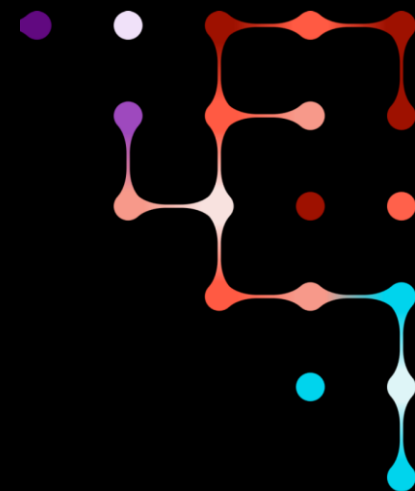




TRUST

The Norwegian Centre for Trustworthy AI



Tillit til KI: Juridiske perspektiver

Malcolm Langford

Professor, det juridiske fakultet, UiO

Nestleder, TRUST: The Norwegian Centre for Trustworthy AI



UNIVERSITY
OF OSLO



Funded by
The Research
Council of Norway







National Research Partners

University of Oslo
SINTEF
Norwegian Computing Centre (NR)
Akershus University Hospital
BI Norwegian Business School
CICERO
Institute of Transport Economics (TØI)
Norwegian Geotechnical Institute (NGI)
NORCE Norwegian Research Centre
Norwegian Institute of Public Affairs (NUPI)
Norwegian Institute of Public Health (NIPH)
Norwegian Police University College
OsloMet
Oslo University Hospital
University of Bergen (UiB)
University of Stavanger (UiS)
UIT The Arctic University of Norway
Volda University College

User partners

ABB Engineering	DNB Bank
DNV	Equinor
Gjensidige	Hydro
IBM,	Kongsberg
Møller Mobility	Oslo Pension Ins, Ruter
Transport	Sparebanken Møre
Statkraft Energy	Data Protection Authority
DigDir	Equality Ombudsperson
Court Administration	Immigration Directorate
National Archives	Tax Directorate
National Archives	National Police
NKOM	Oslo City
Øygarden city	Public Roads
Railway Directorate	Trondheim City
Rana City	Volda City
SE Regional Health	Social Security Agency
Statistics Norway	Aleap
Amnesty	Construction City
Finance Union	NITO Union
Negotia Union	NIFRO Space
OSC	Nuclear Decommission
Standard Norway	StartupLab

International Research Partners

University of California (Berkeley) (USA)	Vector & SRI (Toronto, Canada)
Cambridge University (UK)	IVADO/Mila (Montreal, Canada)
University College London (UK)	Alberta Machine Intelligence Int (Canada)
Singapore Management University	Acceleration Consortium (Canada)
Linköping University (Sweden)	Ministry of Health (Rwanda)
ECTIL (Austria)	CIFAR (Canada)

Research Areas

RA1 Veridical AI
RA2 Responsible uncertainty quantification
RA3 Attribution
RA4 Causal AI
RA5 Representation and alignment
RA6 Foundation models
RA7 Generative AI
RA8 Emergent behaviours
RA9 Secure AI
RA10 Digital twins
RA11 AI engineering
RA12 AI-driven digitalisation
RA13 AI governance
RA14 AI for flourishing

