AI and Human Oversight
Focus on Critical Sectors

Béatrice Schütte

University of Lapland

Introduction

- The use of AI is increasing across all sectors

- From a broad perspective, the maritime sector is no exception

- AI-enabled features can help calculate shipping routes or the amount of combustible needed, track and mitigate pollution and even pilot ships.

- It can also foster environmental protection and sustainability at sea.

- Beneficial effects of AI have been acknowledged by the European Commission in relation to the Green Deal

The new AI Act

- The AI Act entered into force in August 2024

- As a text with EEA relevance, it will likely have influence in Norway

- No final statements as to the extent of its implementation

- A DFØ report from August 2024 provides recommendations for the administrative implementation of the regulation
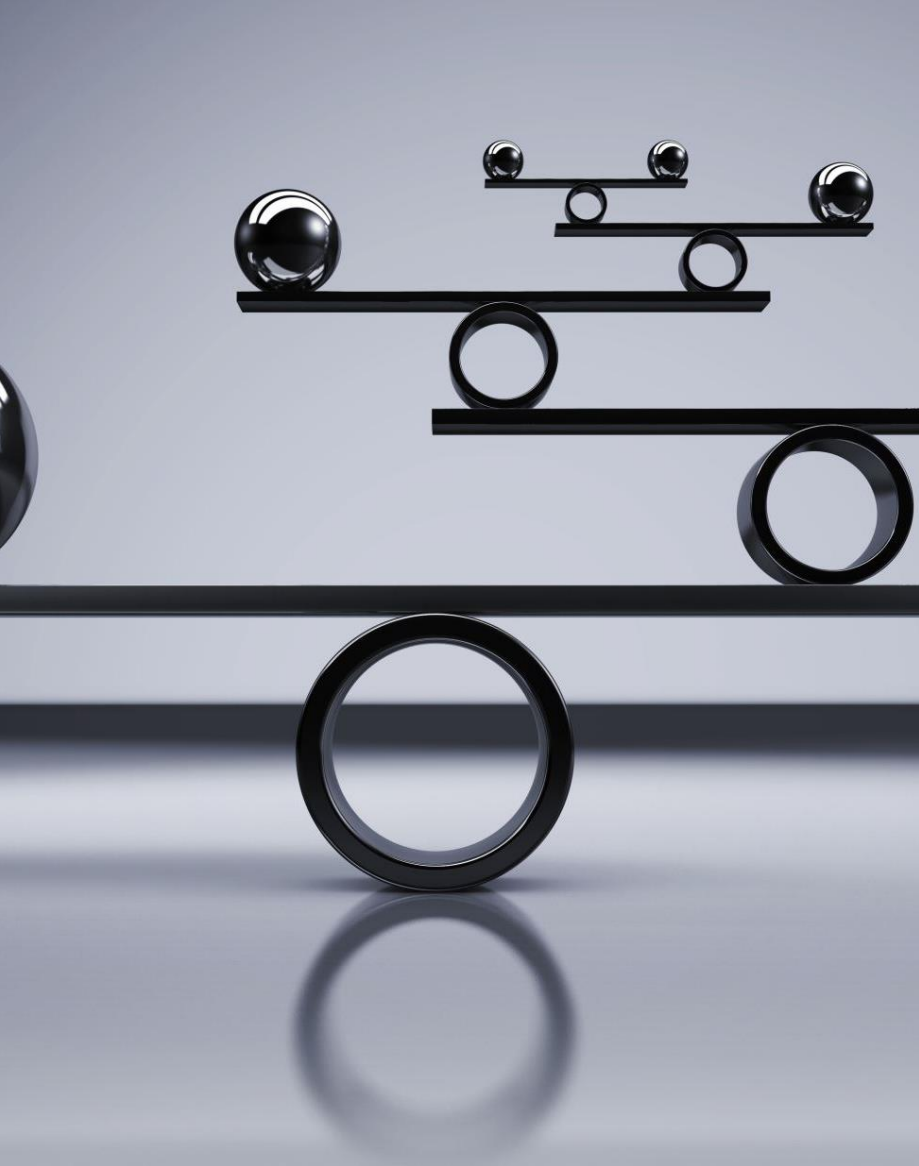
Risk-based Approach

