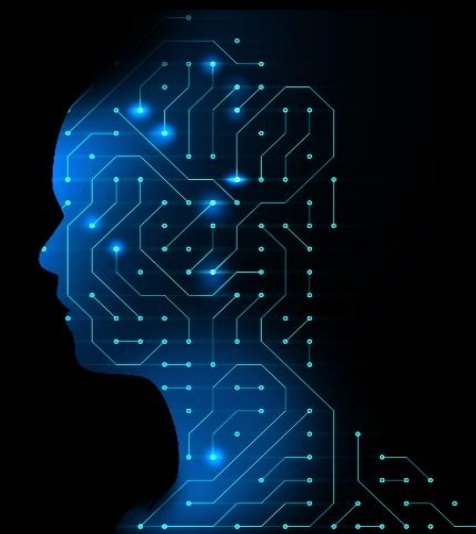
[Faces](#)
[Labels](#)
[Web](#)
[Properties](#)
[Safe Search](#)
[JSON](#)




aslak1.jpg

Web Entities

Aslak Sira Myhre	19.988
Stavanger	8.8
National Library of Norway	3.3128
De Som Bygger Landet	0.7478
Politician	0.6007
Library	0.5691
Red Electoral Alliance	0.539
Bergen Rød Valgallianse	0.4335
Director	0.4193
Male	0.3641
Party leader	0.3372
Author	0.3343
Librarian	0.3328
Erling Sandmo	0.0741
Norway	0.00027