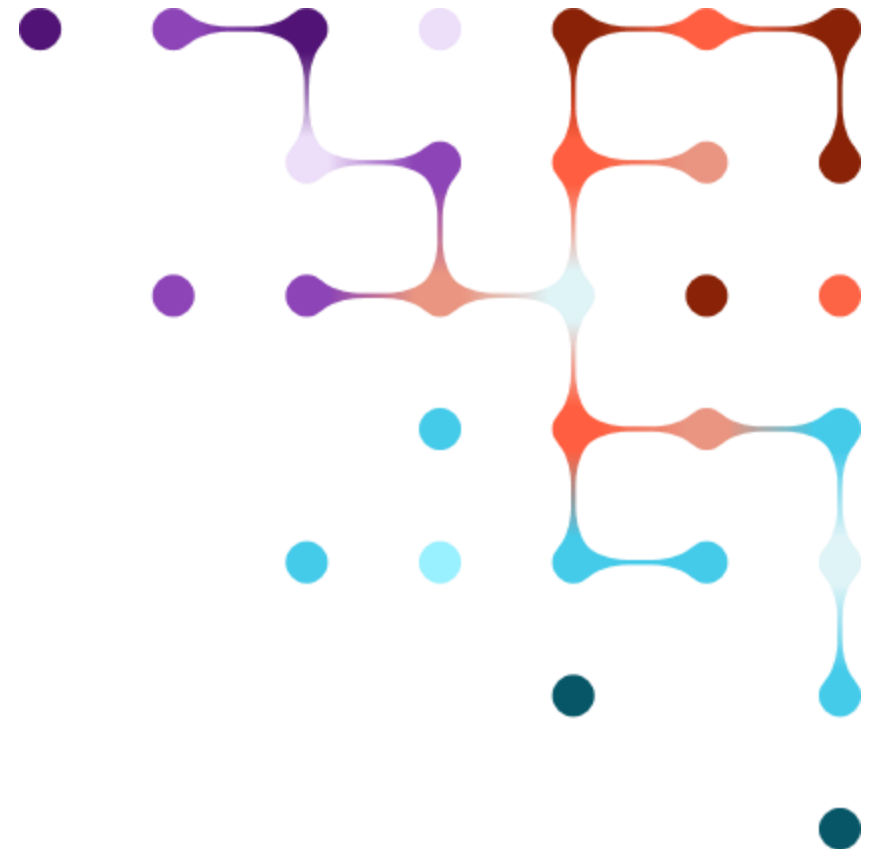
Action Clusters

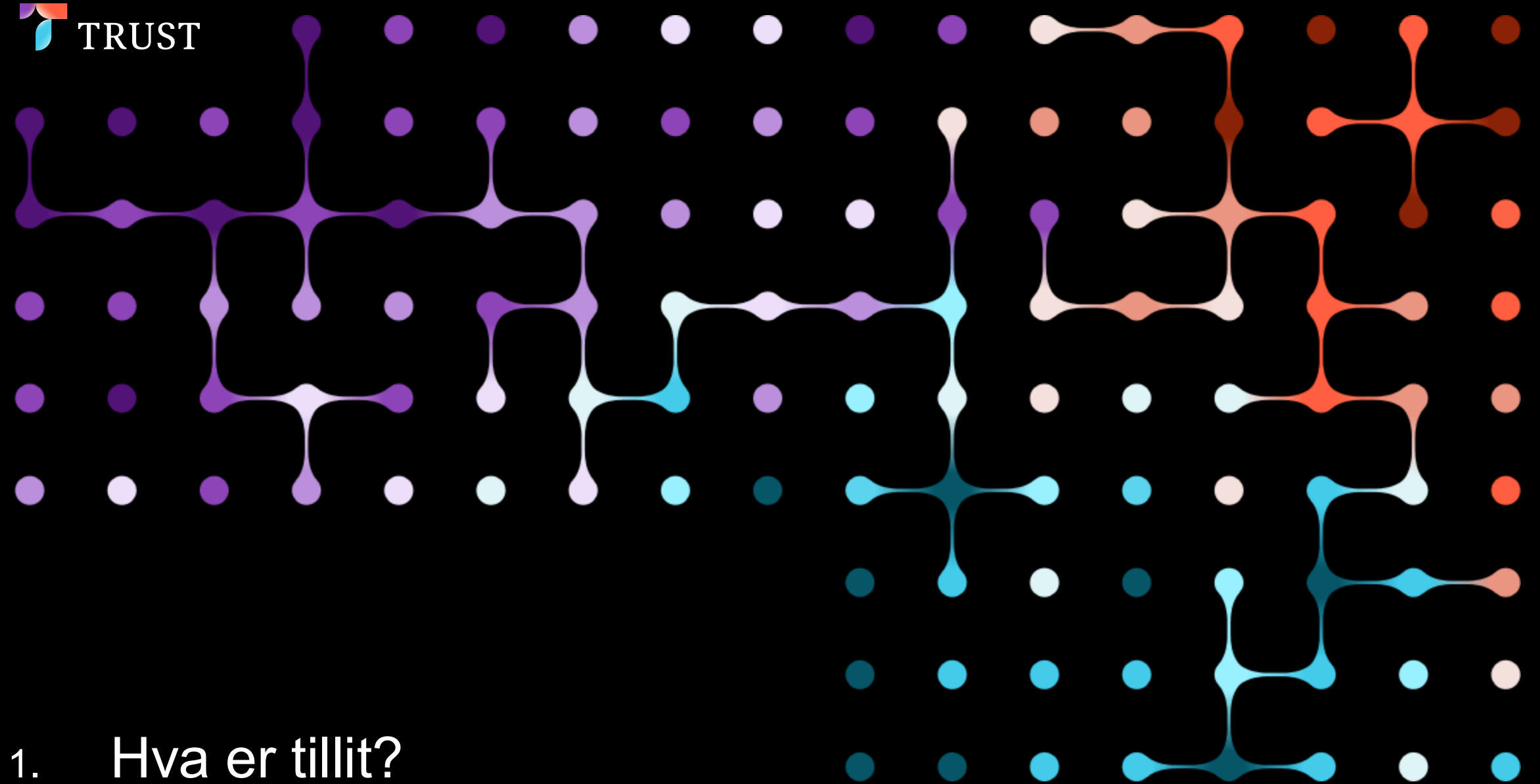
AC1 Healthcare
AC2 Mobility
AC3 Case processing
AC4 Infrastructure
AC5 Armed conflict
AC6 Criminality
AC7 Climate
AC8 Governance
AC9 Assurance
AC10 Anomalies
AC11 Decision making
AC12 Security and preparedness
AC13 Green transition
AC14 Education
AC15 Sciences



Oversikt

1. Hva er tillit?
2. Hva er tillit til KI?
3. Hvordan kan vi forsikre tillit?
4. Hva er jussens rolle?
5. Jus og KI
6. Vurdering av dagens juss



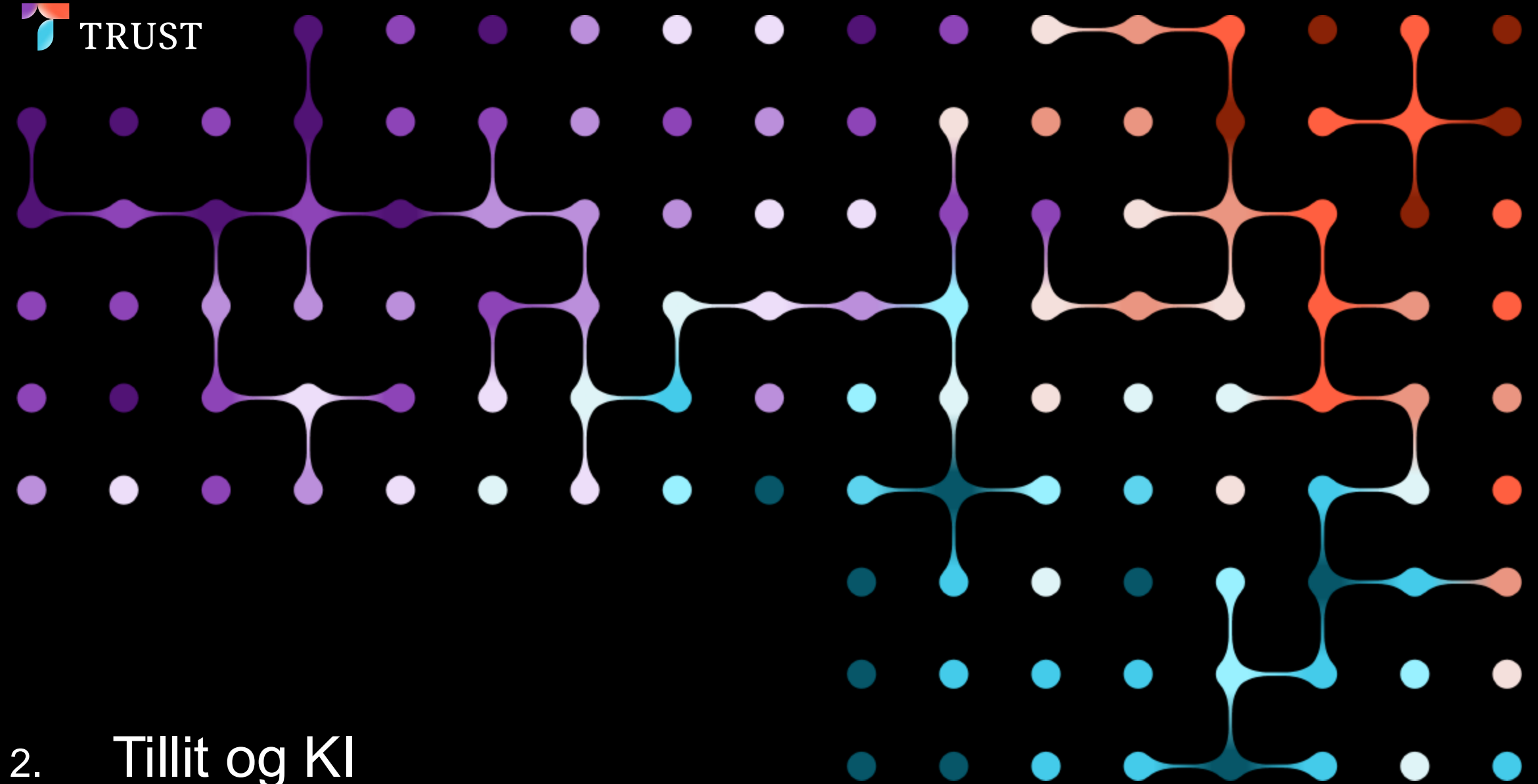


1. Hva er tillit?



- Konsepter:
 - *Trustworthy* – Normativ konsept
 - *Trust* – Empirisk konsept
- Risiko: Feilkalibrering
 - *Overtrust*: $\text{Trust} > \text{Trustworthiness}$ (kostnader)
 - *Undertrust*: $\text{Trust} < \text{Trustworthiness}$ (tapte fordeler)
- Mål: kalibrert tillit
 - ↑ *Trustworthiness*
 - &
 - $\text{trust} = \text{trustworthiness}$