- The AI Act distinguishes between four levels of risk:

➢ Unacceptable (forbidden)

➢ High (specific obligations)

➢ Limited (transparency obligations)

➢ Minimal (no regulation)
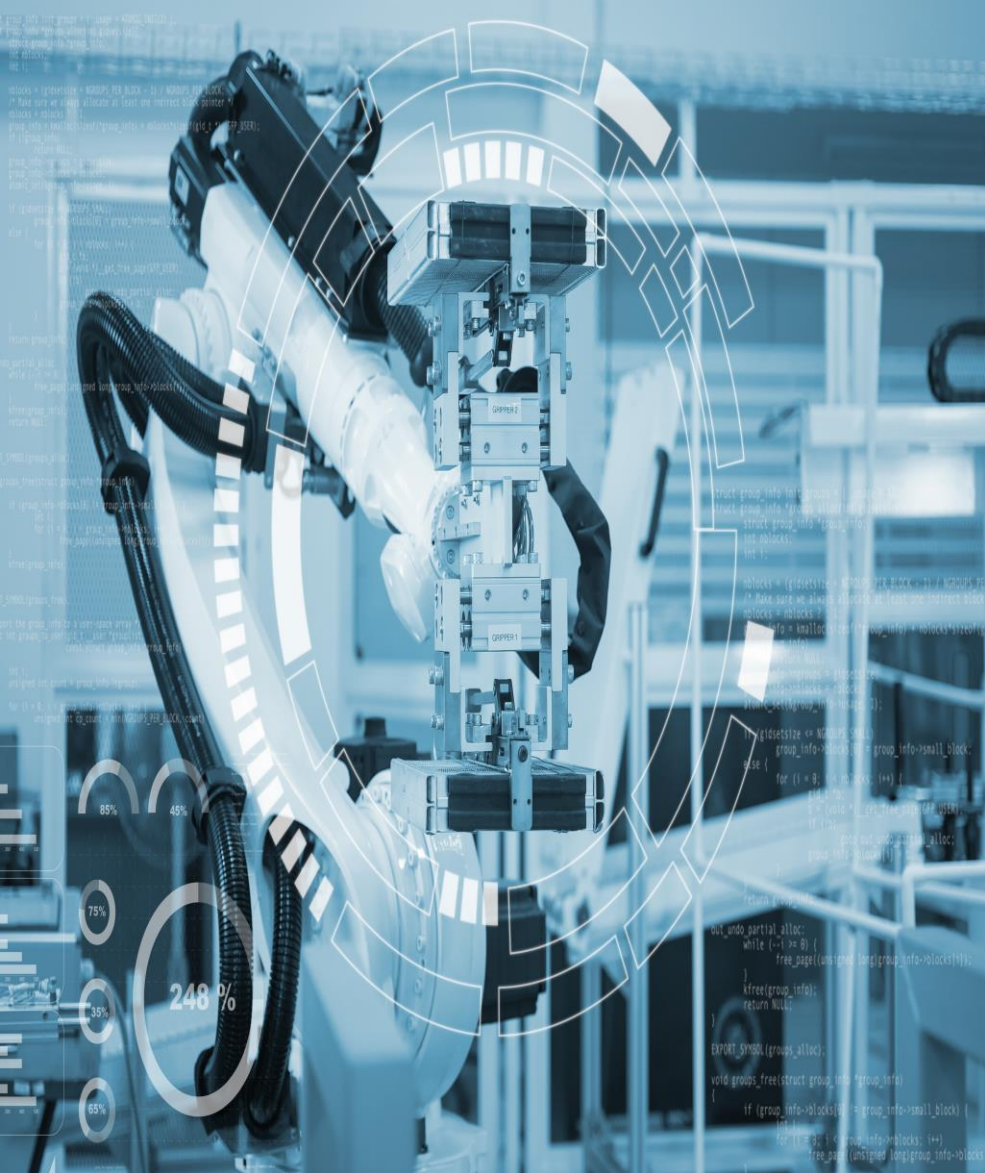
High-Risk AI

- An AI system shall be considered high-risk where **both** of the following conditions are fulfilled (Article 6 AI Act):

- The AI system is intended to be used as a safety component of a product, or the AI system is itself a product, covered by the Union harmonisation legislation listed in Annex I;