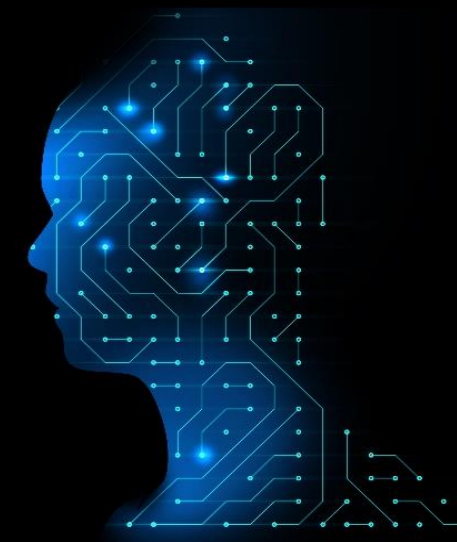


Faces Labels Web Properties Safe Search JSON

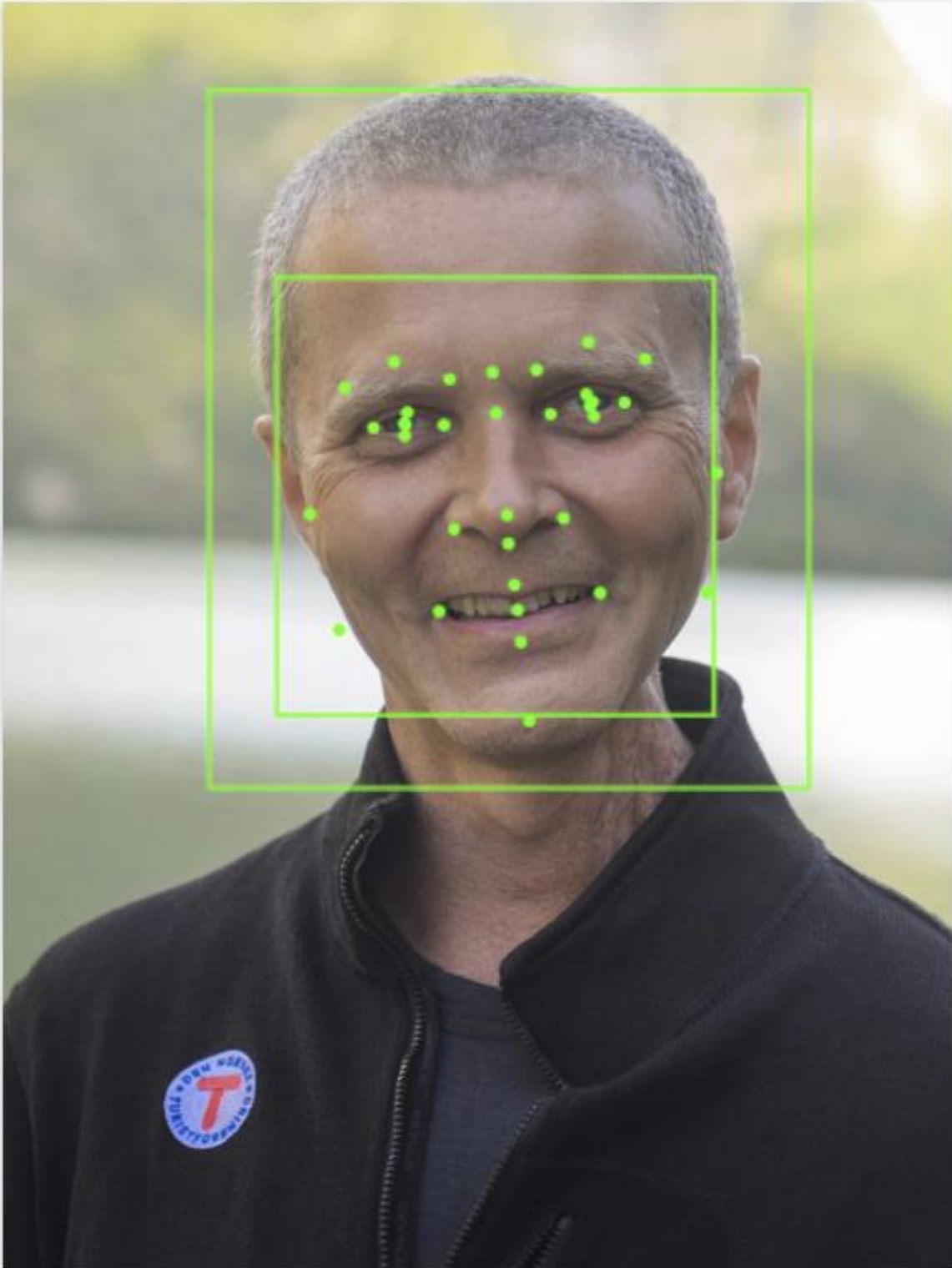


sab.jpeg

Person	96%
Chin	84%
Forehead	78%
Smile	74%
Senior Citizen	69%
Professional	64%
Official	55%
Elder	54%
Vehicle	51%



Faces
Labels
Web
Properties
Safe Search
JSON



sab.jpeg


Joy	■ ■ ■ ■ ■	Very Likely
Sorrow	■	Very Unlikely
Anger	■	Very Unlikely
Surprise	■	Very Unlikely
Exposed	■	Very Unlikely
Blurred	■	Very Unlikely
Headwear	■	Very Unlikely

Roll: -6° Tilt: -2° Pan: -2°

Confidence 100%



[Faces](#)
[Labels](#)
[Web](#)
[Properties](#)
[Safe Search](#)
[JSON](#)



sab.jpeg


Web Entities

James Bond	2.8092
Daniel Craig	1.9986
Sarajevo Film Festival	1.6404
Skyfall	0.8874
Actor	0.7008
Image	0.6226
James Bond Film Series	0.5435
Film	0.4644
Photograph	0.4271
Sienna Miller	0.10988
Judi Dench	0.10196
Daniel Radcliffe	0.09536
The Girl with the Dragon Tattoo	0.08454
Pierce Brosnan	0.08344
Sam Mendes	0.0814

Pages with Matched Images



Labels Web Text Document Properties Safe Search JSON




trikk2.jpeg

Car	98%
Motor Vehicle	98%
Transport	96%
Vehicle	95%
Black And White	88%
Mode Of Transport	88%
Vintage Car	83%
Snapshot	82%
Automotive Design	78%