2. Tillit og KI



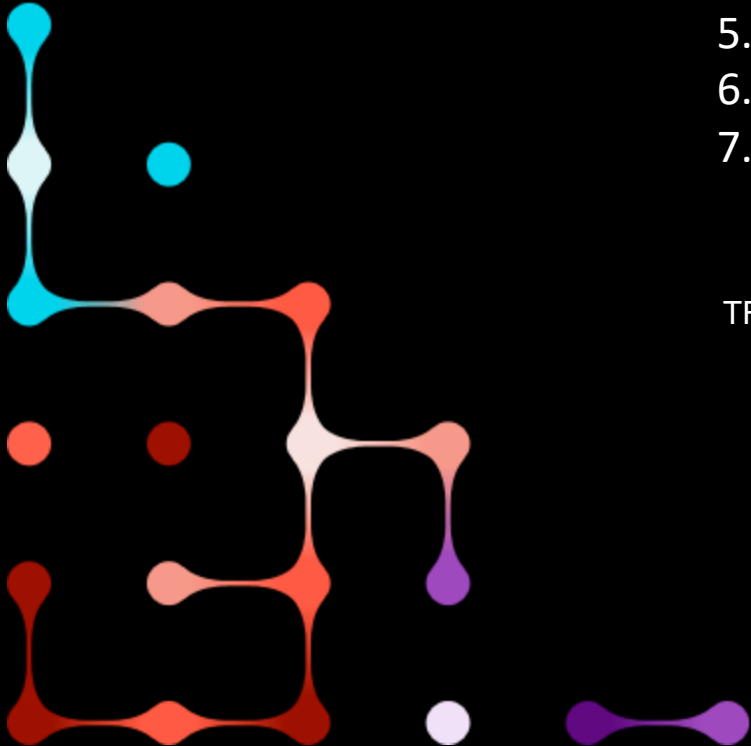
Hva er tillit til KI?

- Fokuseret og samsvarende
 - f.eks. KAIRI framework: sustainability, accuracy, fairness, and explainability. (Giudici et. al. 2024)
- Helhetlig
 - EU HLEG (2019): (1) human agency and oversight; (2) technical robustness and safety; (3) privacy and data governance (4) transparency, including traceability, explainability and communication; (5) diversity and stakeholder participation, non-discrimination and fairness; (6) Societal and environmental wellbeing; and (7) Accountability.

