KiNS 2024, Tromsø

KI og cyber

https://www.mn.uio.no/ifi/english/people/aca/josang/docs/gooey-russisk-5.mp4

Audun Jøsang
Universitetet i Oslo

Om meg

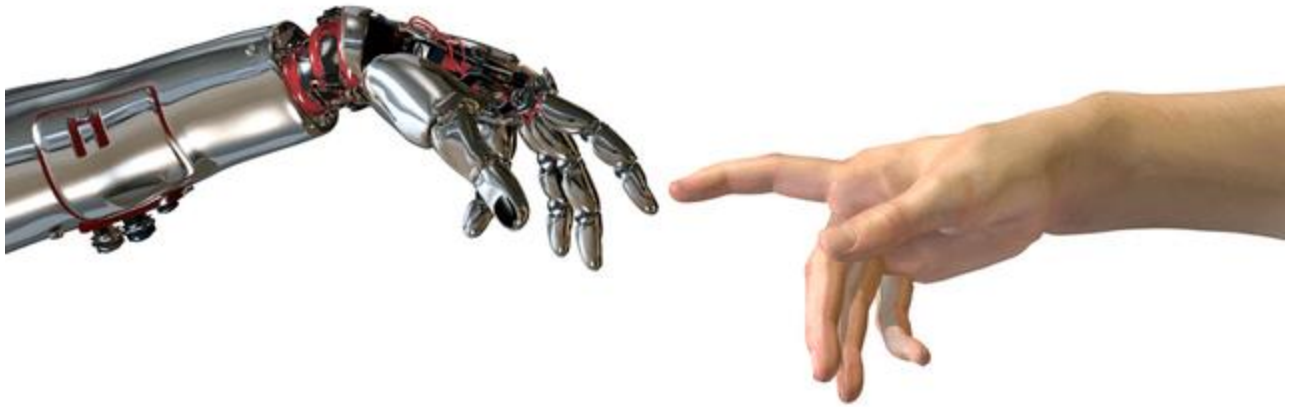
- Prof. Audun Jøsang, UiO
- Jobb
  - UiO, 2008 →
  - QUT, Australia, 2000 – 2007
  - Telenor FoU, 1998 –1999
  - Alcatel, Belgia 1988 –1992

- Lerebok om cybersikkerhet
  - 2. utgave 2023
  - Kahoot

- EVU-kurs: Ledelse av informasjonssikkerhet
  - https://www.uio.no/studier/emner/matnat/ifi/ITEVU4230/
  - Kursstart høst 2024
  - Fagbok av Dinh Uy Tran

Oversikt

- Hva er KI ?
- Offensiv KI
- Defensiv KI
- Sårbarheter og angrep mot KI
- KI-regulering
- Kahoot



Hva er KI ?

- AI-workshop in Dartmouth, Massachusetts, USA 1956

- AI-vintre, mislykkede tilnærminger
  - Logisk resonnering
  - Ekspertsystemer

- Utviklingen fikk opp farten fra ca. 2012
  - Kunstige nevralnett

- Utviklingen eksploderte med ChatGPT i 2022

Hva er KI?

KI
Kunstig intelligens

- Computerapplikasjon som utføre oppgaver som normalt (tradisjonelt) krever menneskelig intelligens

