WHEN TRUST MATTERS

Ansvarlig teknologiutvikling

Sikkerhetsforums årskonferanse 2024

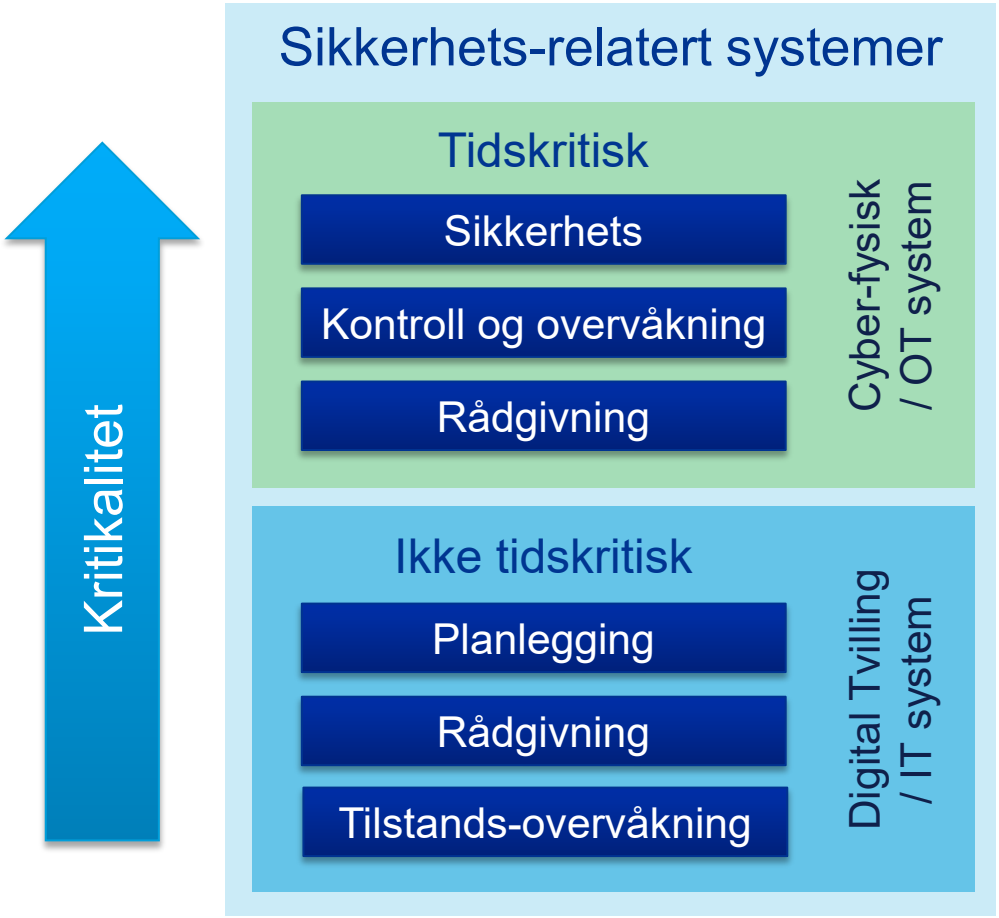
Christian Markussen

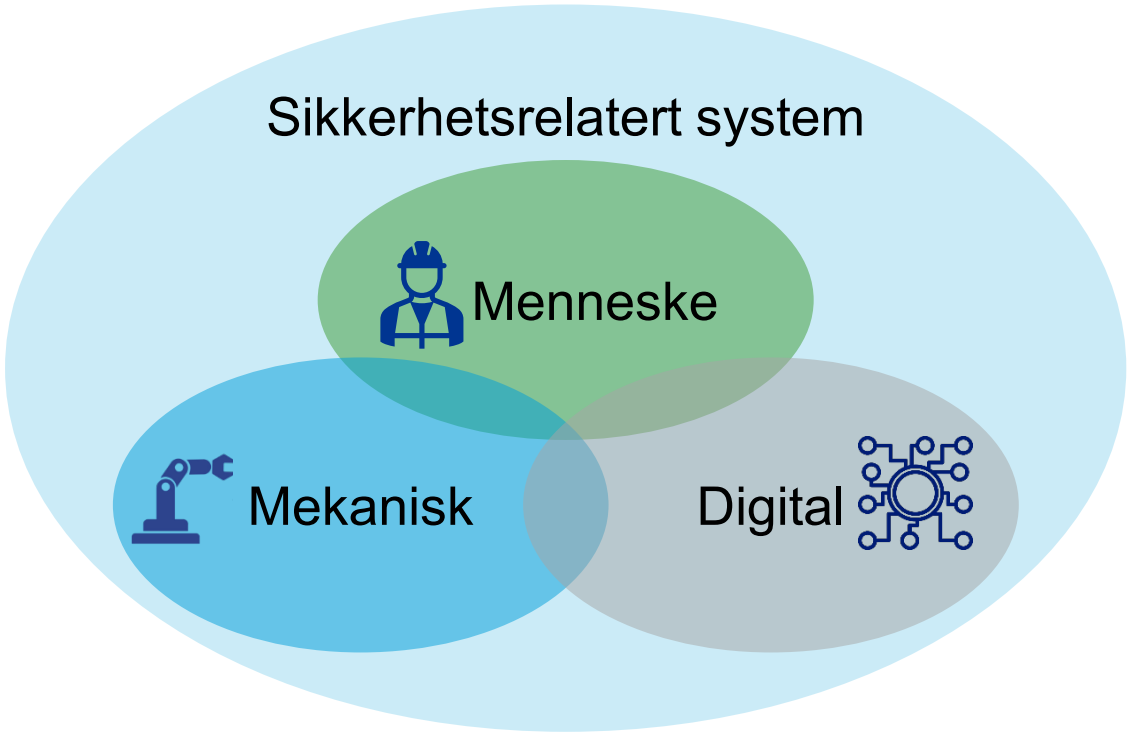
06 June 2024

Hvilken trender ser vi?