Labels Web Text **Document** Properties Safe Search JSON



trikk2.jpeg

+Page 1

+Block 1

TORGET

+Block 2

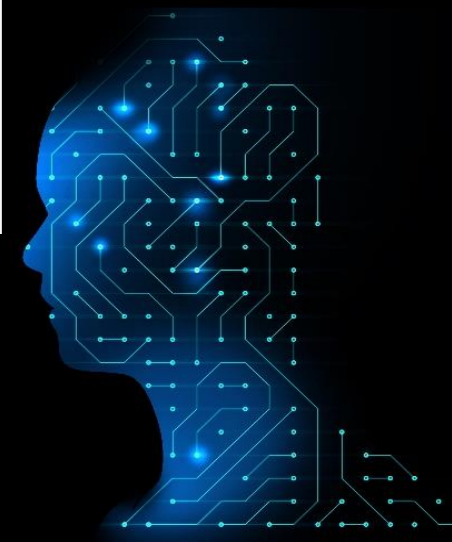
VAL HAGEBY

+Block 3

TORGET

+Block 4

ULLEVÅL HAGEBY



Gerhardsen - Tranmæl

Transcript: orientert og posisjon under ledelse
først og fremst av **Martin tranmæl bondegutten fra
Trøndelag** og Dette var jo da under og mot
avslutning den første verdenskrig etter en russisk
revolusjon og det var en situasjon som appellerte
start leiligheter i unge og de motsetningene som
etter hvert da kommer til å gjøre seg gjeldende i
norsk arbeiderbevegelse de merkede Josefa
bevegelsen One Wheel

Confidence: 0.7010927200317383



Word: **først**, start_time: 843.8, end_time: 844.3
Word: **og**, start_time: 844.3, end_time: 844.3
Word: **fremst**, start_time: 844.3, end_time: 844.7
Word: **av**, start_time: 844.7, end_time: 844.8
Word: **Martin**, start_time: 844.8, end_time: 845.3
Word: **tranmæl**, start_time: 845.3, end_time: 846.1
Word: **bondegutten**, start_time: 846.1, end_time: 847.9
Word: **fra**, start_time: 847.9, end_time: 848.1
Word: **Trøndelag**, start_time: 848.1, end_time: 848.9
Word: **og**, start_time: 848.9, end_time: 849.0



Komplett bibliotek

Hele puslespillet lagt



Nancys største utfordring

- Kan vi lage et digitalt bibliotek basert på bare AI?
- Eksperimentet
 - Komplett innhold fra januar 2011
 - 250 aviser (3.000 utgaver, 4.400.000 artikler, 209.000 foto)
 - To radiokanaler (742 timer audio)
 - En TV-kanal (100 timer video)



