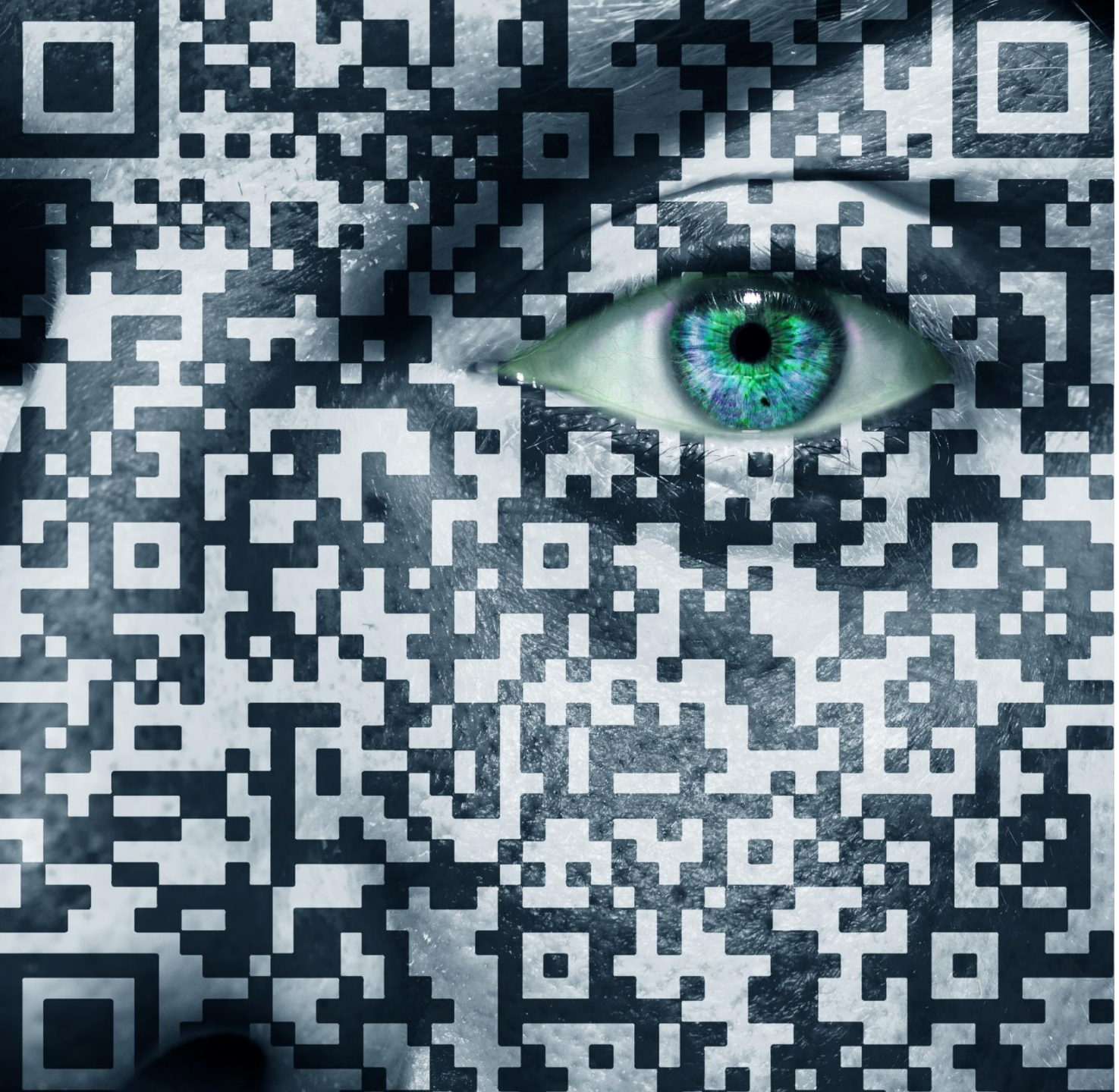


Oppfølging av KI og Human Oversight

Linn I. V. Bergh (PhD), Senior rådgiver

