

Trustworthy and Responsible AI

– How to create and use AI solutions that we can trust?

Presentation at HAVTILs “Innovasjonsdagen 2024”

Frank Børre Pedersen
Programme Director, Digital Assurance
Group Research and Development

06 March 2024

AI is already here

The AI wave is a tsunami

It is already reshaping how we work and creating unprecedented opportunities



AI is now almost everywhere

Industrial AI for safety-critical applications is in the making....

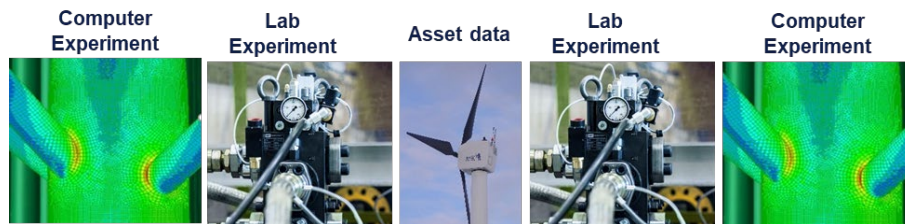
DETECTING CORROSION AND CRACKS



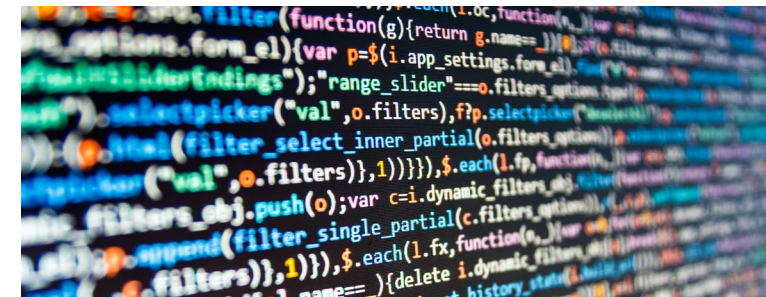
DIGITAL REQUIREMENTS AND VERIFICATION



HYBRID APPROACHES: MODELS + ALGORITHMS

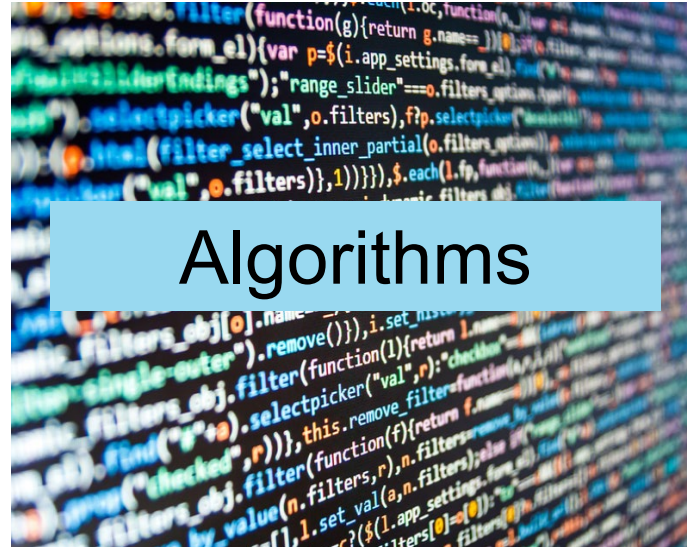


AUTOMATED CODE GENERATION



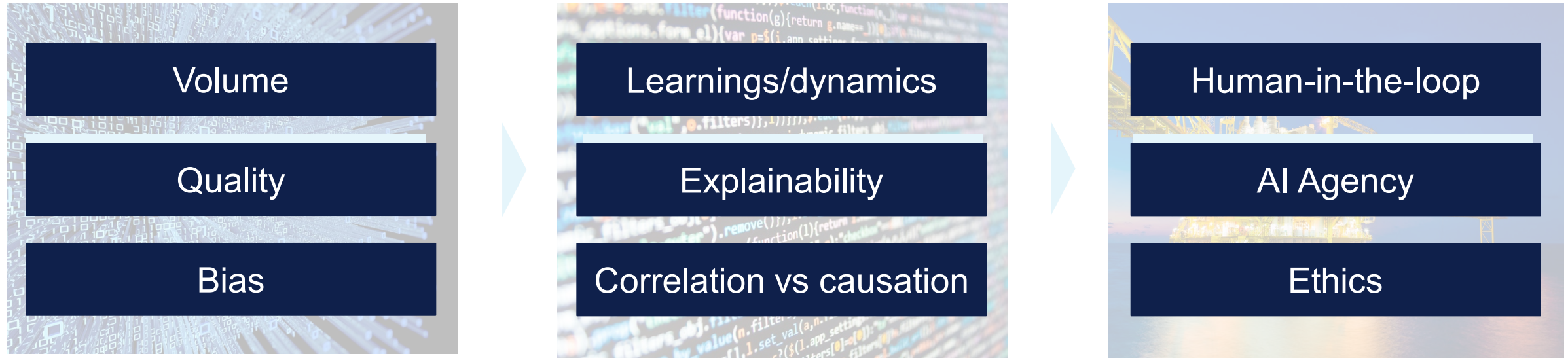
But what is AI really?

AI can be both rule-based and machine learning (and normally a combination)



AI risks are of a different nature

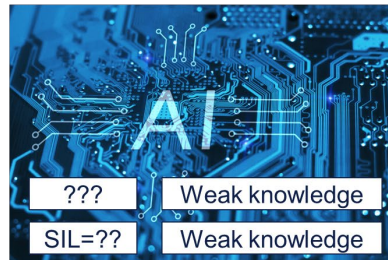
Risks originate along all building blocks of AI



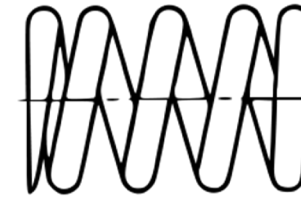
AI challenges our conventional risk approaches