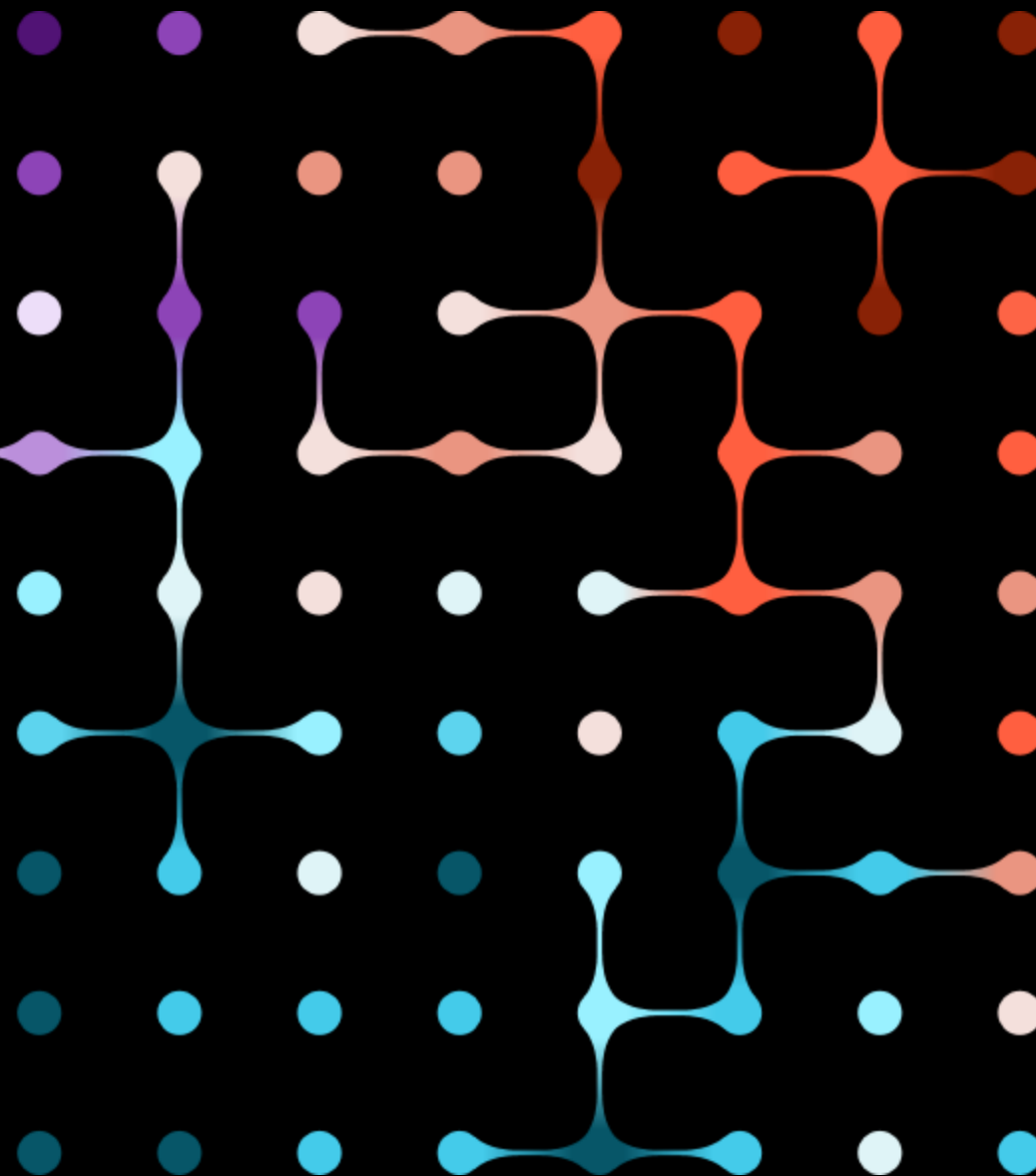


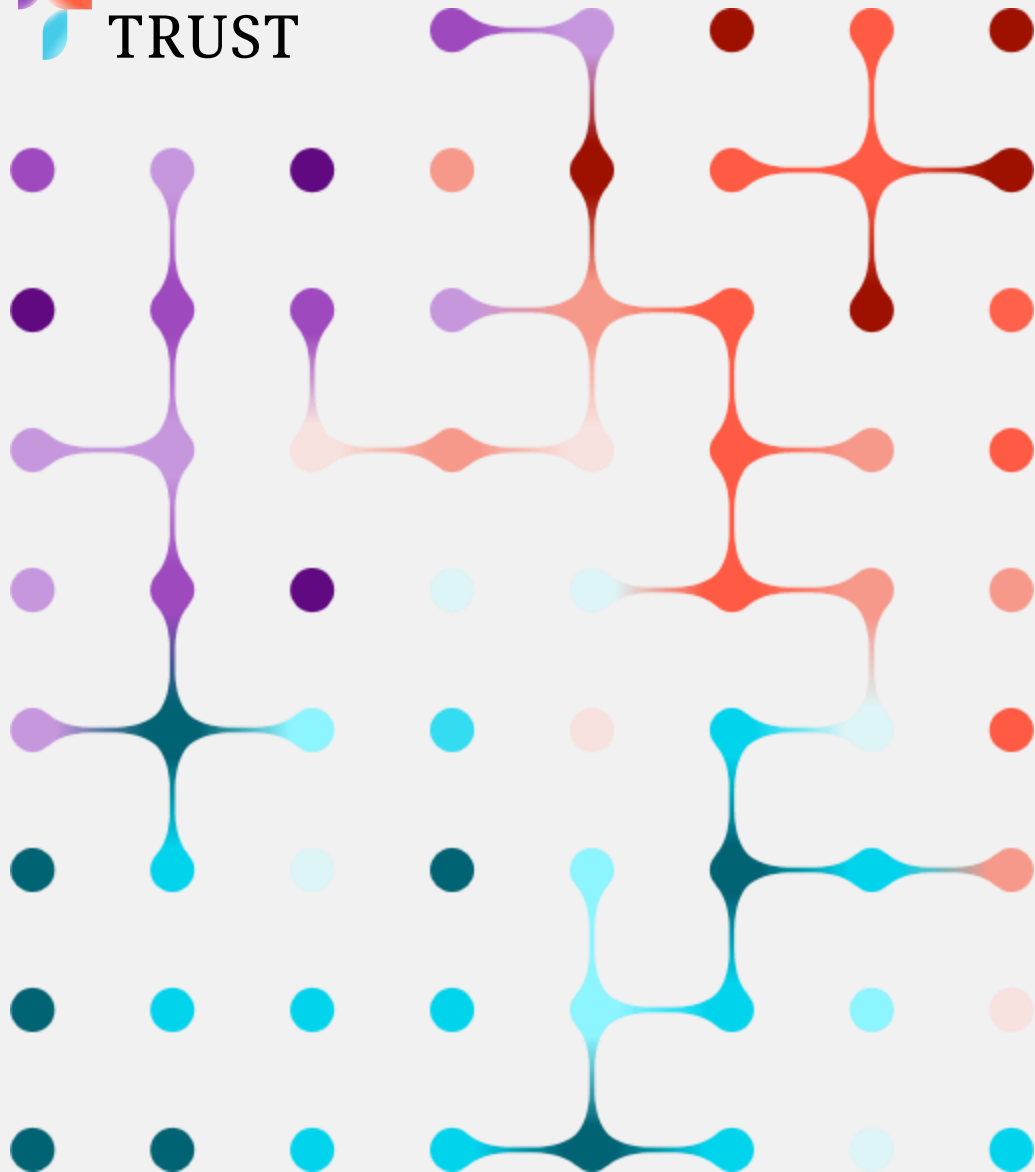
- 1. Accurate
- 2. Interpretable
- 3. Aligned and inclusive
- 4. Safe
- 5. Sustainable
- 6. Democratic and well-governed and
- 7. With the capacity to reach all this at scale