HFC-forum 15-16.10.24

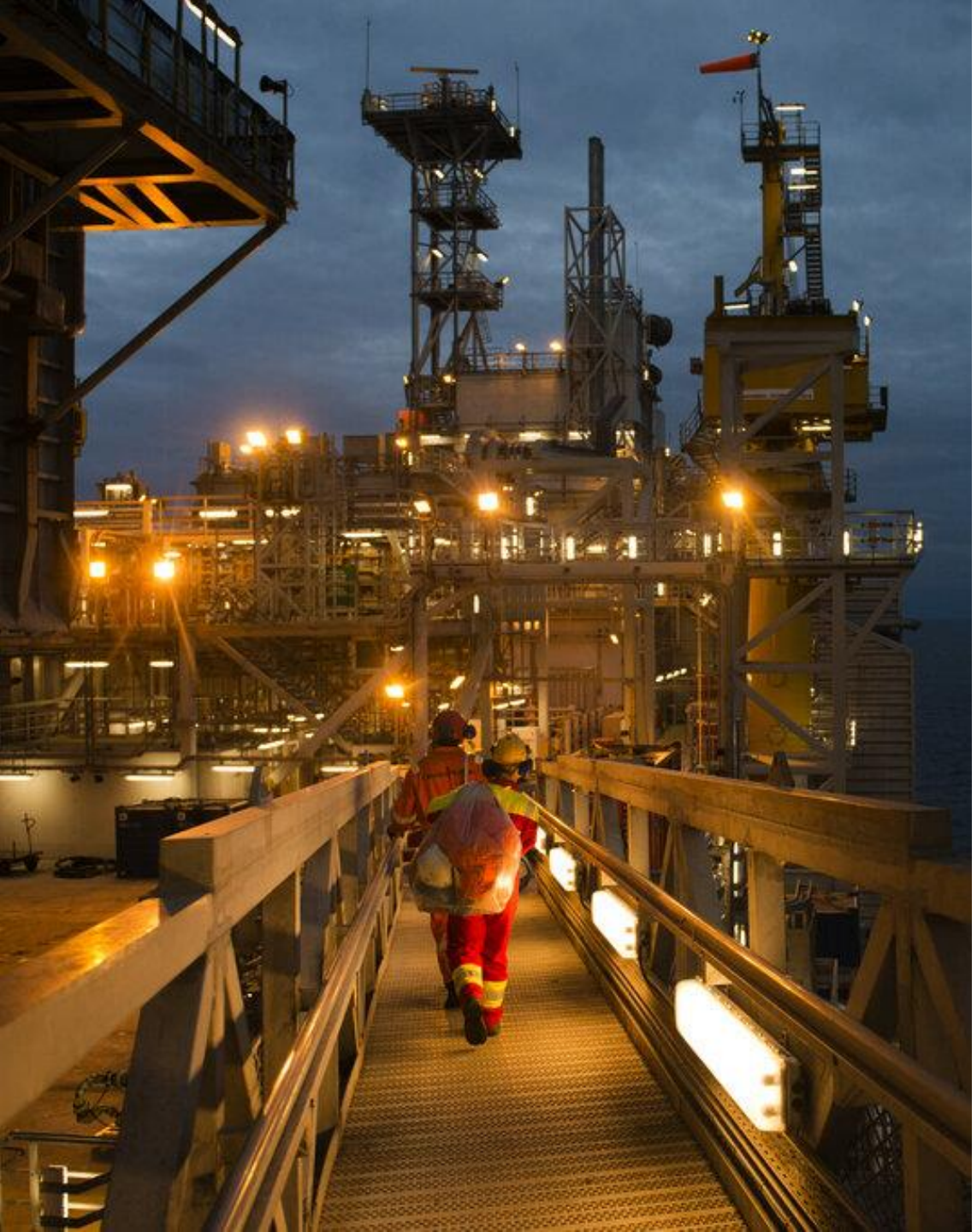
Havtil
Havindustritilsynet

Jeg vil si litt om….