- The product whose safety component pursuant to point (a) is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment, with a view to the placing on the market or the putting into service of that product pursuant to the Union harmonisation legislation listed in Annex I.

- AI systems in Annex III are also considered high-risk
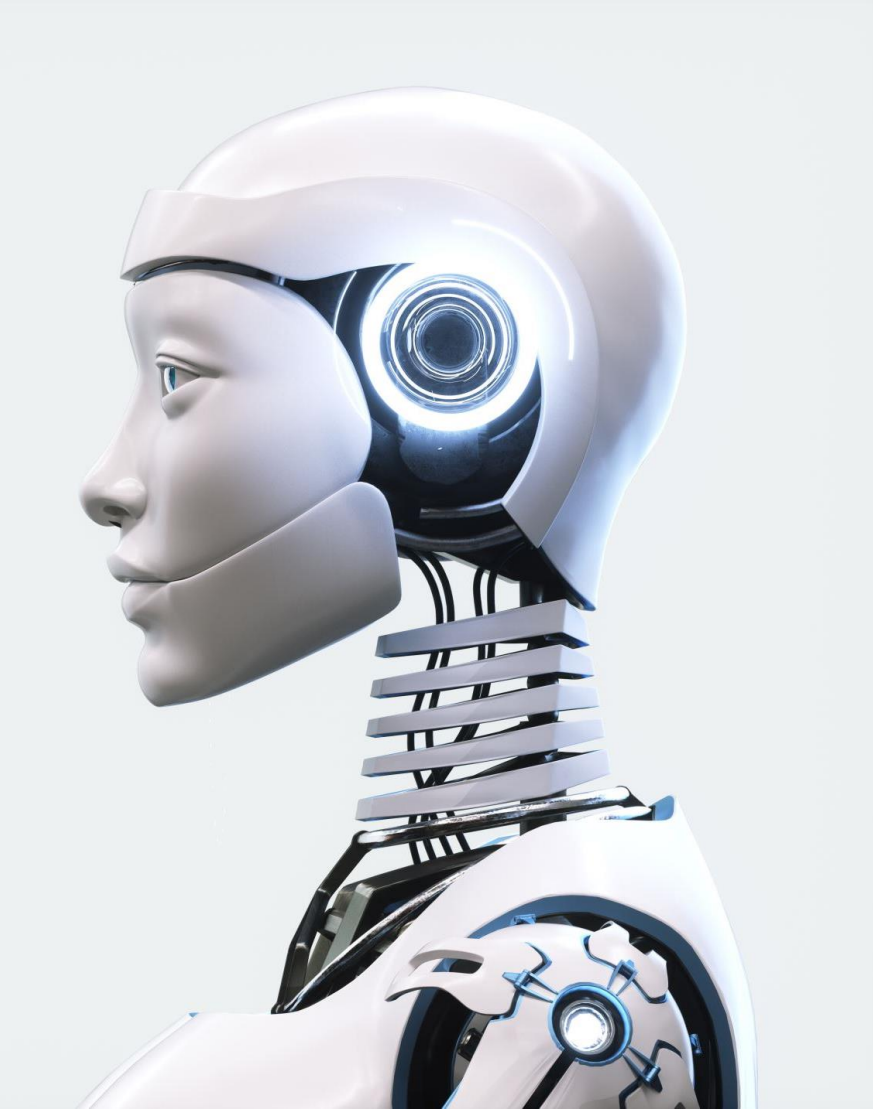
Exceptions from Classification as a High-risk System