TRUST KI-sentres definsjon



3. Hvordan kan vi forsikre
tillit?

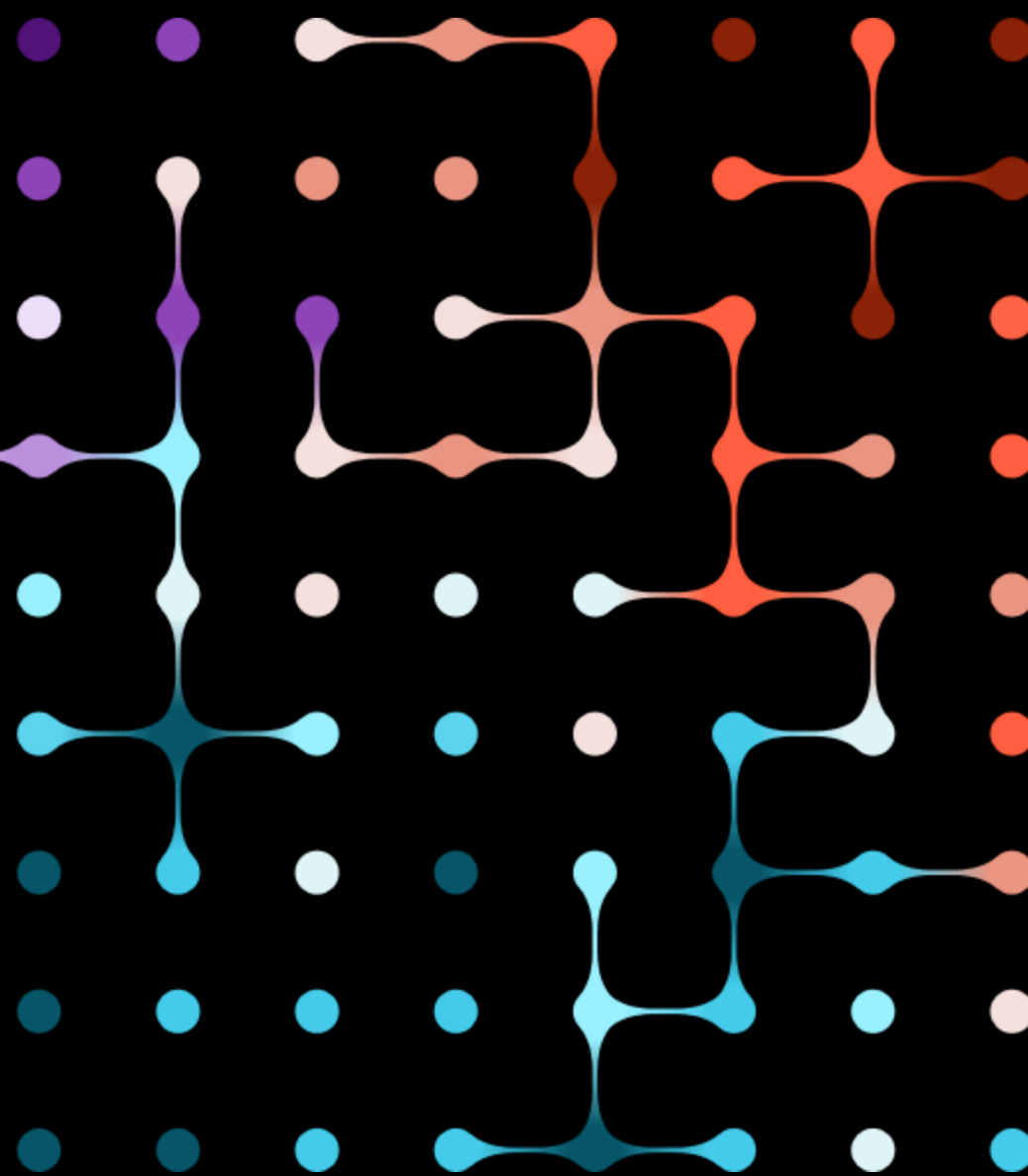


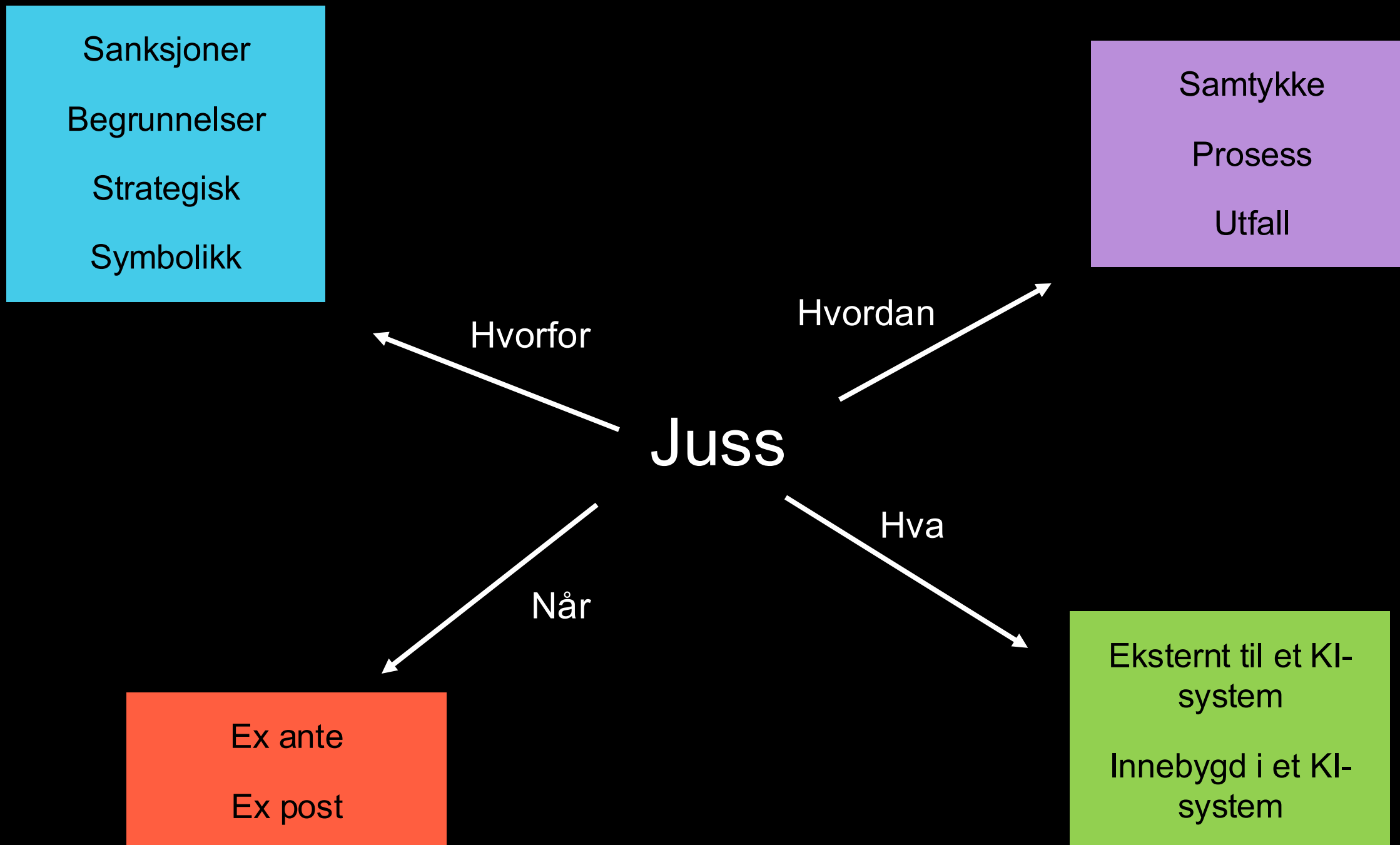


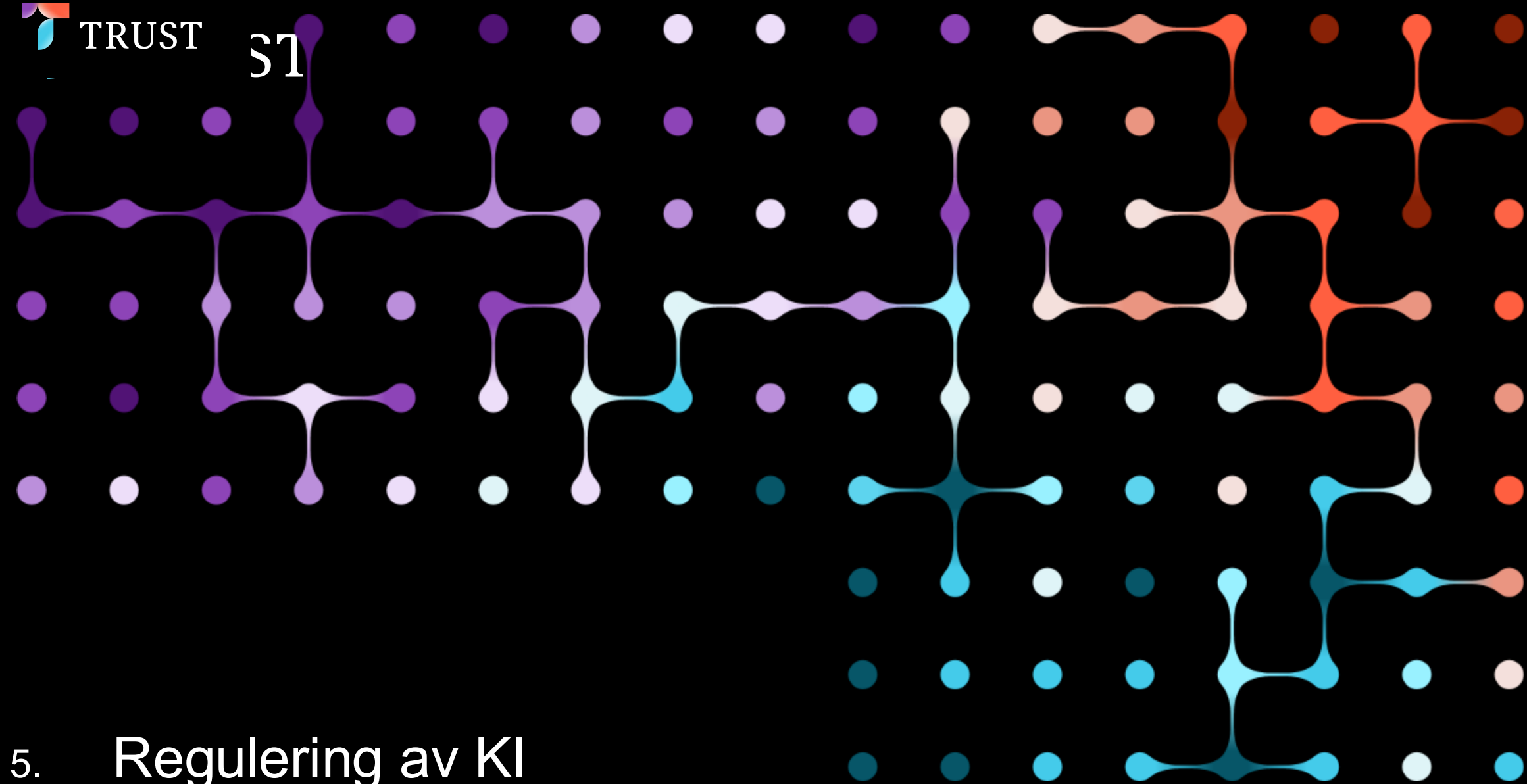
3. Hvordan kan vi forsikre tillit?