- Våre aktiviteter
- KI forordning
- HMS-regelverket og KI
- Veien videre

Havtil
Havindustritilsynet

Havtil
Havindustritilsynet

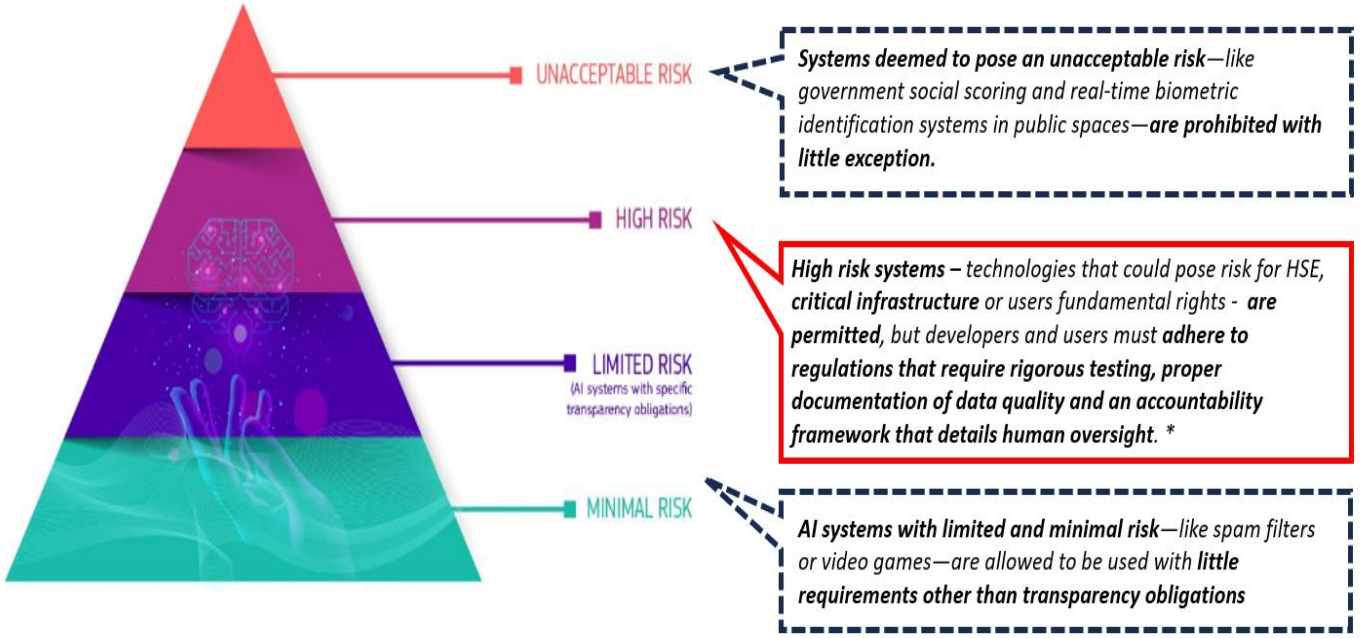


Havtils rolle blir å følge opp selskapenes egen styring når det gjelder utvikling og anvendelse av KI løsninger.

Havtil
Havindustritilsynet

KI-forordningen

- Felleseuropeisk regelverk
- Krav til kvaliteten på systemet i et livssykklusperspektiv
- Risikobasert tilnærming
- Samspill med eksisterende regelverk
- Standarder får stor betydning



Havtil
Havindustritilsynet

Krav til høy-risiko systemer

- Risikostyring
- Datakvalitet og datahåndtering
- Teknisk dokumentasjon
- Logging og sporbarhet
- Transparens og forklarbarhet
- Menneskelig kontroll
- Tekniske robusthet og cybersikkerhet

(Chapter III, Section 2, Article 9-15)

Havtil
Havindustritilsynet

Article 14: Human Oversight

"..high-risk AI systems must be designed in a way that allows humans to effectively oversee them. The goal of human oversight is to prevent or minimize risks to health, safety, or fundamental rights that may arise from using these systems. The oversight measures should match the risks and context of the AI system's use. These measures could be built into the system by the provider or implemented by the user. The AI system should be provided in a way that allows the overseer to understand its capabilities and limitations, detect and address issues, avoid over-reliance on the system, interpret its output, decide not to use it, or stop its operation.."