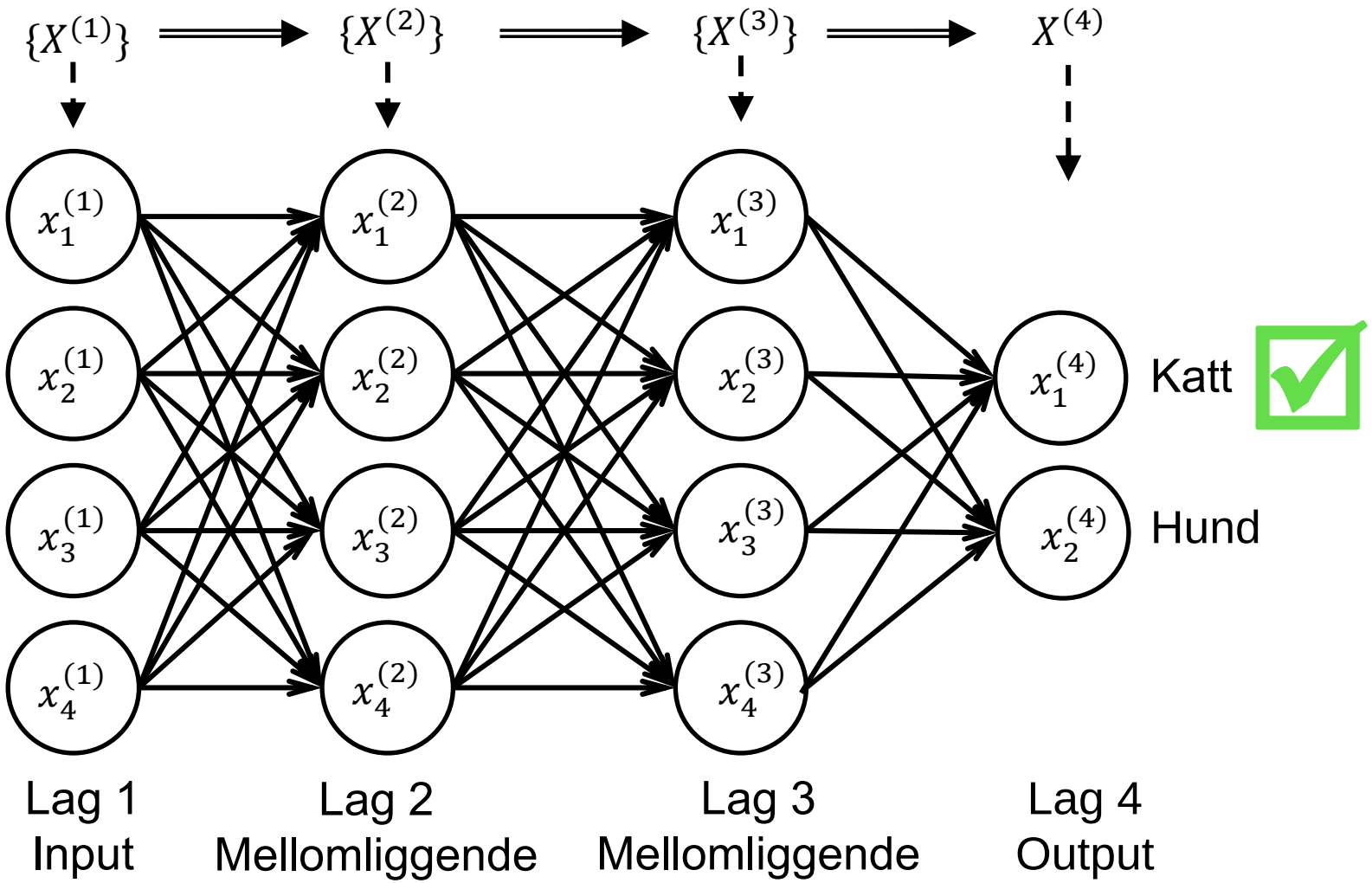
ML
Maskinlæring

- Computerapplikasjon som automatisk lærer fra datasett uten å bli eksplisitt programmert