Produksjon av metadata basert på maskinlæring

Avis



TV



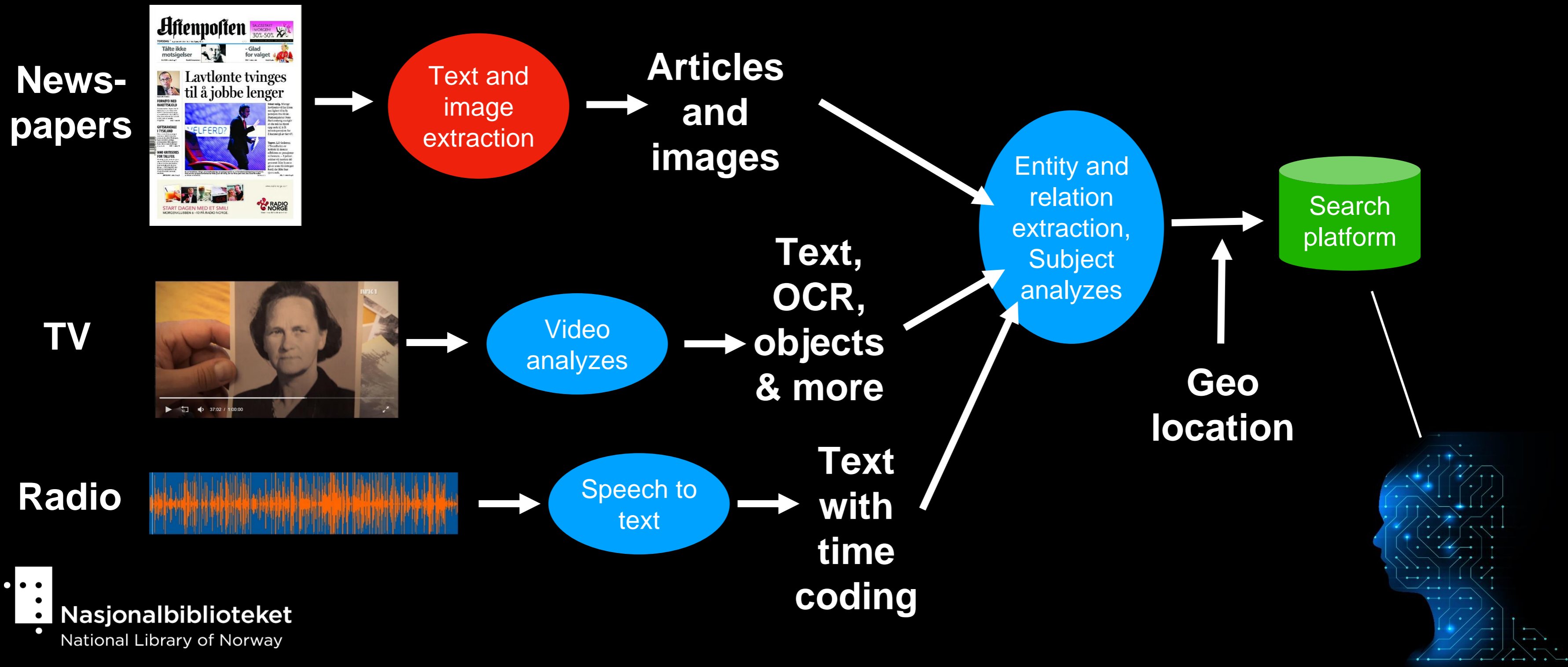
Radio



- Personer (navn)
- Steder
- Organisasjoner
- Tid
- Relasjoner
- Klassifikasjon av innhold



Designprinsipp



Nancy

Ikke sikker

Top hits based on selection

DATE DISTRIBUTION

Saturday 1st January 2011

PEOPLE

- tlf
- ellen degeneres
- seiv
- chris
- amy
- cæsar

LOCATIONS

- norge
- oslo
- sverige
- bergen
- usa
- stavanger

Nancy

Ikke sikker | nbr-freddy.nb.no:7873

Top hits based on current map selection

DATE DISTRIBUTION

Saturday 1st January 2011

Monday 31st January 2011

PEOPLE

- raymond
- hellstrøm
- seiv
- tlf
- ellen degeneres
- chris

LOCATIONS

- norge
- oslo
- usa
- sverige
- bergen
- asker

Search all content by map selection

Map showing search results (pins) overlaid on a geographical map of Norway. The map includes labels for 'Asker', 'Nesoddtangen', 'Bunnefjorden', 'Ski', and 'Gaupsteinmarka naturreservat'. A search bar at the top left of the map area contains a checkmark and the text 'Search all content by map selection'.



Nancy

Ikke sikker | nbr-freddy.nb.no:7873/people

jens stoltenberg

LOCATIONS

- norge
- oslo
- sverige
- russland
- europa
- lofoten

ORGANIZATIONS

- ap
- sv
- stortinget
- ntb
- arbeiderpartiet
- høyre

PEOPLE

- maria amelie
- kristin halvorsen

Vindsklipp

Nyheter

Mandag 24. januar 2011 | Adresseavisen

Sjefredaktør: Arne Blix
 Nyhetsredaktør: Kirsti Husby Samfunnsredaktør: Stein Arne Sæther
 Digital redaktør: Rolf Dyrnes Svendsen
 Adm. direktør: Tove Nedreberg

MIDT-NORGES FRIE STEMME

Politikere på sidelinjen

Politikerne har gitt fra seg styringen over stadig flere samfunnsområder til byråkratiet.