EU Artificial Intelligence Act, The AI Act Explorer

Havtil
Havindustritilsynet

Human Oversight - *noen* sentrale forhold

Tekniske forhold

- Design og vedlikehold av KI løsning
  - Funksjonsallokering
  - Menneskelig kontroll
  - Forklarbarhet og transparens
  - Brukertesting
  - Logging under operasjoner

Organisatoriske forhold

- Integrasjon i styringssystemet
- Klare ansvarsforhold og tilstrekkelig ressurser
- Kompetanse
- *Styring av risiko* knyttet til:
  - Grad av automasjon og betydning for menneske - KI samhandling
  - Kognitiv bias hos sluttbruker

Havtil
Havindustritilsynet

*DALL-E 3, Open AI

KI-system og menneskelig kontroll, noen eksempler

Operatør skal kunne ta over kontroll ved behov, også når flere KI komponenter er integrert i et system

KI skal kunne overlate kontroll til operatør ved behov (systemet selv gjenkjenne når den ikke fungerer optimalt)

Operatør skal kunne monitorere effektivt i sanntid

Gjennomføre analyser i etterkant

Havtil
Havindustritilsynet

*DALL-E 3, Open AI

Transparente og forståelig KI

- Systemet må gi informasjon om hva automatikken gjør, for økt forståelse av prosessen og for å kunne forutse fremtidige handlinger.

- Beslutningstaker må kunne:
  - Forstå hvorfor
  - Forstå hvorfor ikke
  - Vite når man lykkes eller feiler
  - Vite når man kan stole på
  - Vite når man må avslutte

Mica Endsely, 2023

Havtil
Havindustritilsynet

Forståelig for hvem?

- Eksempler på roller som kan ha behov for samhandle med KI systemet i et livssykklusperspektiv:
  - Ledere
  - Designere
  - Operatør i spisse enden
- Brukernes unike perspektiv / arbeidssituasjon er viktig.
- Design for *transparens og forklarbarhet* - noen viktige forhold:
  - valg av datamodell
  - konteksten der løsningen skal brukes
  - type oppgave som skal løses

Havtil
Havindustritilsynet

HMS-regelverket

01

HMS-regelverket i petroleumsvirksomheten er funksjonsbasert, teknologinøytralt og bygger på risikostyring.

02

Regelverket inneholder relevante grunnkrav til forsvarlig virksomhet, risikovurdering og risikostyring som er viktig for ansvarlig og pålitelig utvikling og bruk av KI løsninger.

03

Vår vurderinger er at regelverket er relativt godt anvendbare med tanke på oppfølging av KI-løsninger slik det er i dag, men at det nok mangler henvisninger til normer og standarder som kan gi tilstrekkelig veiledning ved bruk av KI.

Havtil
Havindustritilsynet

Styringsforskriften §11 om beslutningsgrunnlag og beslutningskriterier

- Black-box problemstillinger kan skape utfordringer når det gjelder det å sikre transparente og dokumenterbare prosesser. En potensiell risiko er at man stoler for mye på og ikke har tilstrekkelig oversikt og innsyn i KI-systemets beslutningsprosesser og output.
- Krav til at forutsetninger som legges til grunn for en beslutning, skal uttrykkes slik at de kan følges opp. Det er viktig å synliggjøre hvilke betingelser, forutsettinger og avgrensninger som er lagt til grunn for en beslutning (jf. SF§16).

SF §11 - "Før det treffes beslutninger skal den ansvarlige sikre at problemstillinger som angår helse, miljø og sikkerhet, er allsidig og tilstrekkelig belyst.»

Havtil
Havindustritilsynet

Innretningsforskriften §9 om kvalifisering og bruk av ny teknologi og nye metoder

- Det er en utfordring at man ikke nødvendigvis har vell-etablerte test- og kvalifiseringsmetoder spesifikt for KI-systemer. Dette kan medføre at løsninger ikke i tilstrekkelig grad tar hensyn til brukerens behov. God teknologiutvikling er mer enn bare teknologi, det handler også om mennesker.

- IF§9 stiller krav til hvordan utvikling, testing og bruk av teknologier skal foregå for å ivareta HMS-regelverket. Det kreves at kriteriene for slike aktiviteter skal være representative for de aktuelle bruksforholdene. Kravene til teknologiutvikling vil være relevante for ny teknologi som KI.