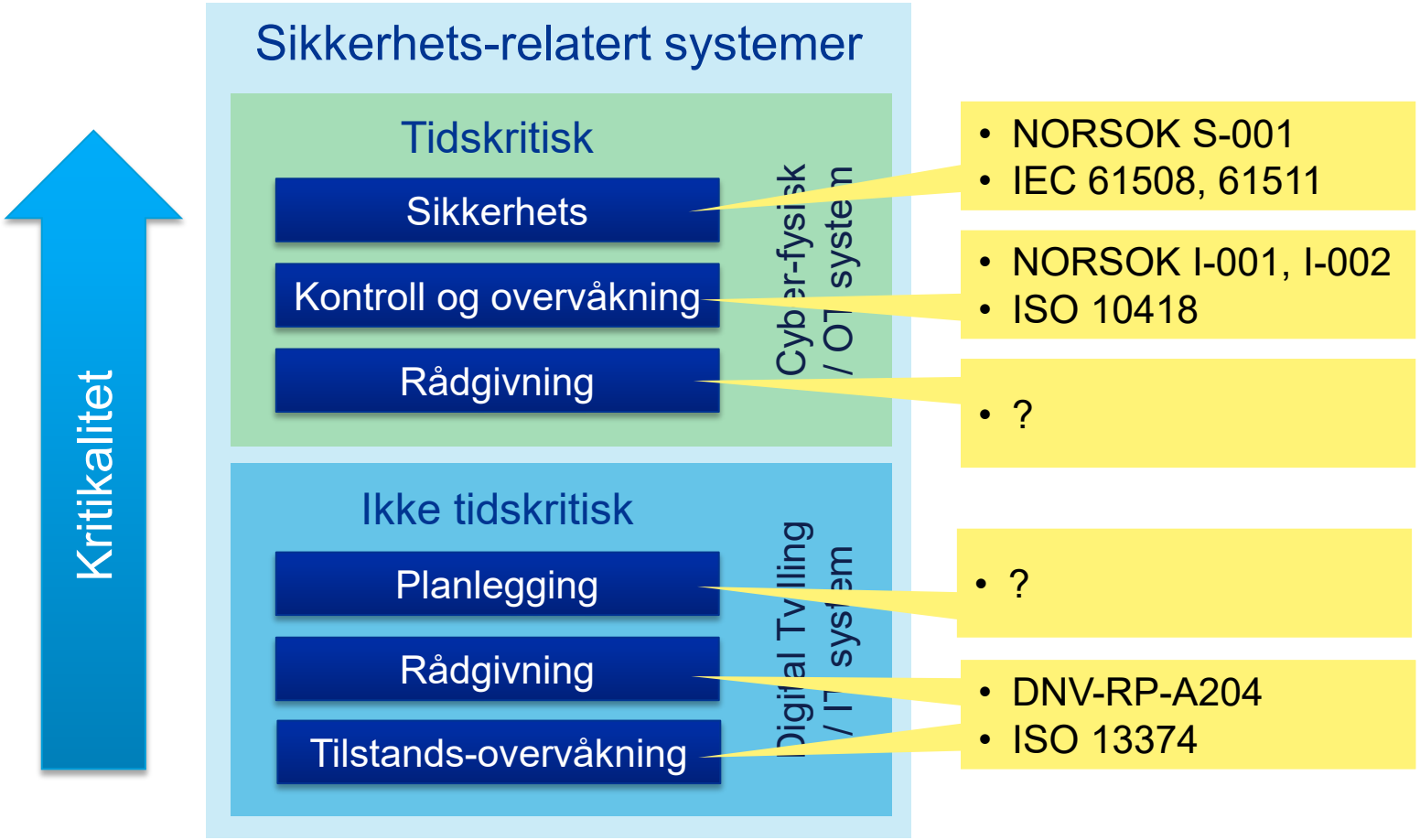
- Mere komplekse systemer
  - Kunstig Intelligens
  - Fysikk-baserte modeller / simuleringsmodeller,
  - Modeller satt sammen til «system-av-systemer»
- Mer integrasjon
  - Data fra mange systemer og disipliner blir brukt av hele organisasjonen
  - Applikasjoner fra forskjellige leverandører blir integrert sammen
  - Integrasjon av OT og IT
- Høyere endringshastighet
  - Utviklings-hastigheten øker i takt med digitaliseringen
  - Benytter «agile» utviklingsprosesser som stammer fra software utvikling
  - Utfordrer organisasjoners styringsmodell for håndtering av data, modeller og teknologiutvikling
- Teknologiene blir mer driftskritisk, både finansielt og sikkerhetsmessig