- Moralitet
- Markedsmekanismer
- Demokratiske prosesser
- Omdømme
- Jussen

4. Hva er rollen til jussen?







5. Regulerig av KI

Regler/Standarder

Prosesser

Institusjoner

KI-spesifikk

Alminnelige

Domene spesifikk

Hvordan

Hva

Regulering

Hvor

Når

Soft law

Hard law

ISO/SN standarder

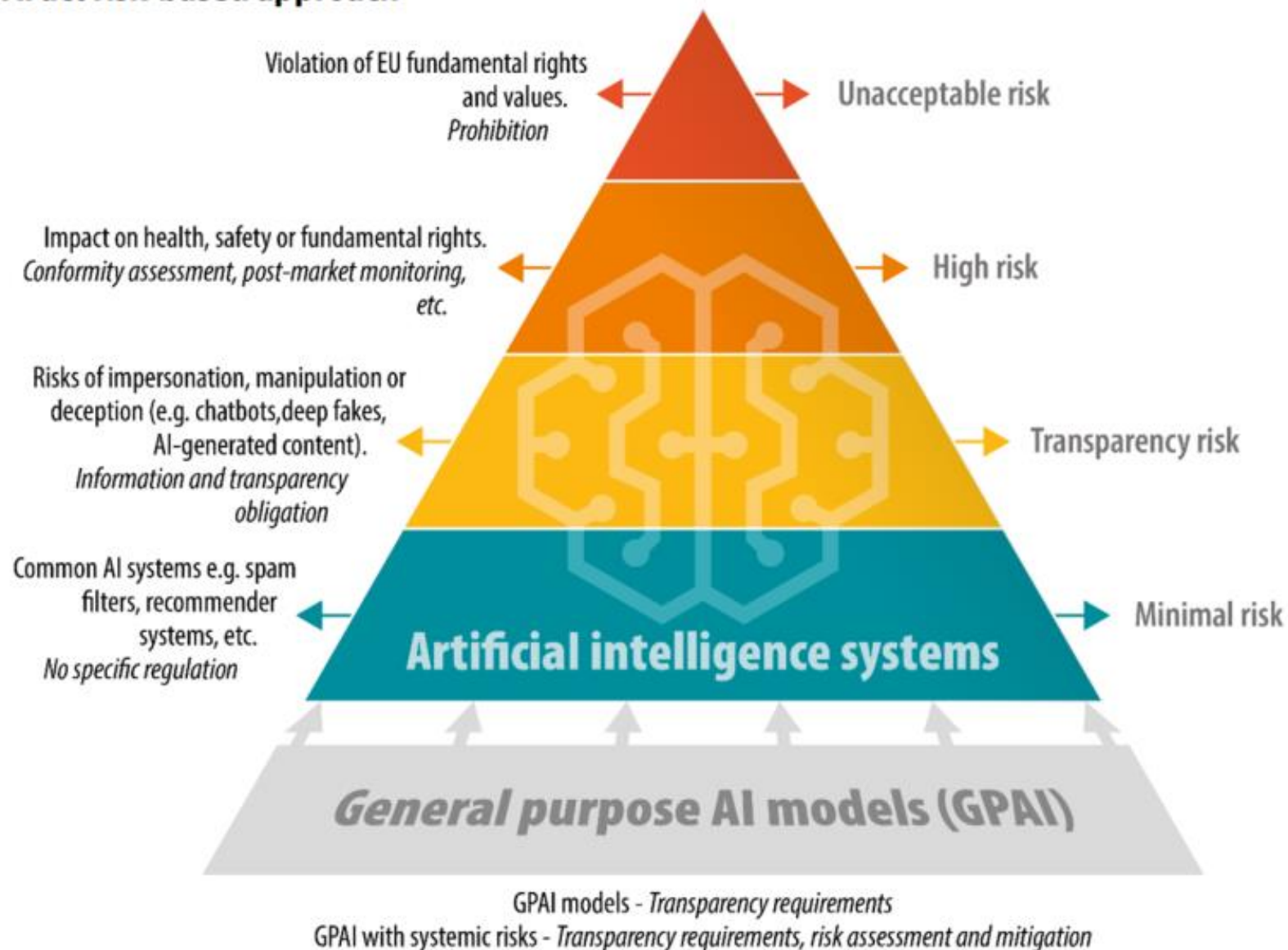
Internasjonalt

Regionalt

Nasjonalt

Lokalt

EU AI act risk-based approach



Data source: [European Commission](https://european-council.europa.eu/media/e3004c0d-3254-476a-990d-699c0613389d/attachment_data/data/eu-ai-act-risk-based-approach.pdf)



Artificial Intelligence

HomeNewsFramework ConventionCommittee on Artificial Intelligence (CAI)Resources ▾

You are here: Artificial Intelligence > Framework Convention

The Framework Convention on Artificial Intelligence

The Council of Europe [Framework Convention on Artificial Intelligence](#) and human rights, democracy and the rule of law is the **first-ever international legally binding treaty** in this field. It aims to ensure that activities within the lifecycle of artificial intelligence systems are fully consistent with human rights, democracy and the rule of law, while being conducive to technological progress and innovation.

 [Framework Convention](#) and its [Explanatory Report](#)

The Framework Convention complements existing international standards on human rights, democracy and the rule of law, and **aims to fill any legal gaps that may result from rapid technological advances**. In order to stand the test of time, the Framework Convention does not regulate technology and is essentially **technology-neutral**.

 [Download the Brochure](#)



How was it elaborated?



What does it require states to do?



Who is covered?