Politikere som overlater viktige eller ubehagelige beslutninger til ulike forvaltningsorgan, gir fra seg mye av sin makt som folkevalgt. Økende byråkratisering er i ferd med å bli et demokratisk problem.

Et resultat av at politisk makt svekkes, er rettsliggjøring, et begrep fra Maktutredningen av 2003. Politikerne vedtar retningslinjer for forvaltningsorganene og overlater konkrete avgjørelser til dem. I konfliktsaker der regler og avgjørelser oppfattes som urettmessige, kan dette bli problematisk. Ofte blir slike saker avgjort av domstolene. Politisk makt blir erstattet av rettsmakt, men også byråkratmakt.

Maria Amelie -saken er en illustrasjon på hva som skjer når politikere abdiserer. For helgen bergset regjeringen seg ut av en håplos skvis ved en unnamanoever. Den «spapirlose» kvinnen må ut av landet i tråd med Utlendingsnemndas (UNE) vedtak. Men hun kan komme tilbake ved å søke om arbeidstillatelse og slippe karantenetiden på inntil fem år.

«Lex Amelie» er en oppmykning av regelverket for arbeidsinnvandring, men ikke av en restriktiv asylpolitikk. Statsminister Jens Stoltenberg har ikke rikkert seg på sitt prinsipp om at Amelie må

behandles på lik linje med andre ulovlige asylsøkere.

Statsministerens hovedargument i Amelie-saken var at regjeringen ikke kan instruere utlendingsmyndighetene. Det er en rettighet politikere selv har gitt fra seg. UNE ble opprettet for at departementet ikke skulle være klageinstans for vedtak fattet av Utlendingsdirektoratet. UNE fungerer i dag som domstolene, hvor politikere ikke kan overprøve avgjørelsene. Spørsmålet er om det er riktig å gi UNE så stor myndighet.

Byråkratiseringen av politisk makt er en foruroligende utvikling. Det ble lagt frem en rapport som viser at statsrådene har gitt fra seg retten til å styre i flere forvaltningsorgan. I partiet av 169 underliggende organer har fått formell eller uformell fristilt makt.

Statsråder som mener at de ikke kan gripe inn overfor forvaltningen i enkeltsaker, snakker seg ut med de grunnleggende prinsippene for styring her i landet. Det gjelder sammenhengende ansvarslinje mellom Stortinget og regjering til de ulike forvaltningsorgan. Statsrådene har det overordnede ansvar for det som skjer innenfor deres fagområder.

Vi ønsker ikke politikere som driver detaljstyring. Men det må ikke føre til at de fraskriver seg ansvar og overfører politisk makt til ansiktsløse byråkrater.

Datalagring

Sosialistisk Venstreparti og Fremskrittspartiet har funnet hverandre i kampen mot EUs datalagringsdirektiv, melder NTB. Deres felles fiende er Høyre. I enkelte saker er det politikernes ytterpunkter som har fornuften i behold.

Nyheter

Tips oss - telefon 07200
 nyhet@adresseavisen.no • MMS/SMS: Kodeord TIPS til 07200

Serievoldtektsmann uten behandling:

Tatt for

Mannen (31) skapte voldtektsfrykt i Trondheim og ble dømt til forvaring etter flere overfall og en voldtekt. I flere måneder har han vært på fri fot uten behandling.

«Lex Amelie» er en oppmykning av regelverket for arbeidsinnvandring, men ikke av en restriktiv asylpolitikk. Statsminister Jens Stoltenberg har ikke rikkert seg på sitt prinsipp om at Amelie må

31-åringen nektet torsdag for anklagene om onani og blotting. Han tok betenkningstid på fengslingsavgjørelsen.

– Han har forklart at han har vært inne og sett etter dvd-filmer, og stått utenfor kiosken. Fornærmede mener å ha observert onani, mens han selv har forklart at han holdt på å knytte buksene.