IF§9 - "Kvalifiseringen eller prøvingen skal demonstrere at gjeldende krav kan oppfylles ved bruk av den aktuelle nye teknologien eller metodene"

Havtil
Havindustritilsynet

Innretningsforskriften §21 om menneske-maskin-grensesnitt og informasjonspresentasjon

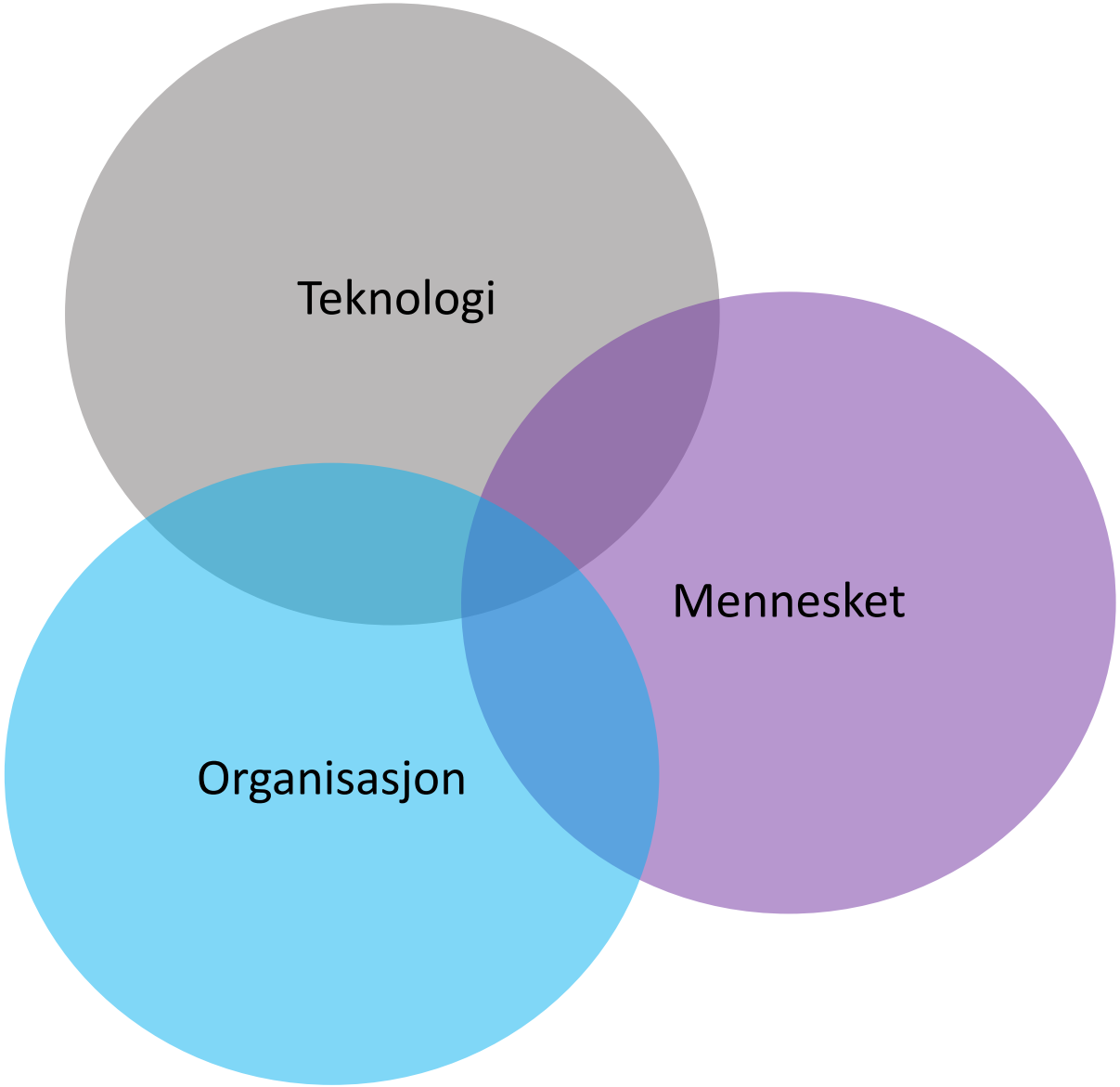
- Selskapene bruker KI-løsninger som beslutningsstøtte og mennesker har fortsatt en «hånd på rattet». En risiko er da er at oppmerksomheten til mennesket som overvåker svekkes eller at den "sannheten" som KI-systemet presenterer, farger menneskets vurdering, uten at man i tilstrekkelig grad forstår usikkerhet.