Sikkerhets-relaterte systemer





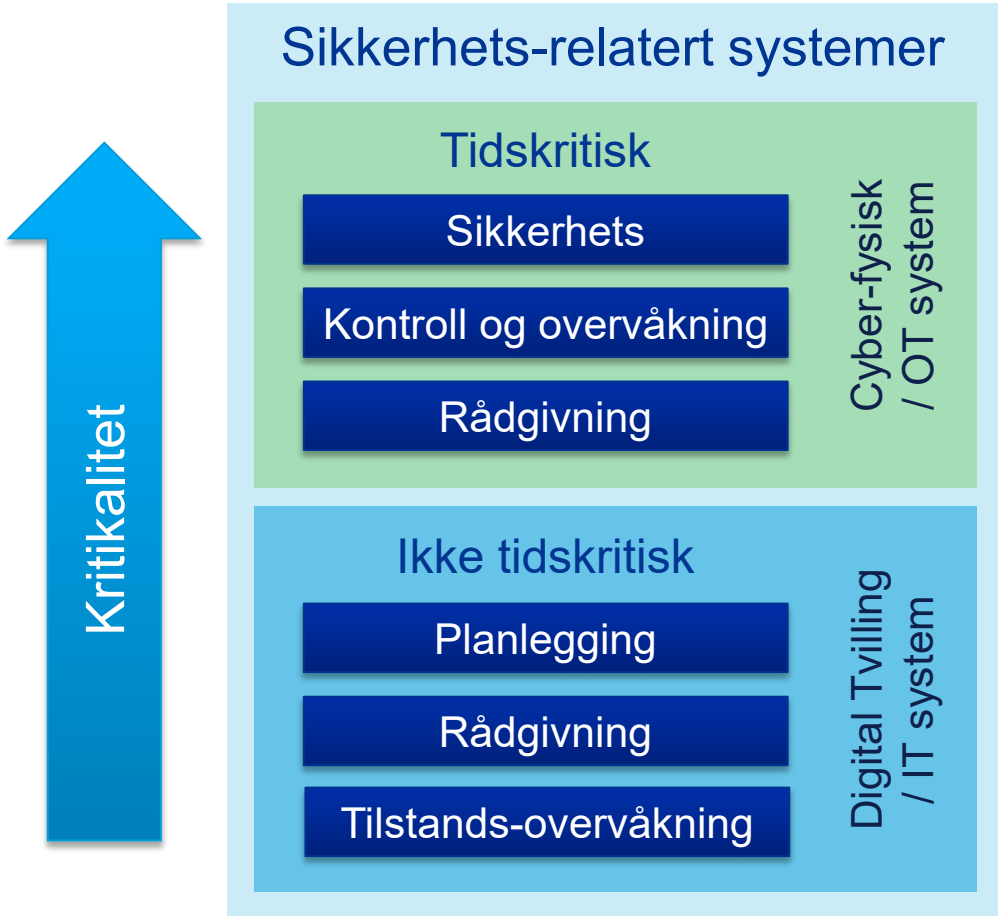
Sikkerhets-relaterte systemer



Hva med integrering av KI og fysikk-baserte modeller?

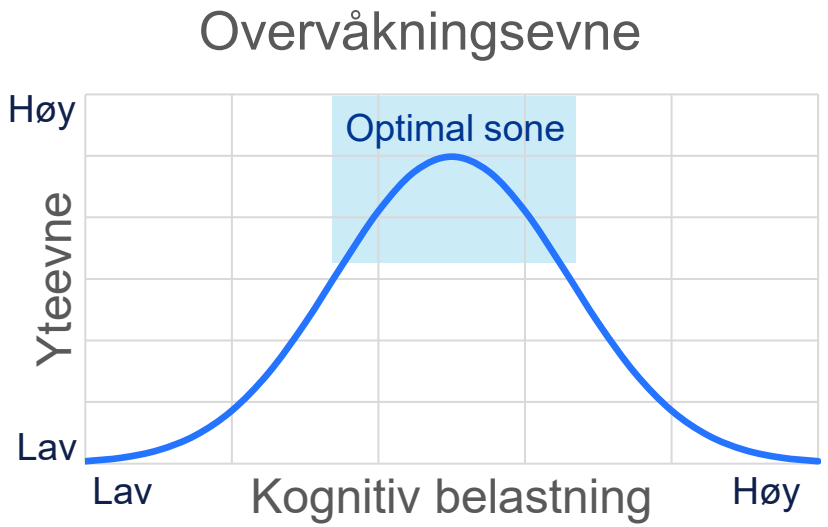
- DNV-RP-A203 Technology Qualification
- ISO/IEC TR 5469 Functional safety
- DNV-RP-0671 Artificial intelligence
- DNV-RP-0665 Machine learning
- ISO/IEC 23053 Framework for AI & ML
- DNV-RP-0497 Data quality
- DNV-RP-0513 Simulation models

Så hva er problemet?



- Manglende standarder og retningslinjer for rådgivning-systemer
- Modeller utviklet og kvalifisert for lavkritiske applikasjoner blir brukt i mere høykritiske applikasjoner
- Rådgivnings-systemer, brukt aktivt i operasjoner med potensiale for storulykker, er implementer i IT systemer / skyløsninger
  - Manglende robusthet og pålitelighet
- «Det er ikke et sikkerhets-system siden vi har menneske-i-loopen»
  - Dette forutsetter at operatøren her riktig situasjons-forståelse basert på informasjon fra systemene
  - Mangler ofte overvåkning av det digitale systemet og metode for å informere operatøren om degradert informasjonskvalitet.
- Integrasjon av avanserte modeller og KI i sikkerhets-relaterte systemer er komplisert og krever en kombinasjon av flere standarder
- Umoden tilnærming på kvalitetssikring av slike systemer
  - Ikke tilstrekkelig å sikre digitale systemer ved «black-box» testing
  - Tre-parts samarbeid ➔ hvordan skal operatøren overholde “påseplikten”?