Safety-critical applications need to be trusted

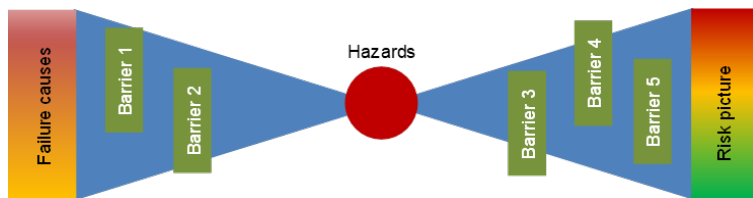
Requirements



Independence and Fail-safe



Barriers and barrier status



Management of change



Steps towards trustworthy AI has been made

One challenge still remains – how to put these legislations into practise

The EU AI Act



Biden's Executive Order



Bletchley Park AI Summit

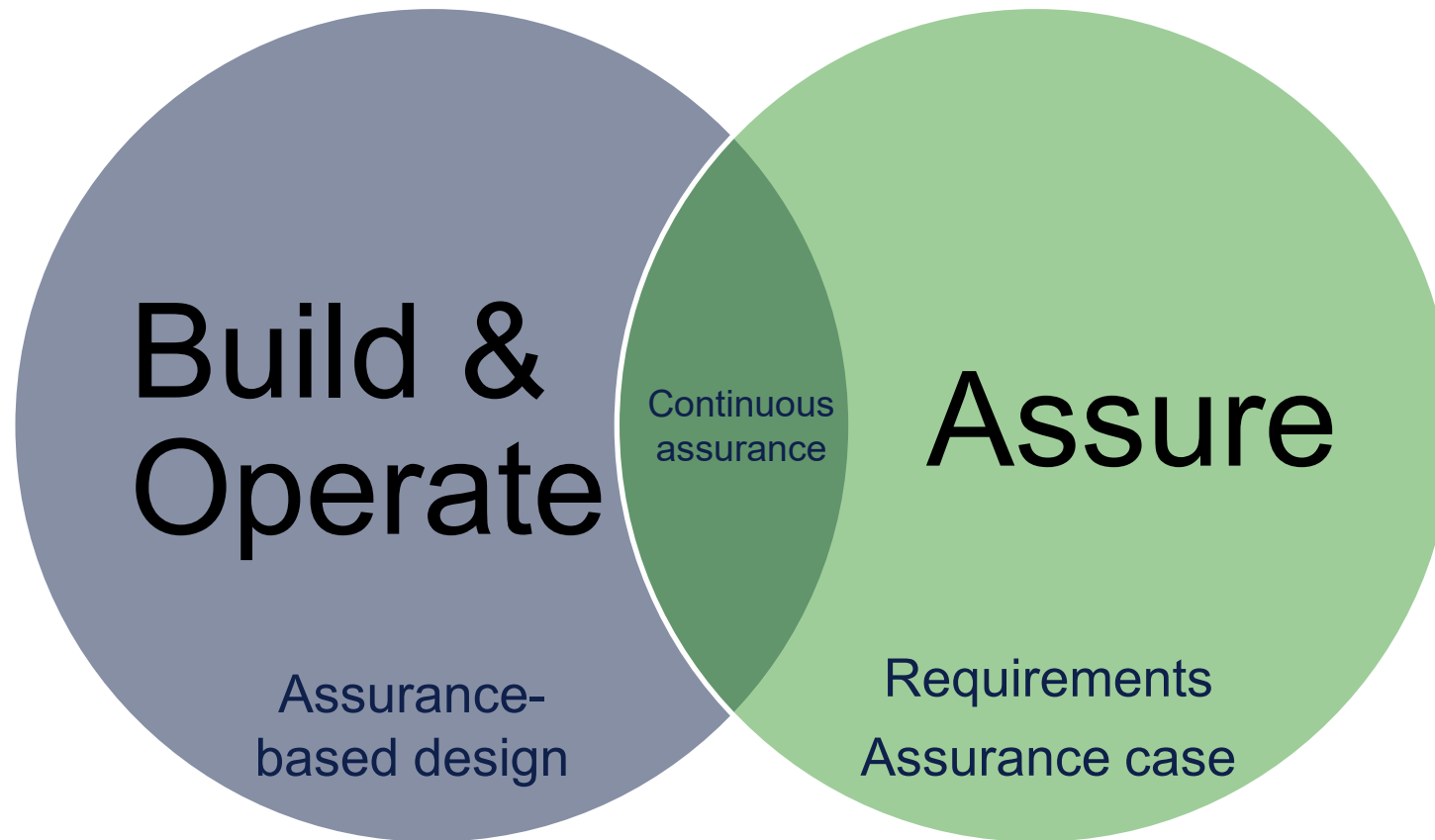


01/11/2023. Bletchley, United Kingdom. Prime Minister Rishi Sunak poses for a family photo with world leaders and business people on day two of the UK AI Summit at Bletchley Park. Picture by Simon Dawson / No 10 Downing Street

Trustworthy and Responsible AI: *Conformity assessments and testing*

Building and assuring becomes more connected

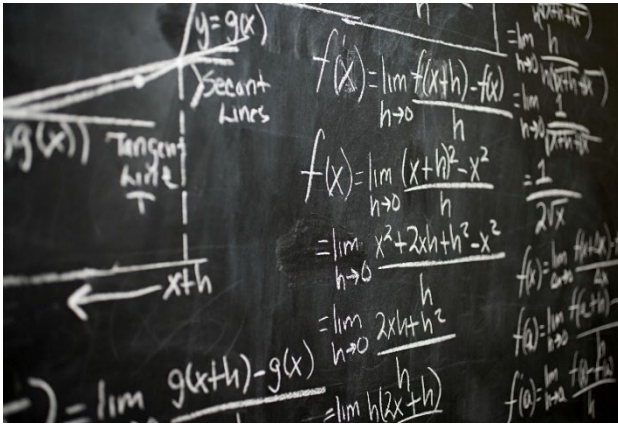
Independence still crucial, but takes a slightly different form



1. Building trust into AI algorithms

Bringing all knowledge into AI – not just data

Science-guided AI



- Combine physical-models with data-driven models (“hybrid-models”)
- Include physical constraints into AI
- Use synthetic data for safety-critical scenarios with little data

Uncertainty-aware AI



- AI that knows when it is uncertain
- Avoid confident mistakes
- Tells you when you cannot trust it

Causal AI



- Understands cause-and-effect
- Distinguishing correlations and causation
- Distinguishing observation from intervention

2. Assure using proper approaches

A combination of technology qualification, risk assessment and assurance case is used

DNV

RECOMMENDED PRACTICE

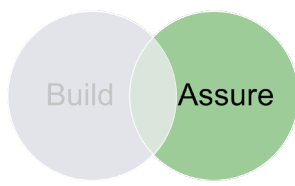
DNV-RP-0671 Edition September 2023

Assurance of AI-enabled systems

The PDF electronic version of this document available at the DNV website dnv.com is the official version. If there are any inconsistencies between the PDF version and any other available version, the PDF version shall prevail.