- Krav til at teknologien presenterer korrekt informasjon til brukeren. Dette inkluderer krav til kvaliteten til presentasjonen og krav til forståeligheten av den. KI-systemet skal kunne overvåkes effektivt mens det er i bruk, og det skal kunne treffes tiltak for å kvalitetssikre systemets output.

IF§21 - "Skjermbasert utstyr og annet teknisk utstyr for å overvåke, kontrollere og styre maskiner, anlegg eller produksjonsprosesser, skal utformes slik at faren for feilhandlinger som kan ha betydning for sikkerheten, reduseres. […] Informasjonen som presenteres, skal være korrekt og lett forståelig."

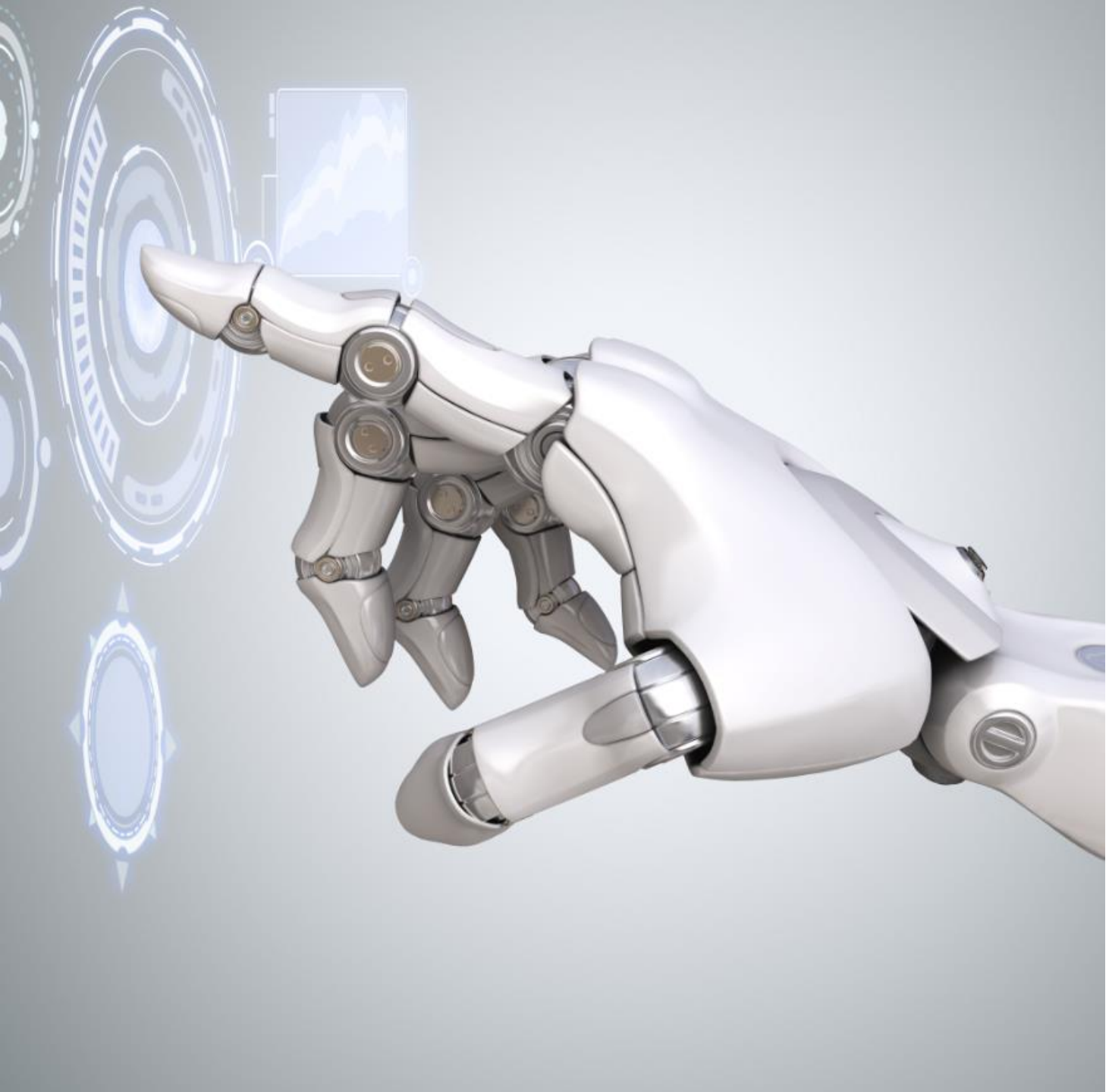
Havtil
Havindustritilsynet

MTO perspektiv og HMS-regelverket



Havtil
Havindustritilsynet

Domenekunnskap og erfaring



KI prioritert tema - Sentrale aktiviteter

Kunnskapsutvikling

- Dialog med selskapenes om deres egen risikostyring

Samarbeid og nettverksarbeid

- Forskning og industri