Menneske-sentrert design



- Cognitive load
- Situational awareness
- Available time
- Stress & fear

- Human centred design
- HMI design
- Alarm system
- Status / quality indicators

Human capacity

Competence

Robustness

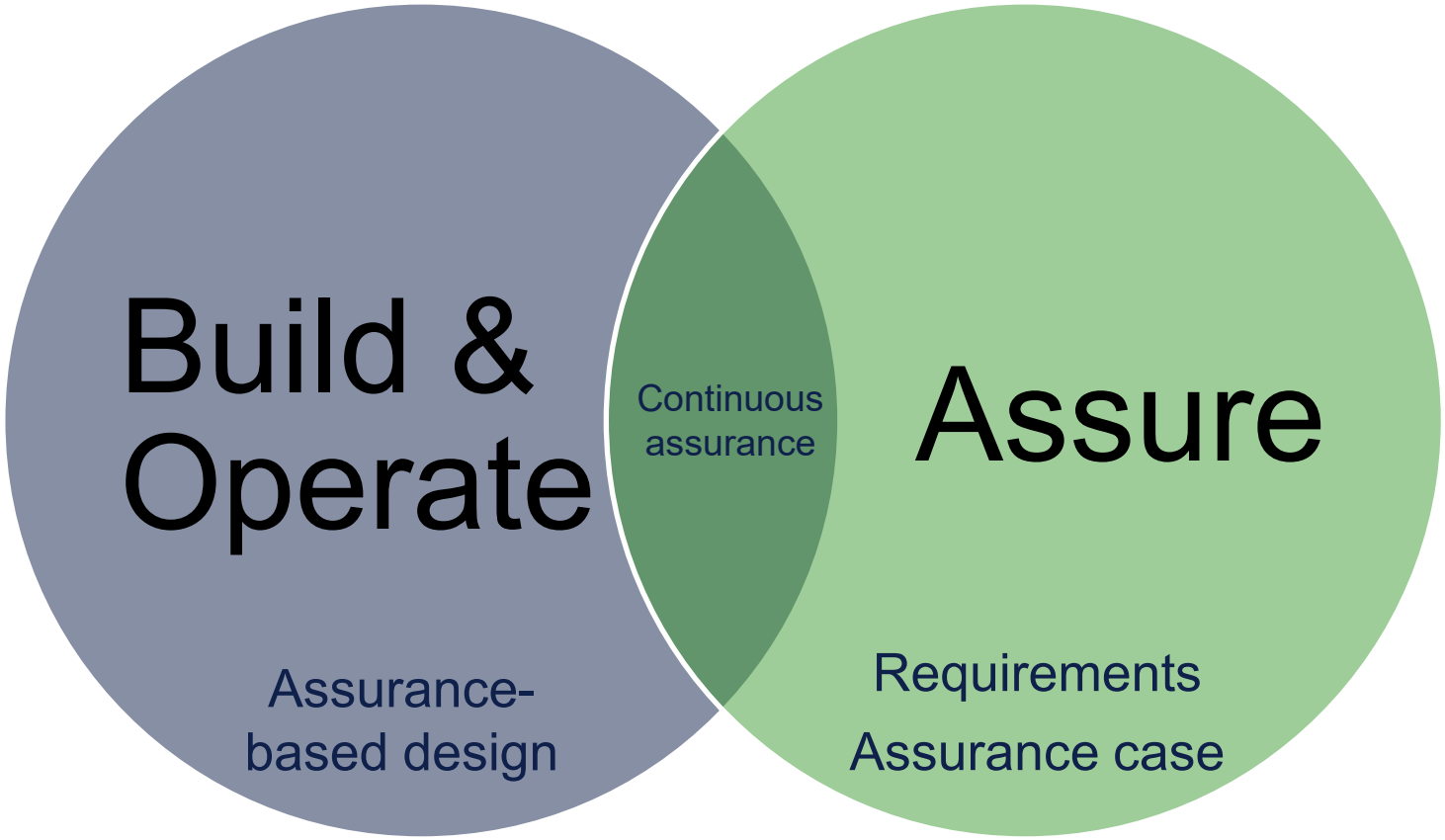
Technology interface

Governance

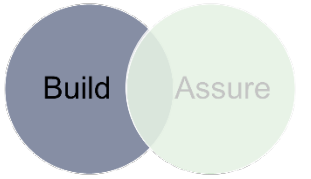
- Competence and training
- Quality of training
- User manuals