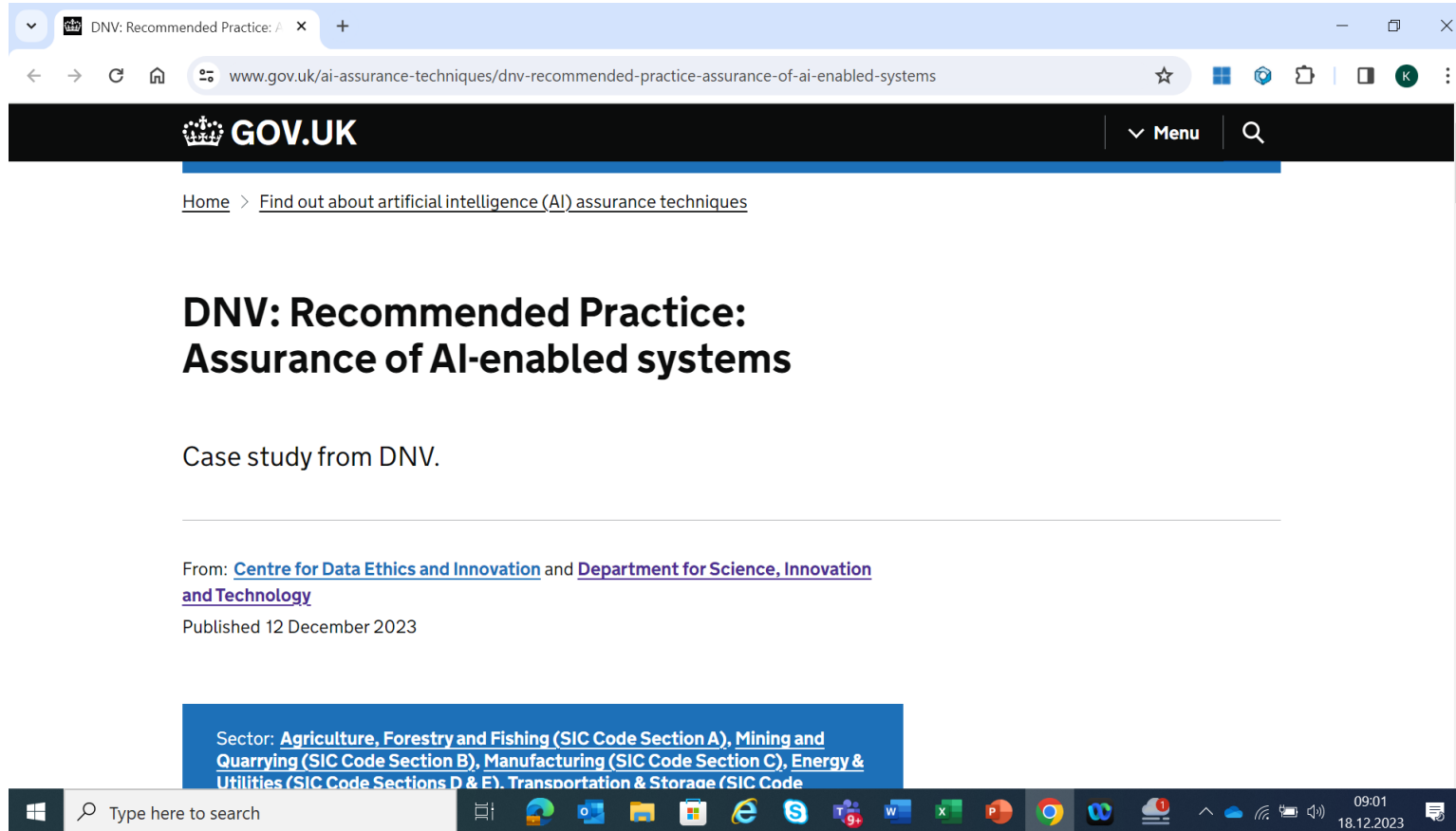
DNV AS





2. Assure using proper approaches

A combination of technology qualification, risk assessment and assurance case is used