Samland mener politiet virker «overivrige» på å fengse mannen igjen.

– Han har fått veldig bra skussmål etter at han kom ut. Han har tatt ansvar for sin fortid og gjort opp for det han har gjort, sier Samland.

Hver gang mannen har begått

Det gjør også leder Lars Høines i Friomsorgen Sør-Trøndelag. Han sier de mangler penger til å finansiere behandlingsopplegg for forvaringsdømte.

– Jeg tror at både justiskomiteen og helsekomiteen på Stortinget burde interessere seg for hva som skjer med denne gruppen når de kommer ut fra forvaringsanstalt og avbryter et påbegynt behandlingsløp. De bør se på det nå, fordi det er nå denne gruppen kommer ut, sier Høines.

PÅLE SOLBERG 986 20 134
 paal.solberg@adresseavisen.no

SEBA BAJOGHLI 900 47 257
 seba.bajoghli@adresseavisen.no

– Hvis påstandene er sanne bekrefter dette at systemet har sviktet.

MARIUS RYVENGEN, mannens fastlege

1.01.2011 - 31.01.2011 SHOW MAP

jens stoltenberg + maria amelie

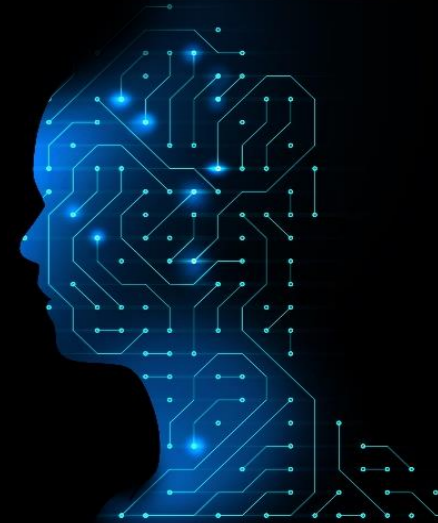
... was mentioned in helgelandarbeiderbla avisbilde on Monday 17th January 2011

jens stoltenberg + maria amelie

... was mentioned in adresseavisen avis on Monday 24th January 2011

nyhet norge statsforvaltning politikk offentlig

Les avisen



Hva kan bibliotek gjøre for AI?

- Åpne datasett for trening og test, forskningsplattform
 - Digitalt innhold med høykvalitets metadata
 - Tilgang til forsknings-/utviklingsmiljø internt i biblioteket som alternativ
- Domenekunnskap
 - Vurdere/måle ytelsen i modeller
- Ferdigtreneede modeller
 - Plug-in modeller for ulike typer innhold



Konklusjoner?

