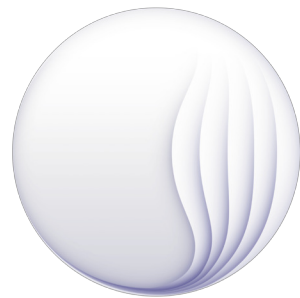


Den danske AI-model – tilliden skal genskabes

Thomas Bolander, DTU

Et par eksempler på dansk AI



corti



LEO[®]
Innovation Lab

insulin insensitivity
in obese children
=
overweight girls may have
reduced hormone responses

ALARM

1 1 2

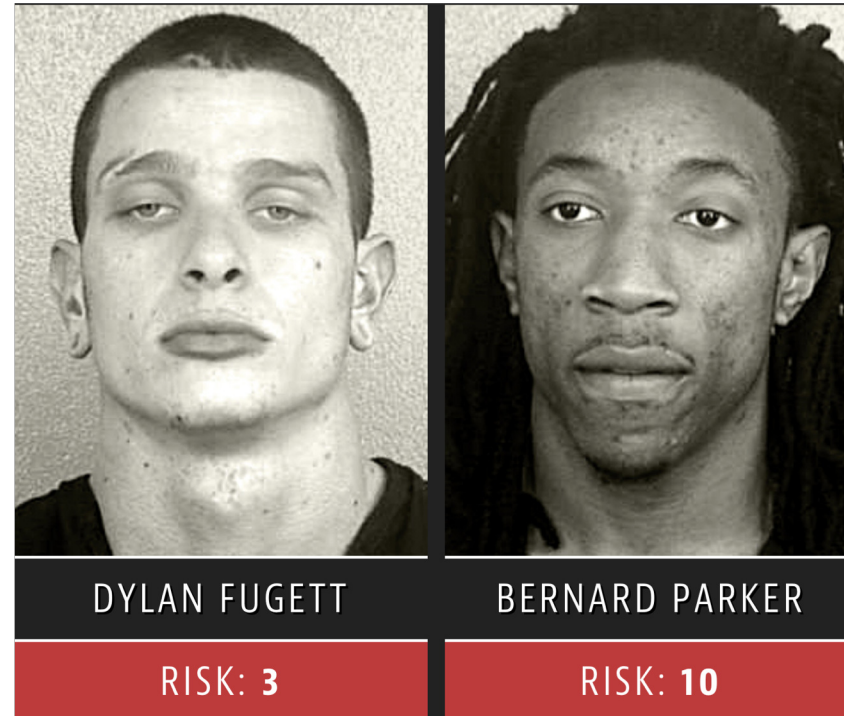
Det direkte nummer
ved akut udrykning og livsfare

Har du akut behov for politi, ambulance, brandvæsen eller miljøberedskab skal du ringe 112, uanset hvor i landet du befinder dig.

POLITI



Bagsiden af AI-medaljen: tilliden svigter



	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

(Angwin et al.: Machine Bias, ProPublica, 23 May 2016)

Bagsiden af AI-medaljen: tilliden svigter



Tesla-ulykke, juni 2016



Uber Volvo-ulykke, marts 2018

”We cannot argue that a pedestrian detector is safe simply because it performs well on a large data set, because that data set may well omit important, but rare, phenomena (for example, people mounting bicycles). We wouldn’t want our automated driver to run over a pedestrian who happened to do something unusual.”

(Russell & Norvig: *Artificial Intelligence – A Modern Approach*, 3ed, 2010)

Hvordan genskabes tilliden?

Tilliden er udfordret pga:

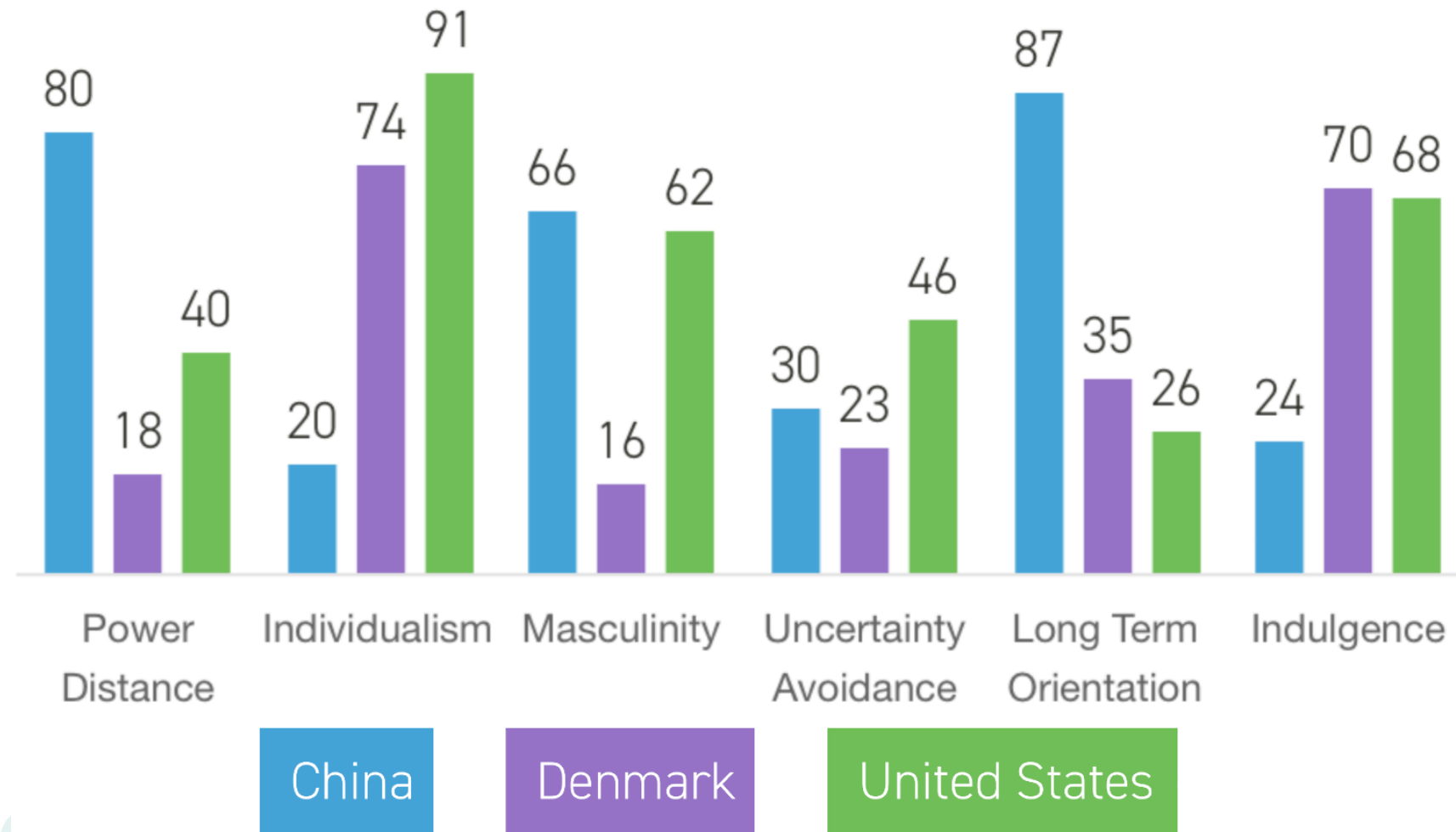
- **Mangel på robusthed:** Systemerne begår (uundgåeligt) fejl.
- **Mangel på forklarlighed:** Fejlene er uigennemskuelige og vi får ingen forklaringer.

Problemerne kan ikke løses med eksisterende AI-teknikker alene. Kræver at vi:

1. Tænker grundigt over **hvad vi bruger AI til.**
2. **Forsker** i at løse problemerne omkring **forklarlighed.**

Hvad vi bruger AI til – et dansk perspektiv

- Mennesket i centrum.
- Human-in-the-loop.
- Samarbejde, tværfaglighed, sociale aspekter.
- Transparens, etik, demokrati.



(Geert Hofstede's "Dimensions of Culture": www.hofstede-insights.com)

Løse problemerne omkring forklarlighed

- **Grundforskning** har skabt de teknikker indenfor AI som har givet os et overfløddighedshorn af betydningsfulde anvendelser.
- **Grundforskning** skal også til for at forbedre eksisterende teknikker og udvikle nye, som kan løse de problemer som den nuværende generation af AI-teknikker står overfor.

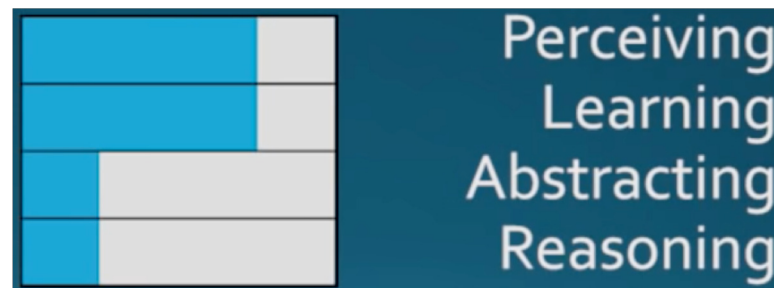


Løse problemerne omkring forklarlighed

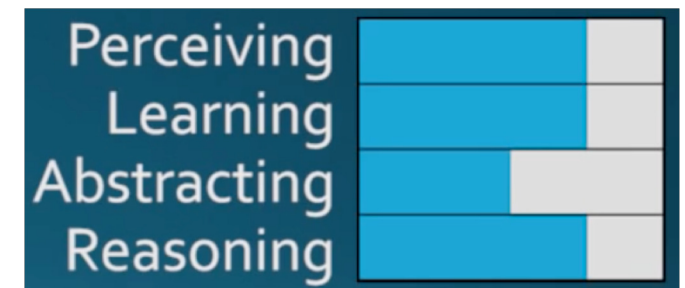
- Forskning i bedre integration af de teknikker som bruger sprog og eksplicite model (første bølge af AI) med de teknikker som opbygger implicite modeller via læring (anden bølge af AI).
- Målet er: læring **og** forklarlighed/robusthed (tredje bølge af AI, som pt forskes i).



Første bølge i AI:
Håndkodet viden



Anden bølge af AI:
Statistisk læring



Tredje bølge af AI:
Kontekstuel tilpasning

Den danske model: AI rundet af danske værdier

Teknologi skal øge gennemsigtighed, retfærdighed og forklarlighed, ikke sænke dem.

"The measure of success for AI applications is the value they create for human lives. In that light, they should be designed to enable people to **understand AI systems successfully, participate in their use, and build their trust.**"

"AI technologies already pervade our lives. As they become a central force in society, the field is shifting from simply building systems that are intelligent to building intelligent systems that are **human-aware and trustworthy.**"

(One Hundred Year Study on AI: 2015-2016, Stanford University, 2016)