- Management engagement
- Roles & responsibilities
- Procedures

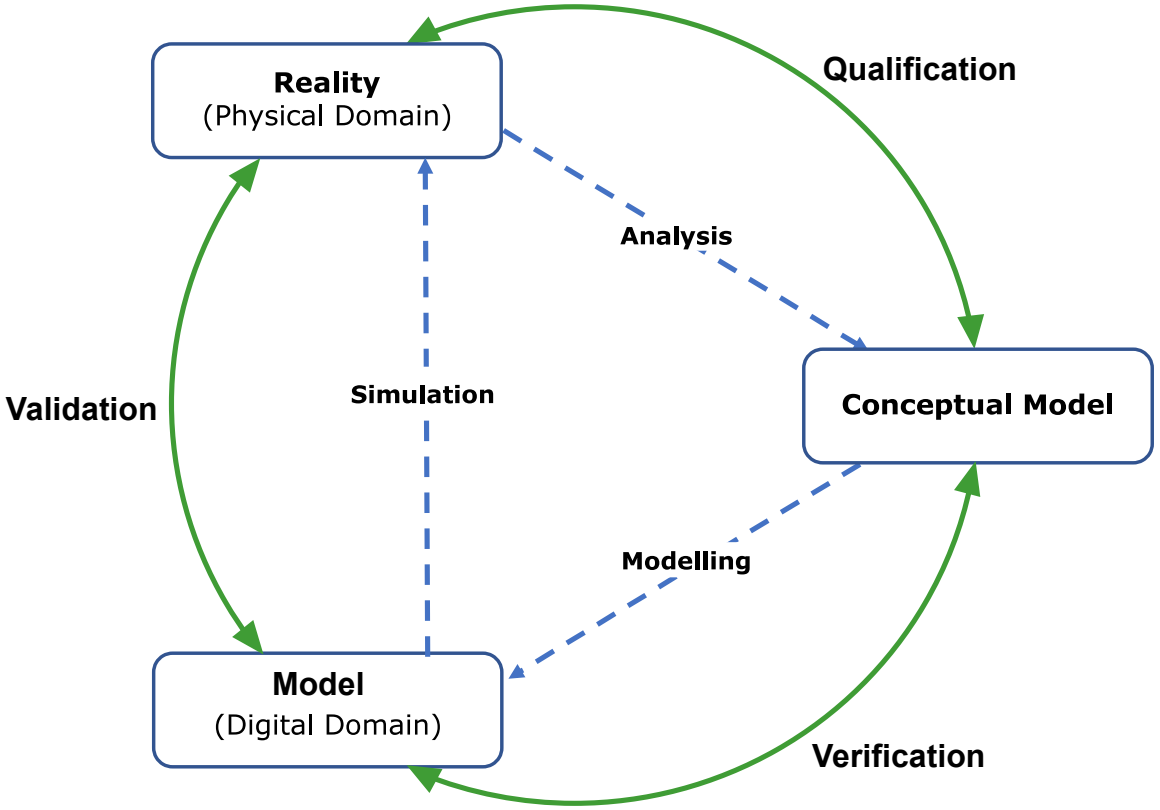
AI – Integrating building and assurance



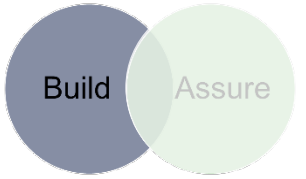
Conceptual Model



- From experience, most modelling problems stems from a lack of a **Conceptual Model**
- Conceptual Model
  - bridge between reality and the model
  - ensures a common understanding of the model objectives, requirements, assumptions, etc.
  - bridge the gap between the domain expert, the modeller and the end user
- Purpose:
  - provide a model specification for the intended application
  - basis for verification and validation of the model
    - Black-box → Gray or white box verification
  - element in the “duty to ensure”



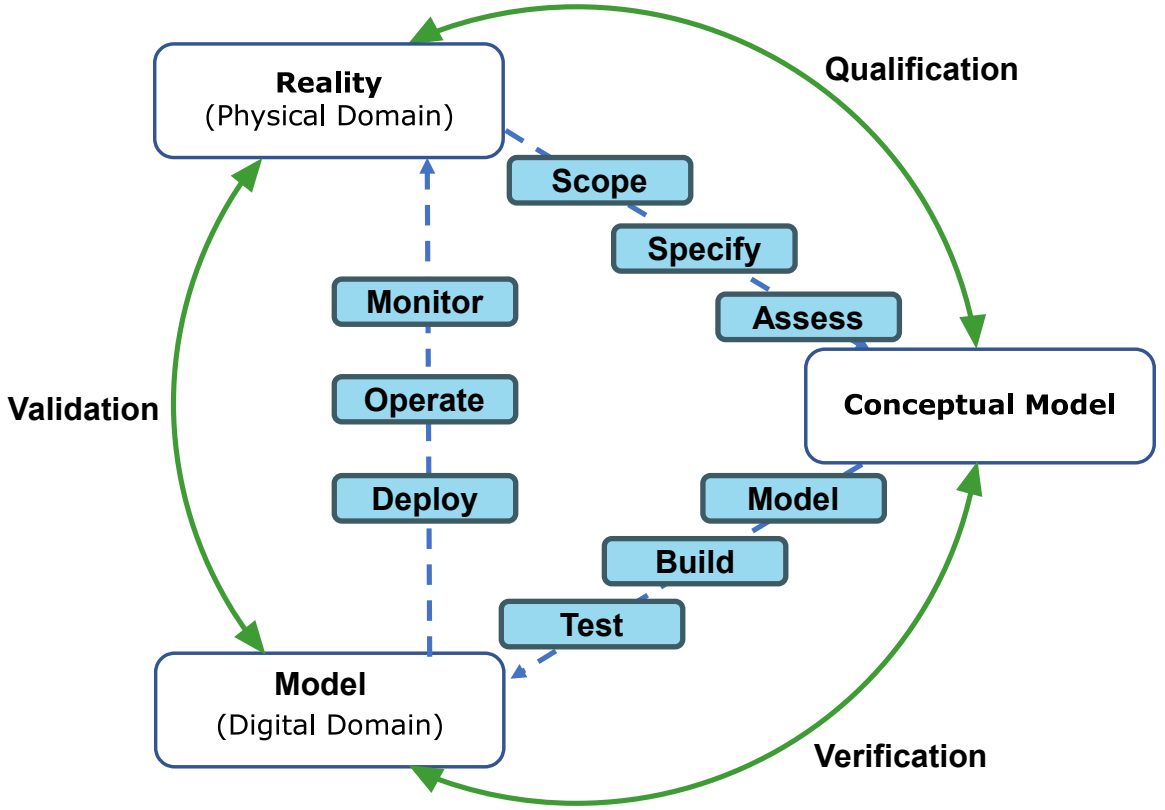
Conceptual Model



Includes:

