

# Hallusinerer chatGPT?

Øyvind Ryan

Førstelektor

Matematisk institutt

Universitetet i Oslo

# Kunstig intelligens (KI) – Ekko fra en gammel historie.



WIKIPEDIA  
The Free Encyclopedia



Create account

Log in



## Clever Hans

🌐 29 languages ▾

Article [Talk](#)

[Read](#) [Edit](#) [View history](#)

From Wikipedia, the free encyclopedia

*For Grimm's tale, see [Clever Hans \(fairy tale\)](#).*

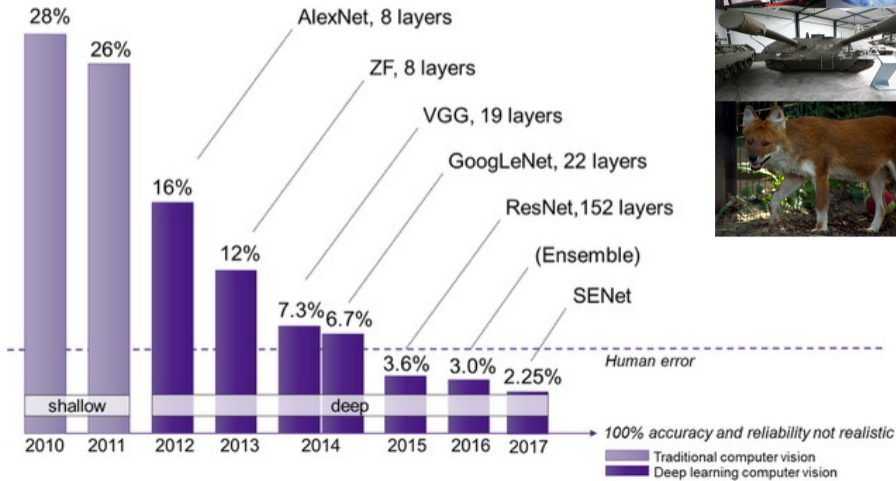
**Clever Hans** (*German*: *der Kluge Hans*; c. 1895 – c. 1916) was a horse that was claimed to have performed [arithmetic](#) and other intellectual tasks. After a formal investigation in 1907, [psychologist Oskar Pfungst](#) demonstrated that the horse was not actually performing these mental tasks, but was watching the reactions of his trainer. He discovered this artifact in the [research methodology](#), wherein the horse was responding directly to involuntary cues in the [body language](#) of the human trainer, who was entirely unaware that he was providing such cues.<sup>[1]</sup> In honour of Pfungst's study, the



Clever Hans performing in 1904



# KI før og etter 2012...



# Hva lærer KI?

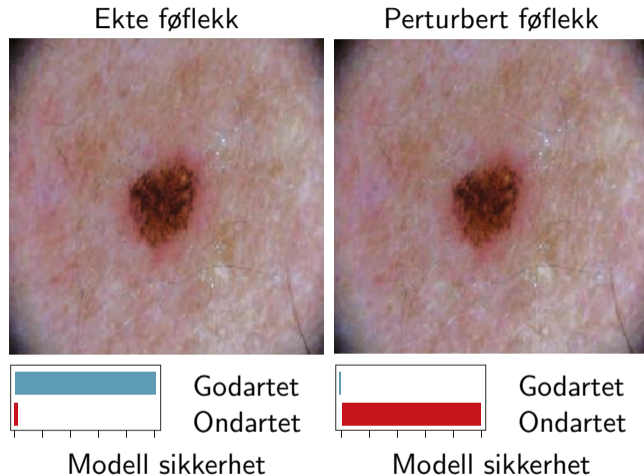


# Medisinske bilder

Vi har trent opp et KI system til å klassifisere føflekksvulster som godartede/ondartede. Det er 4 muligheter:

- (i) KI-systemet sier at svulsten er ondartet, og svulsten er ondartet i virkeligheten og.
- (ii) KI-systemet sier at svulsten er godartet, og svulsten er godartet i virkeligheten og.
- (iii) KI-systemet sier at svulsten er ondartet, men svulsten er i virkeligheten godartet (falsk positiv)
- (iv) KI-systemet sier at svulsten er godartet, men svulsten er i virkeligheten ondartet (falsk negativ)

# KI kan ha en ikke-menneskelig oppførsel...



**Figur fra:** Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L., & Kohane, I. S., 2019, 'Adversarial attacks on medical machine learning'. *Science*, 363(6433), 1287-1289.

# Verden er full av universelle grenser

- ▶ Det finnes mange teoretiske grenser i matematikk
  - ▶ Grenser for hvor mye vi kan komprimere tekst.
  - ▶ Grenser for hvor høye frekvenser vi kan representere med digital lyd.
  - ▶ Grenser for hvilke problemer vi kan løse på datamaskin.
- ▶ Nå trenger vi å forstå begrensningene til KI.
- ▶ KI er kjempenyttig! ...men på lang sikt vil det være sunt å forstå begrensningene.

# Ressurser

- ▶ Kloke Hans
- ▶ Imagenet challenge
- ▶ Farene ved Tesla's 'Full Self-Driving mode'. vist under Super Bowl 2023