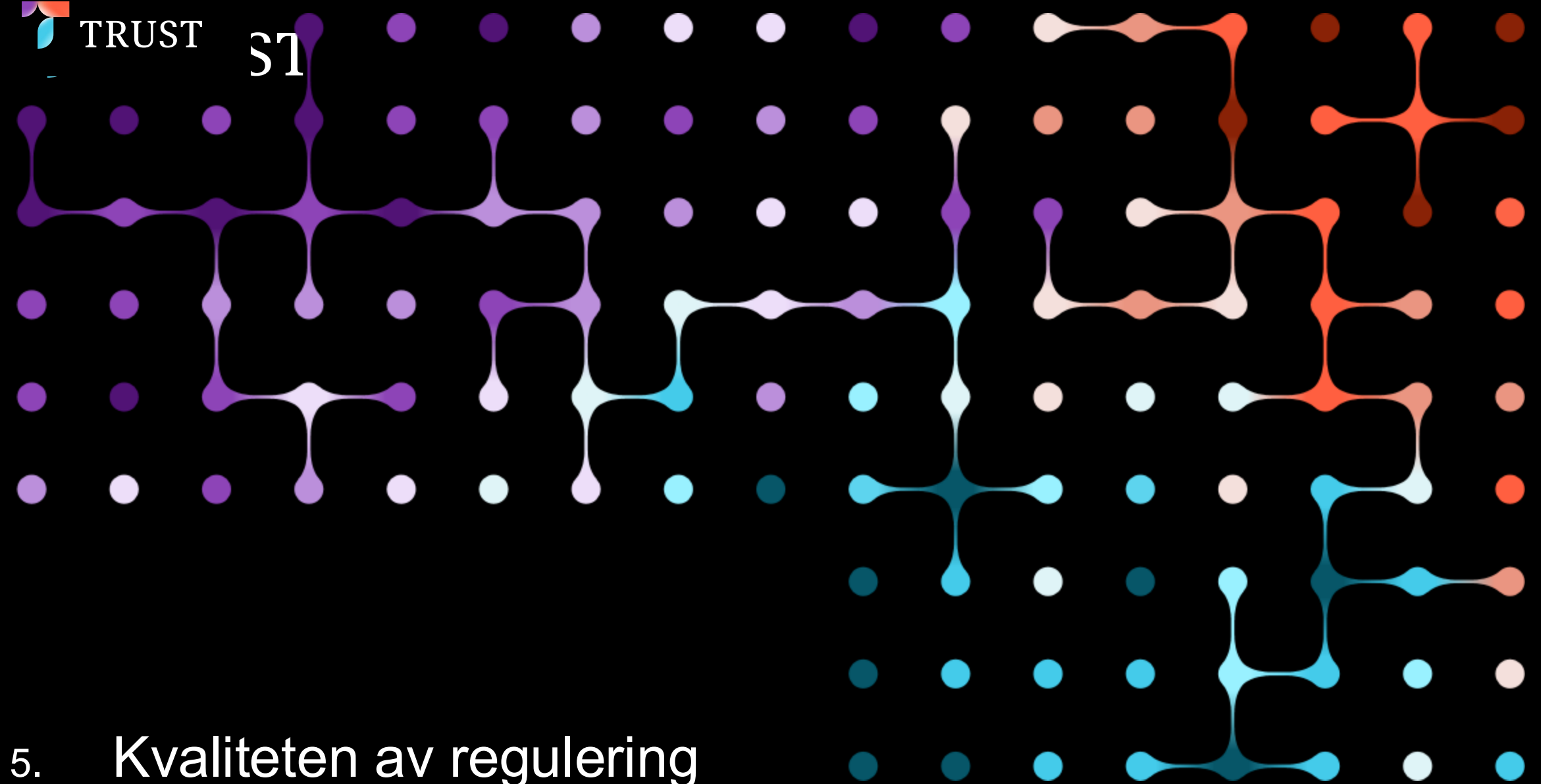


How is the implementation monitored?