- For AI systems listed in Annex III

➤ Not to be considered high-risk, if it does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making

➤ *The AI system is intended to perform a narrow procedural task;*

➤ *The AI system is intended to improve the result of a previously completed human activity;*

➤ *The AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review*

➤ *The AI system is intended to perform a preparatory task to an assessment relevant for the purposes of the use cases listed in Annex III.*

Potential Risk Classification of AI in the Maritime Sector (and Petrol)

- AI-powered equipment could fall under the scope of the Machinery Directive (MD)

- Machinery (Article 2 (a) MD): an assembly, fitted with or intended to be fitted with a drive system other than directly applied human or animal effort, consisting of linked parts or components, at least one of which moves, and which are joined together for a specific application

- Machinery is to be interpreted broadly: extends to safety components, interchangeable equipment, lifting accessories etc.

Potential Risk Classification of AI in the Maritime Sector (and Petrol)

- Could also fall under Annex III AI Act

- Annex III lists some critical use cases for AI

- Critical infrastructure: AI systems intended to be used as safety components in the management and operation of critical digital infrastructure, road traffic, or in the supply of water, gas, heating or electricity. (Annex III, no. 2)

Requirements for High-risk AI Systems

- Articles 8 to 27 of the AI Act set out a number of obligations in relation to high-risk AI systems.

➢ Obligations relating to the design and development

➢ Obligations of providers and deployers

➢ Obligations of importers and distributors

➢ Appointment of an authorised representatives

➢ Obligations along the value chain

➢ Fundamental rights impact assessment
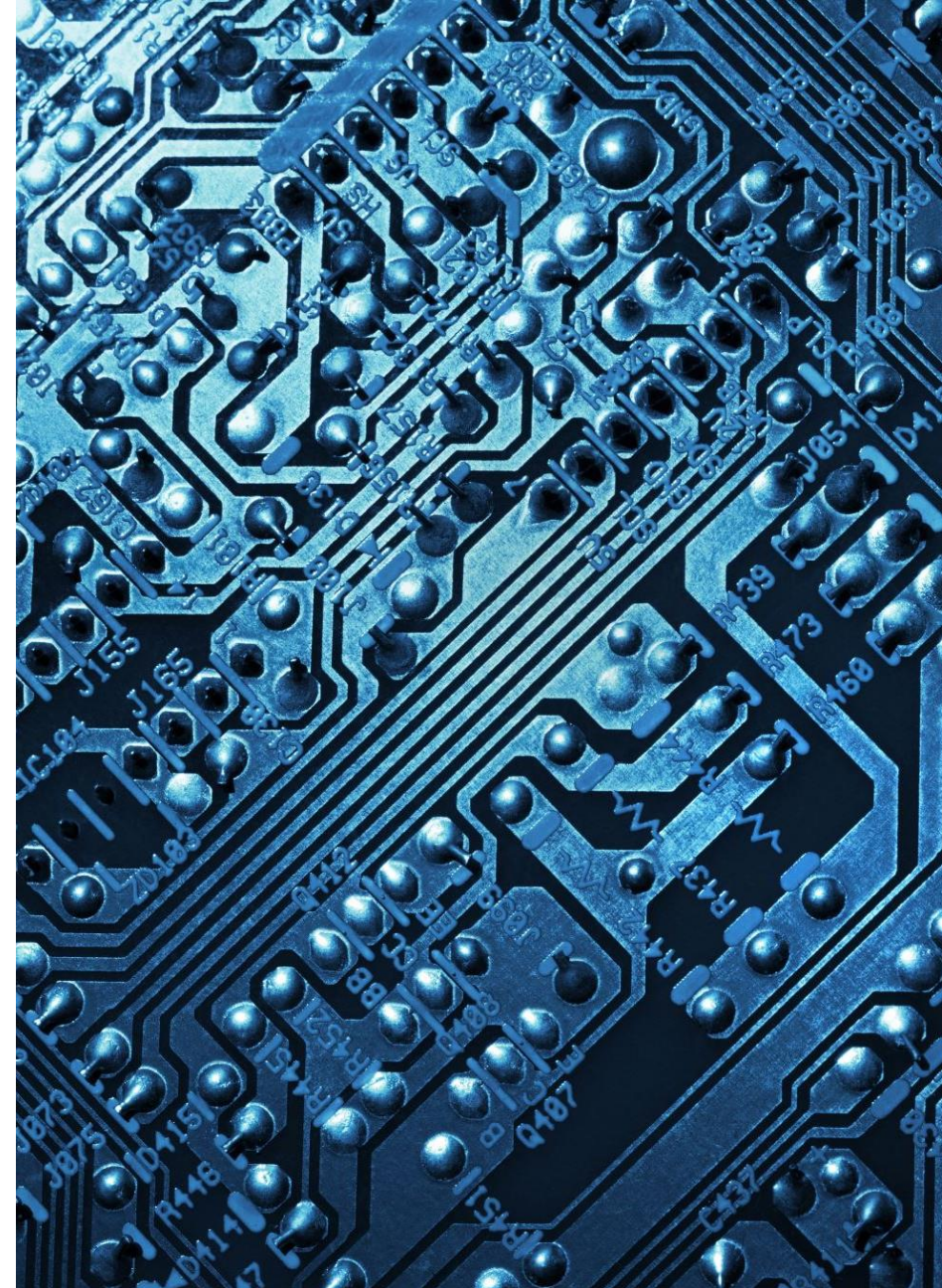
Obligations relating to Design and Development of High-risk AI