- Standardarbeid (nasjonalt og internasjonalt)

Tilsyn og regulering

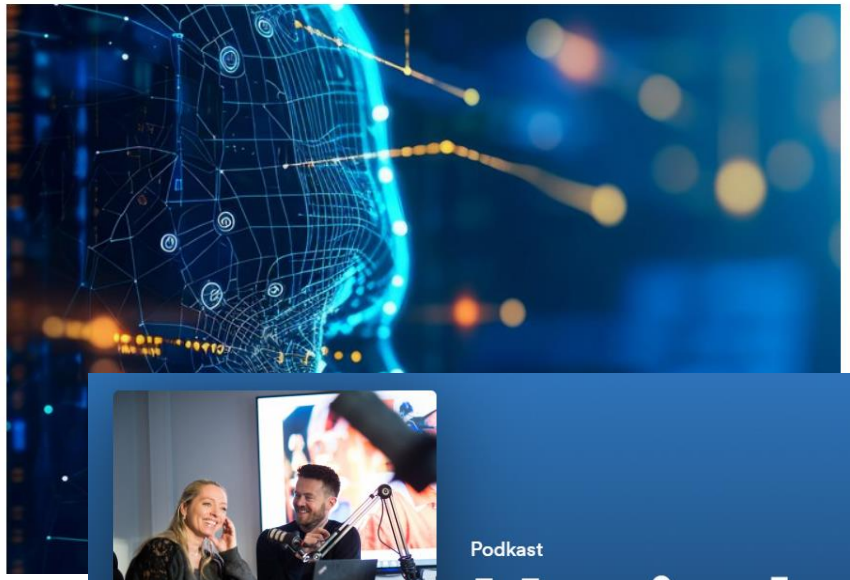
- HMS regelverket – KI forordningen
- Maskindirektivet – KI forordningen

Havtil
Havindustritilsynet

Hvordan kan vi bruke opparbeidet kunnskap i møte med økt bruk av KI i næringen?

Havtil
Havindustritilsynet

§ Regelverk → Tilsyn → English →



Tema: Kunstig intelligens →

Bruk av ku
petroleum
Havindust

Podkast

Havindustritils

Havindustritilsynet

Følg • • •

Om

En podcast fra Havindustritils

De 5 KI-paradokser med Mica Endsley

Havindustritilsynet

Hvor intelligent er kunstig intelligens? Hva må en bransje som ønsker å innføre KI-løsninger vite om teknologien, og hvordan kan vi innføre det på en ansvarlig måte? I denne episode...

13. mars · 31 min 28 sek

Innovasjonsdagen

Innovasjonsdagen 2024

Havindustritilsynet (Havtil) inviterer til en spennende dag om forsvarlig bruk av KI under årets Innovasjonsdag 2024.