Department for Science, Innovation and Technology (DSIT)

Centre for Data Ethics and Innovation (CDEI)

Include description of the RP and links to dnv.com:

- Digital trust by DNV
- Artificial intelligence

Links to gov.uk:

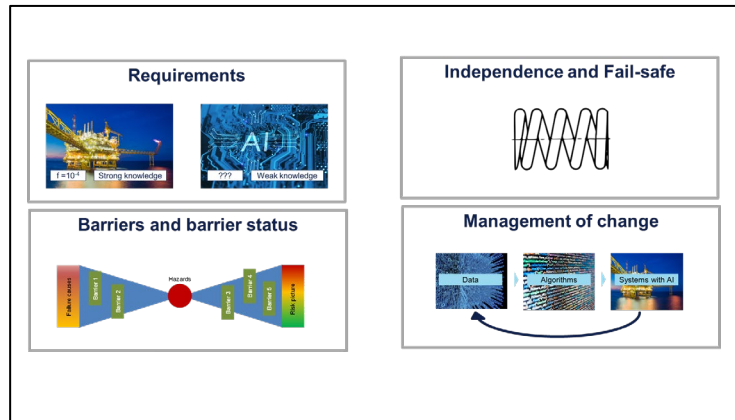
<https://www.gov.uk/ai-assurance-techniques/dnv-recommended-practice-assurance-of-ai-enabled-systems>

<https://www.gov.uk/guidance/cdei-portfolio-of-ai-assurance-techniques>

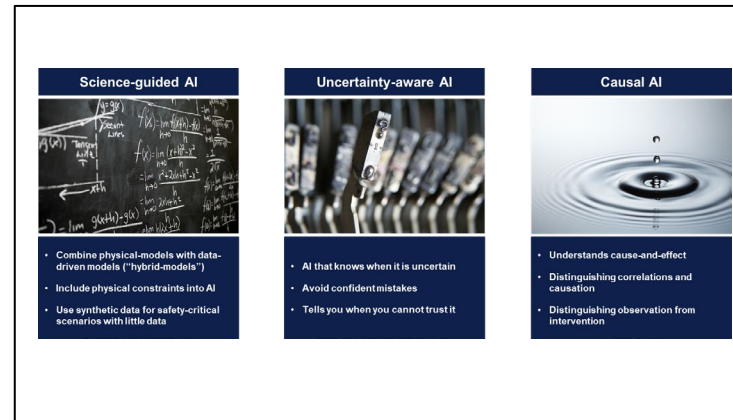
Summary

By designing proper AI and using novel assurance approaches we can have sufficient trust in AI

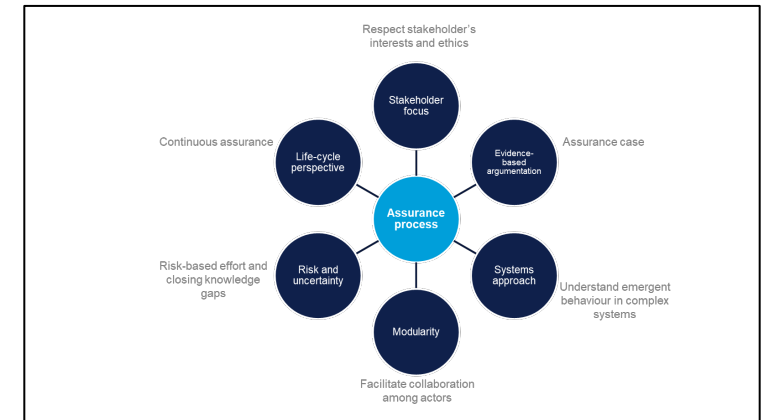
1 AI is difficult to trust initially



2 Build AI that can be trusted



3 Assure so you can trust it



Thank you for your kind attention.

Visit www.dnv.com/research

Frank.Borre.Pedersen@dnv.com

Mobile +47 90552666

www.dnv.com