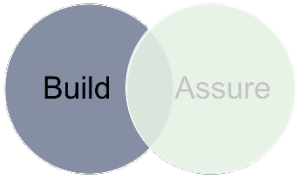
- Objectives
- System boundary
- Requirements
- Modelling concepts
- Assumptions
- Risks
- Validation plan (white- or gray-box vs. black-box)

➔Model specification (verified and agreed)

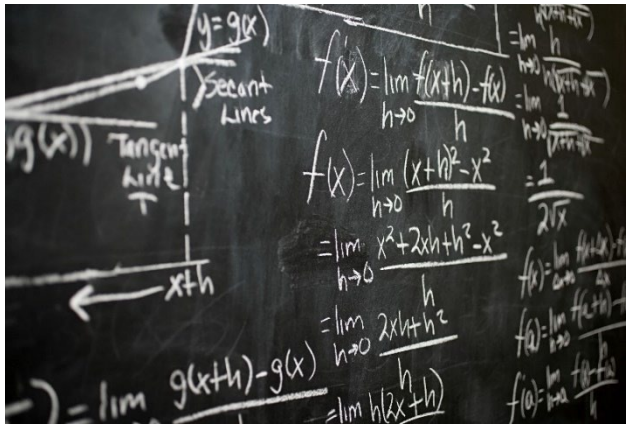


Building trust into AI algorithms

Bringing all knowledge into AI – not just data

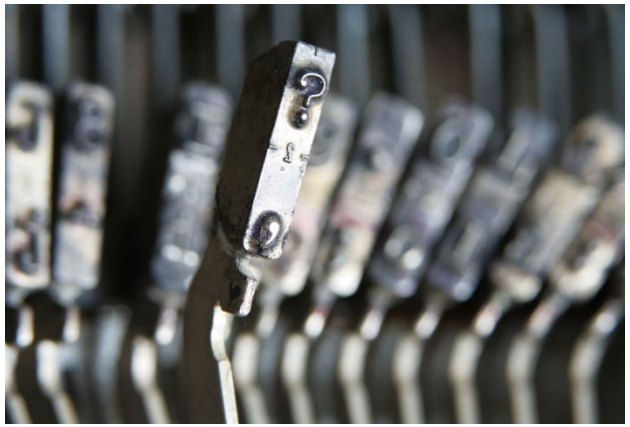


Science-guided AI



- 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

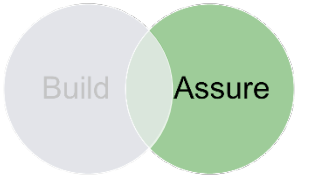
Causal AI

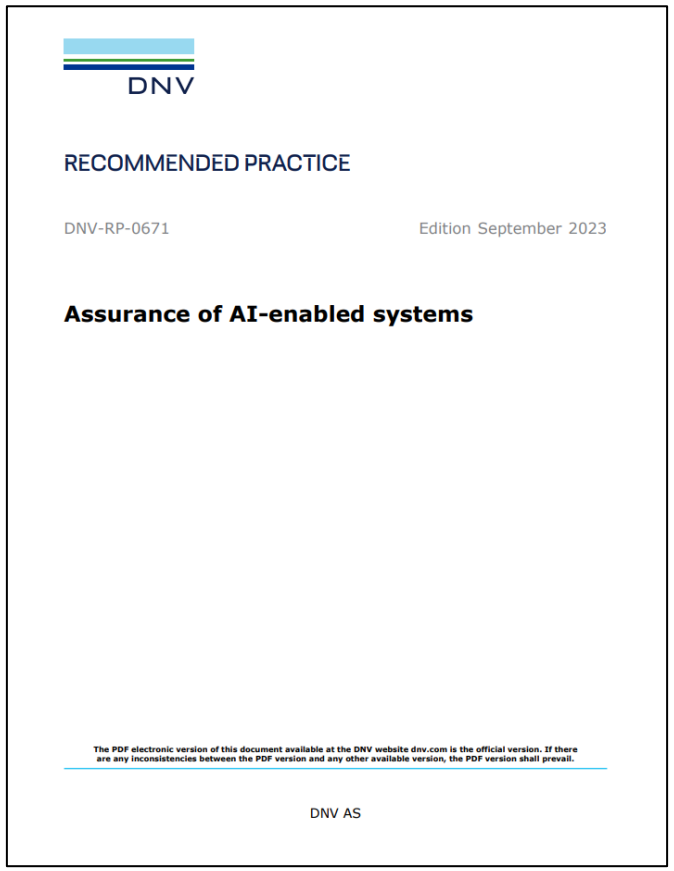
- Understands cause-and-effect
- Distinguishing correlations and causation
- Distinguishing observation from intervention