Dato: Onsdag 6. mars 2024
Sted: Solastranden Gård Maskinhuset
Påmeldingsfrist: 5. mars 2024

Innovasjonsdagen er gratis for registrerte deltagere, men dersom du melder deg på fysisk deltakelse og ikke gir beskjed om at du ikke kan komme innen 24 timer før arrangementet, vil du bli fakturert et no-sho gebyr pålydende 1000 kr, eks. mva. Dette gjør vi for å redusere matsvinn og sørge for at flest mulig får mulighet til å delta fysisk på våre

AI safety: A regulatory perspective

Linn Iren Vestly Bergh and Kristian SolheimTeigen
Dept. of Process Safety
Norwegian Ocean Industry Authority
Stavanger, Norway
Email: Linn.iren.bergh@havtil.no

***Abstract*— Companies on the Norwegian Continental Shelf (NCS) are integrating advanced technologies to supplement and replace human-performed tasks both offshore and onshore. The Norwegian Ocean Industry Authority (HAVTIL), through ongoing follow-up, see a rising use and reliance on Artificial Intelligence (AI) for critical operations. In the coming years it is expected that the utilization of AI will increase within the petroleum sector. Based on developments in the petroleum industry, HAVTIL will over the next three to five years be devoting greater attention to AI-related safety. This paper will outline the main priorities for HAVTIL going forward.**

Keywords—artificial intelligence, petroleum industry, safety, regulations

I. Introduction

The Norwegian Industry Ocean Authority (HAVTIL) is a governmental agency overseeing safety, the working environment, emergency preparedness, and security within petroleum, renewable energy, CO_2 transport and storage, and ocean minerals. As an authority, HAVTIL's responsibility is to develop regulations, supervise companies in the industry, and communicate expertise as well as provide specialist advice to the ministry [1]. HAVTIL's goal is to follow-up that the petroleum activity gives high priority to safety, health and working environment when digital technology such as AI solutions is developed and implemented in the companies.

Companies operating on the Norwegian Continental Shelf (NCS) are adopting complex technologies to complement and replace tasks previously performed by humans—both offshore and onshore. Through various follow-up activities, HAVTIL observe an increased use of Artificial Intelligence (AI) in critical operations and applications impacting safety. It is expected that the use of AI will increase in the coming years [2]. We see and expect to see an increased use of AI in behavioural anomaly detection, automated threat and vulnerability discovery, augmentation of incident response within cyber security. We also expect to see increased application of AI in digital well planning, drilling automation and monitoring where historical and real time data is leveraged to predict well control incidents or equipment failure.

Despite high ambitions and increased use in the industry, the use of AI is currently in an early phase of testing and product development.

A. Artificial Intelligence (AI) risk in the petroleum industry

HAVTIL has a risk-based approach when following up petroleum activities. This entails directing efforts towards issues where the risk is highest, particularly with regards to major accident potential. The use of new technology such as AI, with potential challenges regarding explainability, transparency and documentation may cause such increased risk.

In the petroleum industry, each company must take ownership of and manage the risk related to the implementation of new systems and technological solutions. The companies must assess vulnerability and risk from an integrated perspective which includes human, technological and organisational (HTO) aspects. An AI system that is capable of directly influencing risk of major accidents, such as when applied in safety systems and barriers, will according to the management regulations require a risk reduction-based approach. This will also apply to the use of AI models in operations, planning, and decision support where safety is indirectly affected. This could be critical tasks and operations pertaining to working environment, emergency preparedness, and preparedness against deliberate attacks [2].

Novel solutions like AI may contribute to reduced risks, however that depends on the company's understanding and following up of inherent and contextual AI risk factors. The risks posed by AI differ from traditional industrial risk factors and conventional IT solutions. For example, machine learning relies on accurate training data, and errors or inaccuracies in the input data can lead to incorrect results and decision support to a human operator. Even small discrepancies between training data and actual data can cause errors. Furthermore, AI models may have little experience with rare or uncommon situations, leading to a lack of recognition of such conditions and output from AI systems can be challenging to interpret, and the documentation may be inadequate [3].

Technology development is not only about technology. The contexts where AI is applied is also often complex, making it difficult to detect and respond to errors when they occur. The development of AI take place within an organizational and social context. Consequently, the AI solutions are intricately linked with social structures, organizational practices, work processes, and employees' competence [4].

Emphasizing the important relationship between social and technological dimensions is also the foundation of the social technical approach originally proposed by the Tavistock Institute in the 1950's and the later research preformed in countries like Norway, UK and USA [5]. Managing and accessing vulnerability and risk therefore needs to include an integrated perspective, covering human, technology, and organizational aspects.

A significant driver of internal risk in AI systems are their complex life cycles with maintenance and follow-up needs. This

979-8-3503-4976-4/24/$31.00 ©2024 IEEE


Havtil
Havindustritilsynet