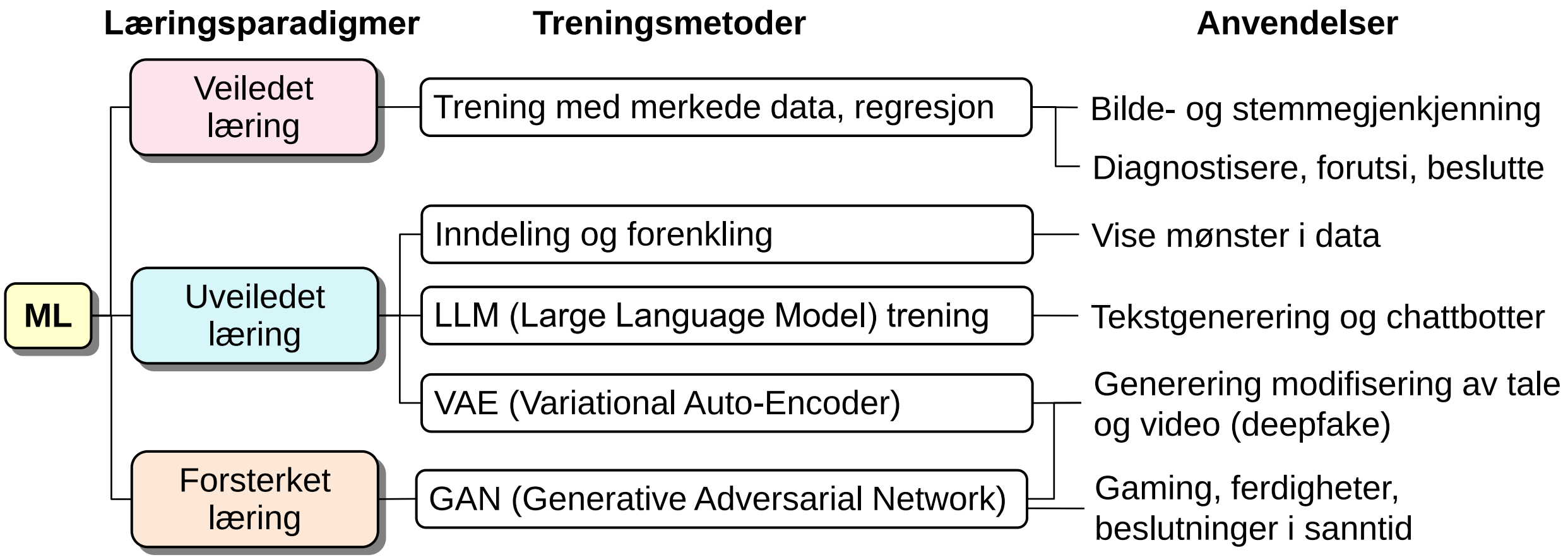
DL
Dyp læring

- ML-modeller basert på kunstige nevralnett med minst to mellomliggende lag, og som lærer fra store datasett

Kunstige nevralnett (eng. ANN)



Maskinlæring - paradigmer, metoder og anvendelser



KI og cyber

Offensiv KI

- Deepfakes
- Produksjon av skadevare
- Angrepsautomatisering

Defensiv KI

- Inntrengningsdeteksjon
- Analyse av skadevare
- Digital trusseletterretning
- Hendelsesrespons

Sårbarheter og angrep mot KI

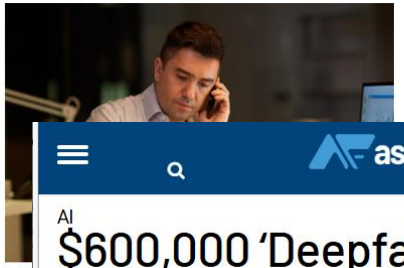
- Forgiftning av læringsdata
- Forurenset læring
- Leveransekjederisikoer
- Synsbedrag
- Lekkasje
- Injeksjonsangrep

TREND | Business

Unusual CEO Fraud via Deepfake Audio Steals US$243,000 From UK Company

September 05, 2019

An unusual case of CEO fraud used a deepfake audio, an artificial intelligence (AI)-generated audio, and was reported to have conned US$243,000 from a U.K.-based energy company. According to a report from the Wall Street Journal, in March, the fraudsters used a voice-generating AI software e[illegible]xecutive of the company's Germany facilitate an illegal fund transfer.



InSight Crime

Menu

News | Cybercrime

Fraud Groups Use Deepfakes to Enhance Imitation Scams in Peru

by *Gavin Voss*
21 Jul 2023

CNN Business | Watch | Listen | Live TV | Sign in

British engineering giant Arup revealed as $25 million deepfake scam victim

By Kathleen Magramo, CNN
3 minute read · Updated 4:53 AM EDT, Fri May 17, 2024

asia financial | NEWSLETTER

AI

$600,000 'Deepfake' Fraud Heats Up AI Debate in China

May 22, 2023

"He chatted with me via video call, and I also confirmed his face and voice in the video. That's why we let our guard down," the victim said

Gadgets 360
An NDTV venture

HOME | GUIDE | AUTO | NEWS | REVIEWS | FEATURE
MOBILES | TELECOM | HOW TO | GAMING | ENTERTAINMENT

Home > Cryptocurrency > Cryptocurrency News > Crypto Scammers Using A

Crypto Scammers Using AI Deepfakes to Spoof KYC Verification on Exchanges, Binance Security Chief Says

Deepfakes are artificially generated photos or videos designed to can convincingly replicate the voice as well as facial features of an individual.