DNV

Assure using proper approaches

A combination of technology qualification, risk assessment and assurance case





DNV

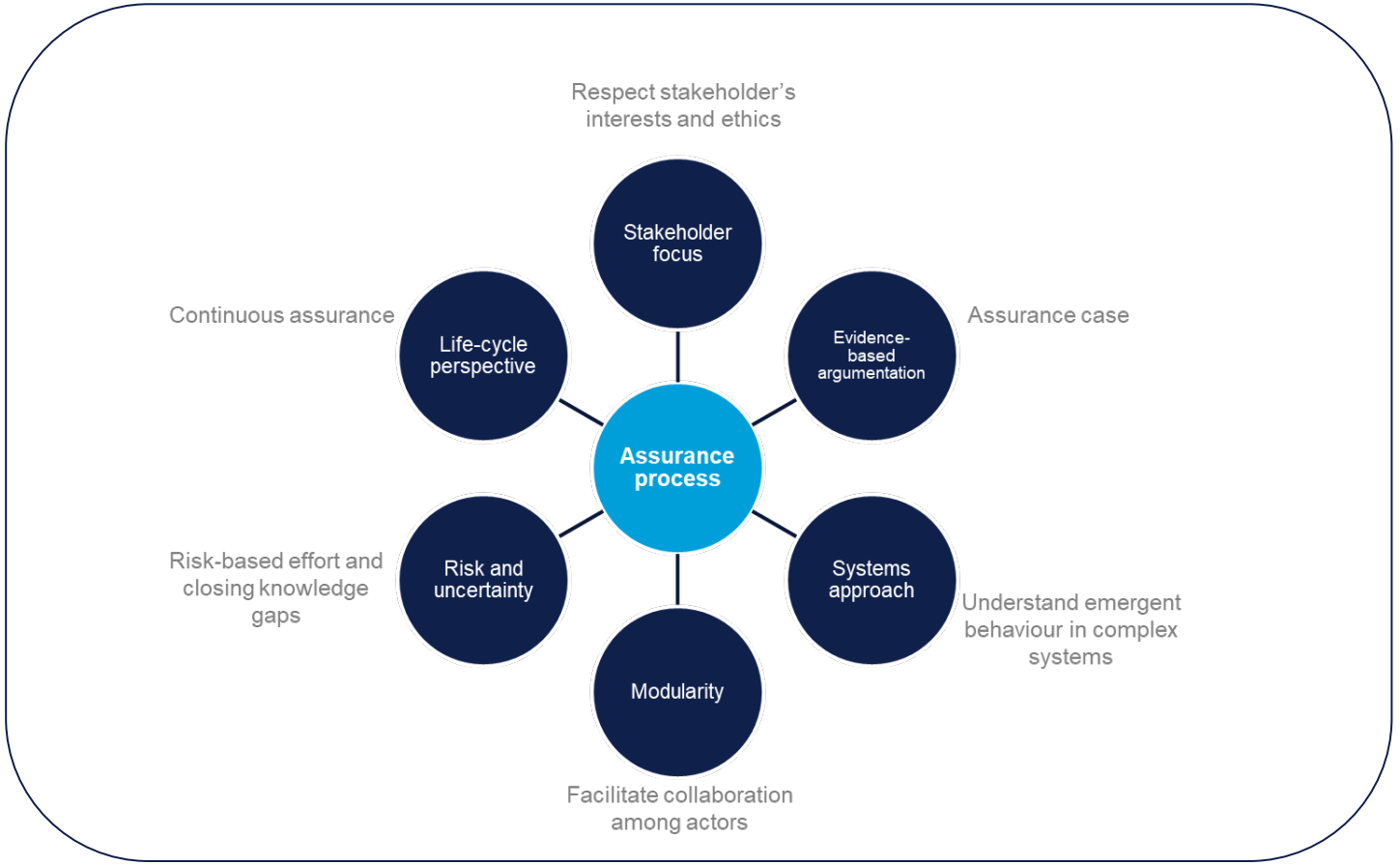
RECOMMENDED PRACTICE

DNV-RP-0671 Edition September 2023

Assurance of AI-enabled systems

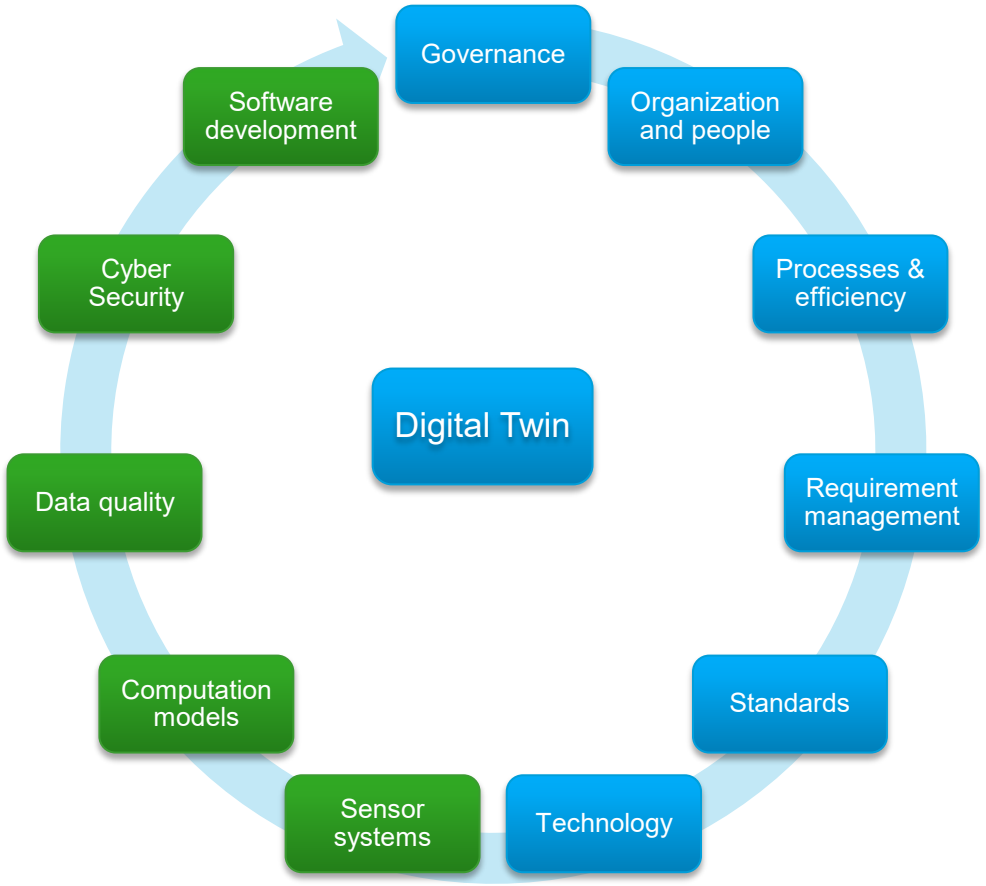
The PDF electronic version of this document available at the DNV website dnv.com is the official version. If there are any inconsistencies between the PDF version and any other available version, the PDF version shall prevail.