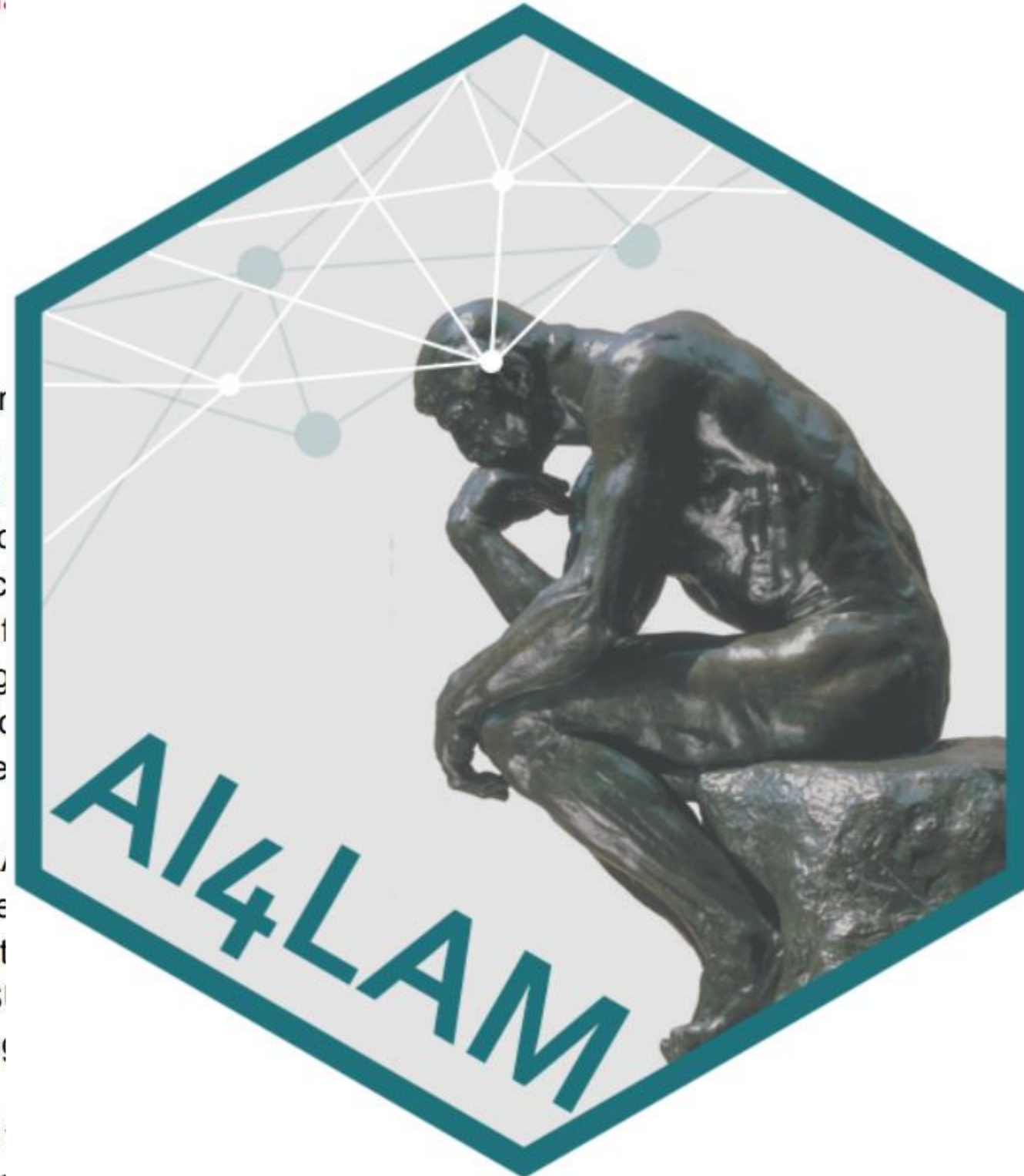
- Vi forstår at vi ikke helt forstår, men at vi bør forstå
- Intern arbeidsflyt kommer til å endres
- Vårt tradisjonelle syn på metadata vil endres betydelig
- Samlingene vil bli bedre tilgjengelig
- Biblioteket vil bidra bedre til kunnskap og forståelse
- Vi blir bedre bibliotek – nye verktøy for tilgang til og produksjon av kunnskap



Stortingsmelding 8: Kulturens kraft

«Nasjonalbiblioteket ønsker å byggje eit miljø som kan vere ein ressurs for fleire innanfor arkiv-, bibliotek- og museumssektoren knytt til maskinlæring og kunstig intelligens.»





Libraries, as
knowledge,
information
drive cars to
will be affect
application of
of knowledge
of informati
they are the

However, L
what has be
happen, but
Libraries (S
emphasizin

Inspired by

foundation as well as the community around it, INLIL & SOL will, over the next year,

we
ow we
s (LAMs)
ts
monwealth
ollections
vertise;
umankind.

services in
s not
niversity
AI, thus
vledge.

cal

- Internasjonalt nettverk
- ai4lam, Google Group

Husk

NLN AND STANFORD LIBRARIES JOINT

FANTASTIC FUTURES 2. AI CONFERENCE

DEC 4-5, STANFORD UNIVERSITY PALO ALTO



Takk

Svein Arne Brygfjeld

svein.arne.brygfjeld@nb.no