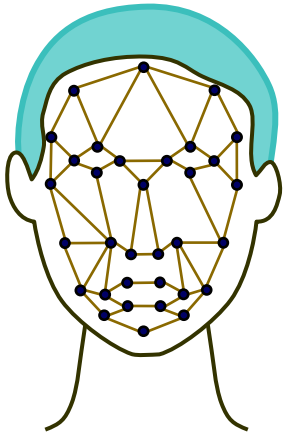
Written by Radhika Parashar, Edited by David Delima | Updated: 24 May 2023 15:13 IST

Share on Facebook | Tweet | Snapchat | Share | Reddit | Email | Comment | Google News

Jøsang

KI og cyber

State-of-the-art Deepfake



- Deepfake av Mette Frederiksen statsminister i Danmark, laget av Morten Messerschmidt i Dansk folkeparti
  - https://twitter.com/MrMesserschmidt/status/1783882247323492725
- Deepfake av Fredrik Solvang
  - https://www.nrk.no/norge/ki-stunt-i-debatten-_-deepfake-av-fredrik-solvang-pa-direkten-1.16901230
- Olga Loiek stålet avatar i Kina
  - https://www.youtube.com/watch?v=3FQSFnZpsqw
- Microsofts VASA-1 (Visual Affective Skills Avatar)
  - https://www.microsoft.com/en-us/research/project/vasa-1/

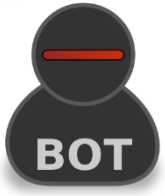
  «We have no plans to release an online demo, API, product, additional implementation details, or any related offerings of VASA until we are certain that the technology will be used responsibly and in accordance with proper regulations.»

Deteksjon av deepfake

- Deteksjon avhenger av at det ser **uekte** ut
- Deepfake lages for å se ekte ut
- Uløselig problem

- Kryptografisk autentisering av video?