DNV AS



DNV

A mature organisation is the foundation for success



the degree to which an organization has processes, procedures and capabilities in place addressing the organizational aspects of developing, maintaining and operating a digital asset

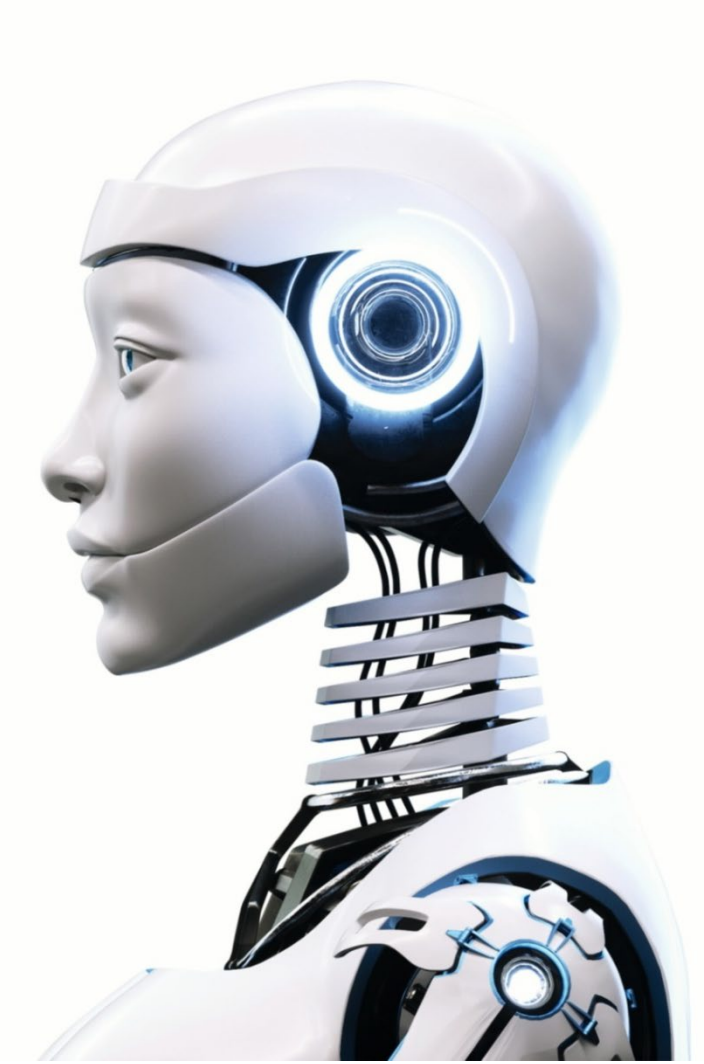
Increasing maturity

5 - Optimized	• Process management includes deliberate process optimization / improvement
4 - Managed	• The process is quantitatively managed in accordance with agreed-upon metrics
3 - Defined	• The process is well defined and used as a standard business process
2 - Repeatable	• The process is documented sufficiently such that repeating the same steps is attempted
1 - Initial	• Ad hoc, individual initiatives – The starting point for a new or undocumented repeating process

DNV

Oppsummering

- Ta menneske-sentrert design på alvor under teknologiutvikling
  - Operatøren som sikkerhetsbarriere → forutsetter riktig situasjons-forståelse
  - Spesifiser hvordan feiltilstander skal detekteres og formidles til operatøren
  - Inkluder spesialister på området
- Integrer kvalifisering og verifikasjon i utviklingsprosessen
  - Planlegg hvordan “påseplikten” skal ivaretas
  - Etabler gode konsept-modeller som alle interessehavere forstår
  - Unngå kvalifisering av komplekse digital modeller som en «black-box»
- Organisasjons-modenhet
  - Høy endringstakt stiller krav til organisasjonsmodenhet
  - Moden organisasjon – en forutsetning kvalitetssikring og vedlikehold av komplekse systemer
  - Gjelder både operatører og leverandører



Takk for oppmerksomheten

Spørsmål?

Christian.Markussen@dnv.com
+47 94831307

www.dnv.com