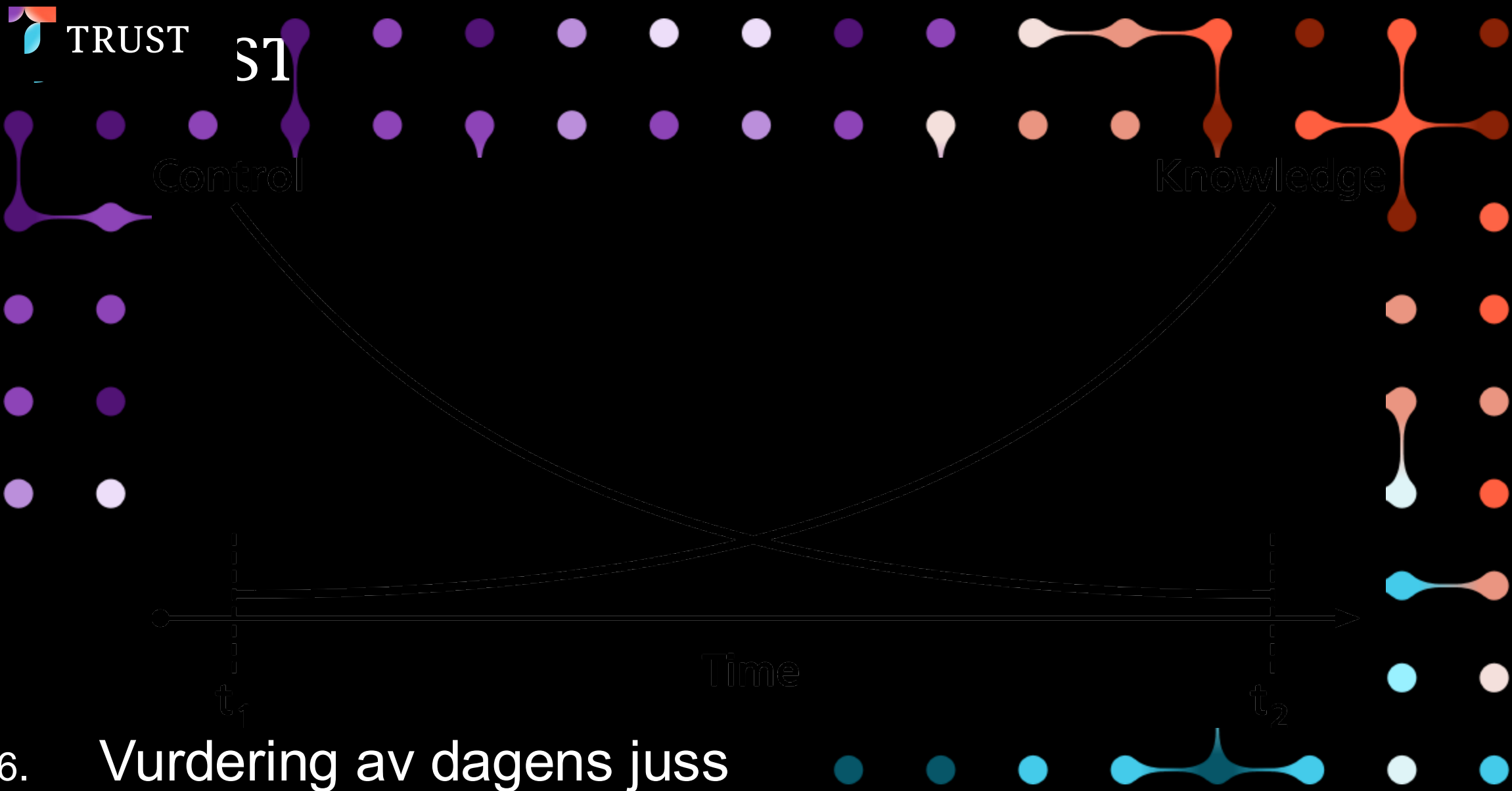


TRUST

§1



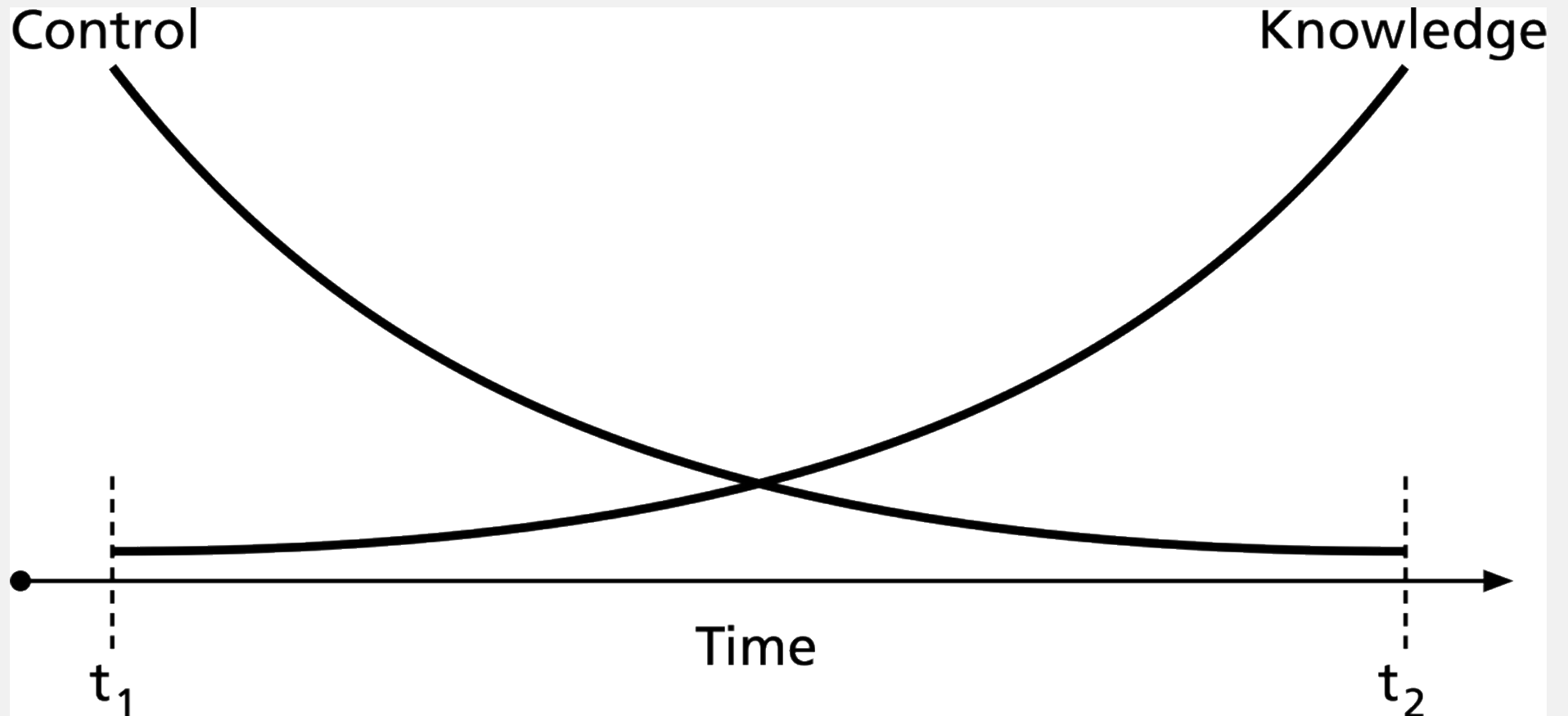
5. Kvaliteten av regulering





TRUST

Collingridge Dilemma





Governing AI

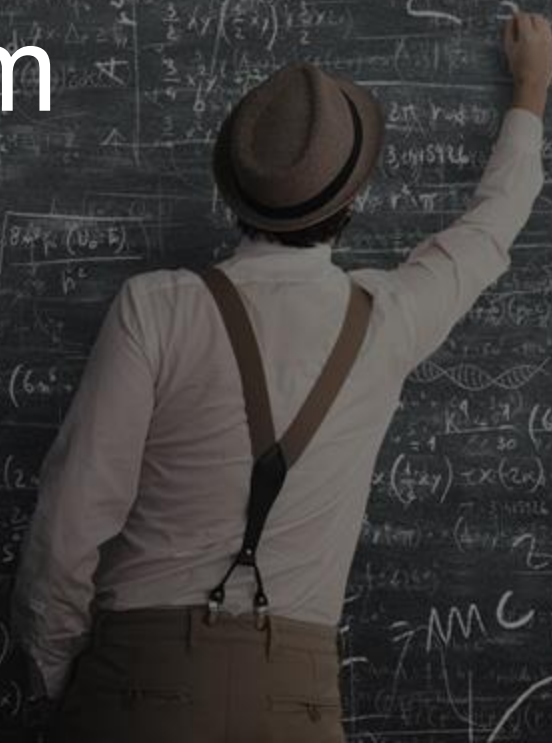
- Vertical
- Horizontal
- Co-regulation





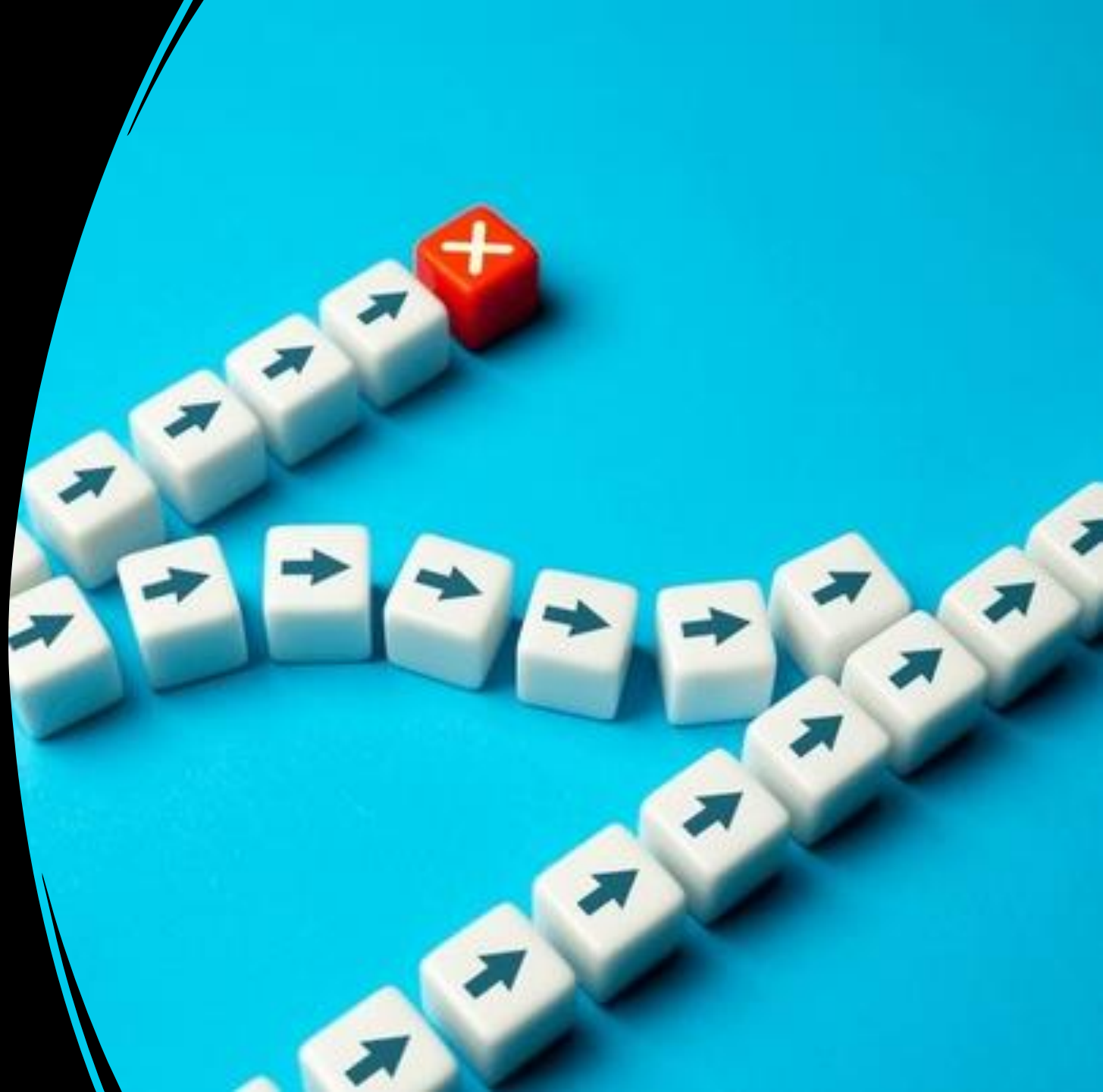
TRUST

Long problem



TRUST Anticipatory governance

1. Foresight
2. Experimenting (diagonal)
3. Integrating





malcolm.langford@jus.uio.no

Takk for meg