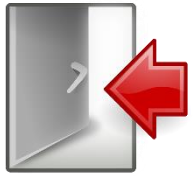
Produksjon av skadevare

- Datavirus 
- Løsepengevirus 
- Spionvare 
- Bott-program 

- Exploit 

- Makro-virus 

- Trojaner 
- Dataorm 
- Rootkit 

- Bakdør 
- Skadelig JavaScript 

- Logisk bombe 

Angrepsautomatisering

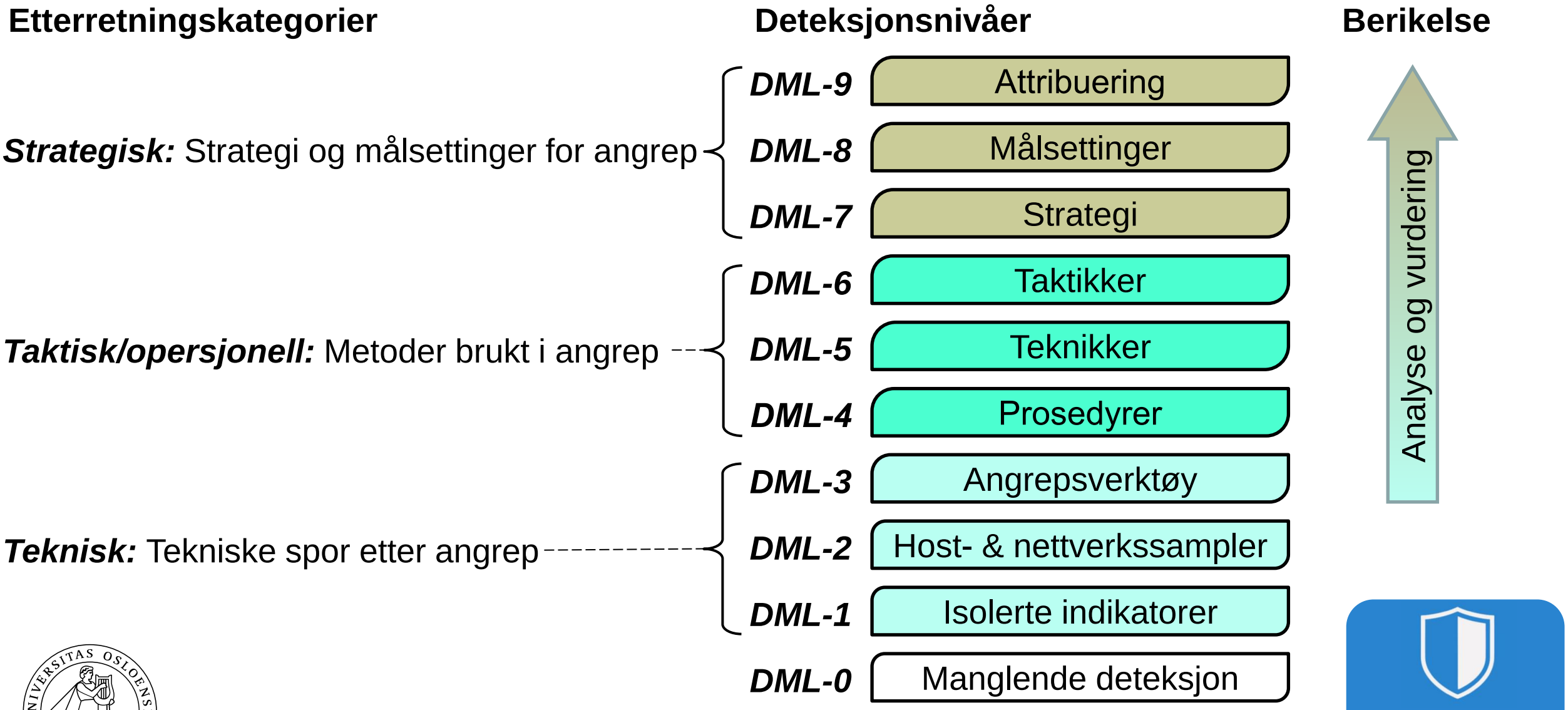
Inntrengningsdeteksjon

Defensiv

Analyse av skadevare

Defensiv

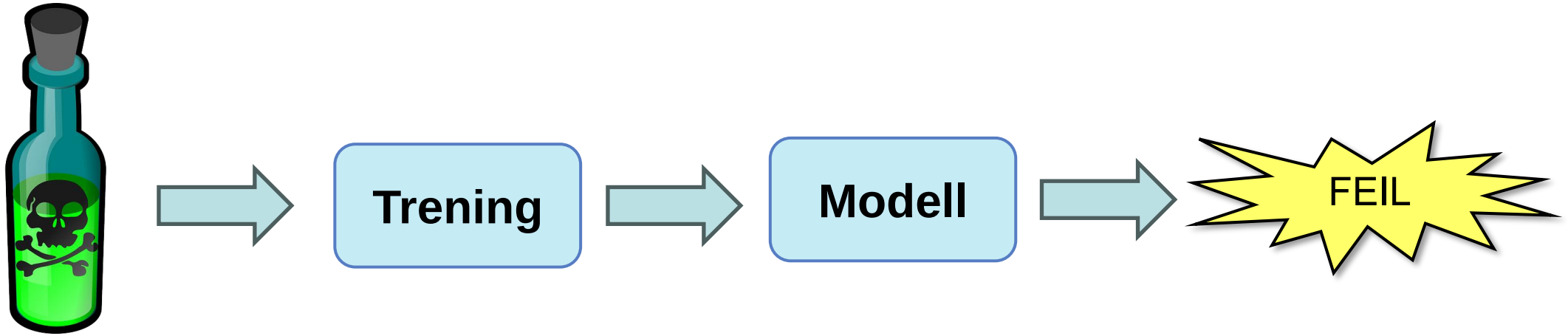
Digital trusseletterretning



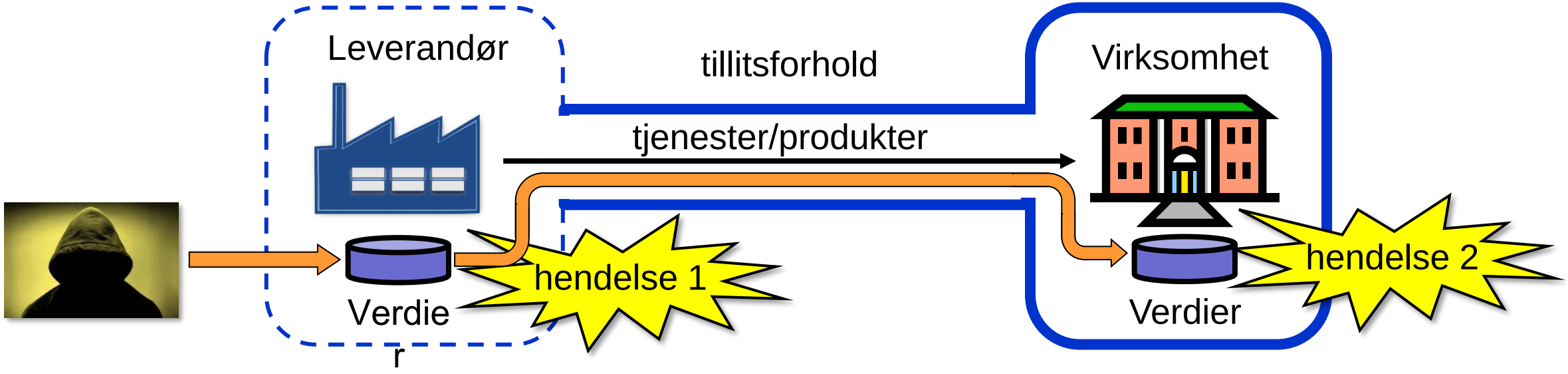
Hendelsesrespons

Defensiv

Forgiftning og forurensing av læringsdata

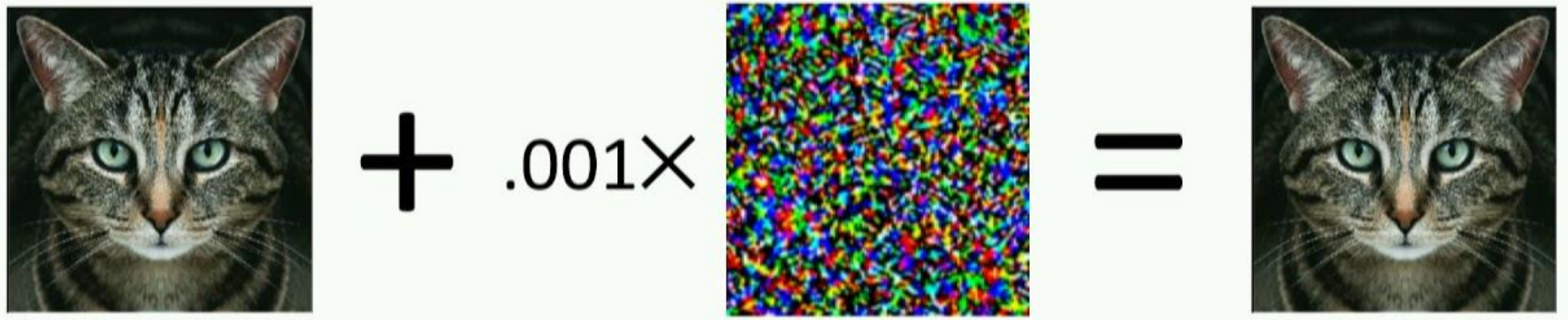


Angrep på leveransekjeden



Synsbedrag

#RSAC



I. J. Goodfellow, J. Shlens and C. Szegedy. Explaining and harnessing adversarial examples. 2015

RSAConference2019

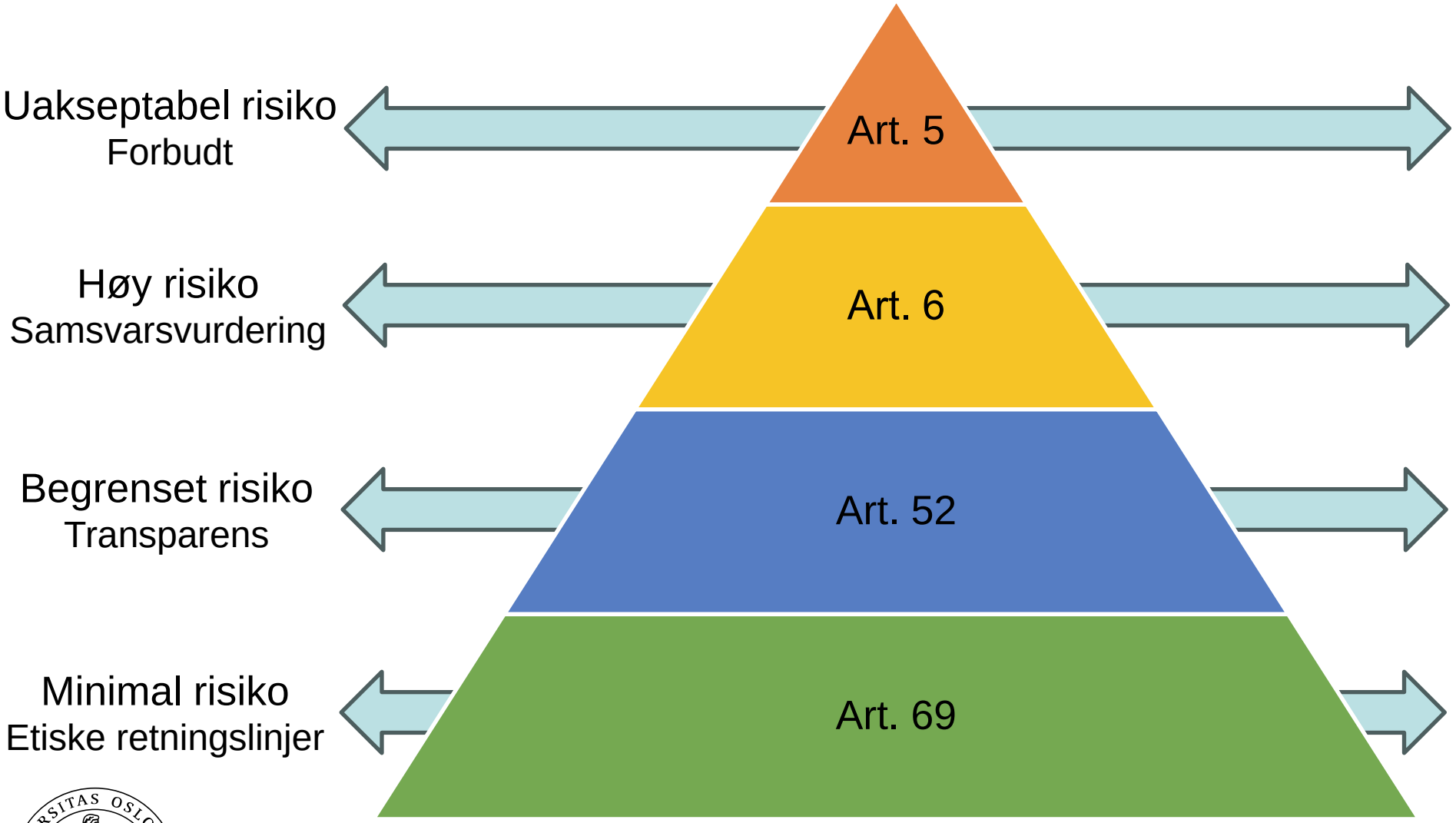
Synsbedrag

Human: 100.0 % stop sign
Machine: 99.7 % stop sign

Human: 100.0 % stop sign
Machine: 0.9 % stop sign

Jailbrake, lekkasje og injeksjonsangrep

EUs KI-forordning (AI Act)



Figur etter:
https://www.hjort.no/bruk-av-kunstig-intelligens-i-virksomheter/

EUs KI-forordning (AI Act)

- Lagt fram av Europakommisjonen 21. april 2021
- Vedtatt av Europaparlamentet 13. mars 2024
- Vedtatt av Europarådet 21. mai 2024
- Iverksatt juni 2024
- Håndhevelse av forbudt KI desember 2024
- Innføring av praksisnormer mars 2025
- Håndhevelse av generell KI juni 2025
- Håndhevelse av høyrisiko KI juni 2027

- Det forventes at EØS (Norge, Island, Lichtenstein) vil innføre KI-forordningen

https://create.kahoot.it
https://kahoot.com