- Article 9: risk-management system

- Article 10: data and data governance

- Articles 11 and 12: technical documentation and record-keeping

- Article 13: transparency

- Article 14: human oversight

- Article 15: accuracy, robustness and cybersecurity

Human Oversight

- The regulatory approach to AI in the EU strives to ensure trustworthy and human-centered AI

- AI systems shall not undermine human autonomy or cause adverse effects

- Human oversight shall prevent or minimize risks to health, safety or fundamental rights

Human Oversight – Article 14 AI Act

- AI systems must be designed and developed in a way that they can be effectively overseen by natural persons while in use

➤ e.g. by means of interfaces

- Oversight measures must correspond to risk, level of autonomy and context of use

- One or both types of the following measures

➤ Measures identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service

➤ Measures identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the deployer

Human Oversight – Article 14 AI Act

- Natural persons, to whom human oversight is assigned, must be enabled to the following:

➢ Proper understanding of the relevant capacities and limitations of the high-risk AI system

➢ Duly monitoring its operation, including in view of detecting and addressing anomalies, dysfunctions and unexpected performance

➢ Awareness of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (automation bias), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons

➢ Correct interpretation of the high-risk AI system's output, taking into account, for example, the interpretation tools and methods available

➢ Deciding not to use the high-risk AI system or to otherwise disregard, override or reverse its output

➢ Intervening in the operation of the high-risk AI system or interrupt the system through a 'stop' button or a similar procedure that allows the system to come to a halt in a safe state.

Criticism raised against Human Oversight

- No legal or other definition of what a human centric approach to AI implies – neither from a societal or regulatory perspective

- It is unclear what 'effective' human oversight requires

- Black-box effect may jeopardize human oversight

- Some applications may be opaque even to their programmers

- Human oversight lacks feasibility – it may not be possible to 'fully' understand an AI system

- One-size-fits-all approach makes the provision vague
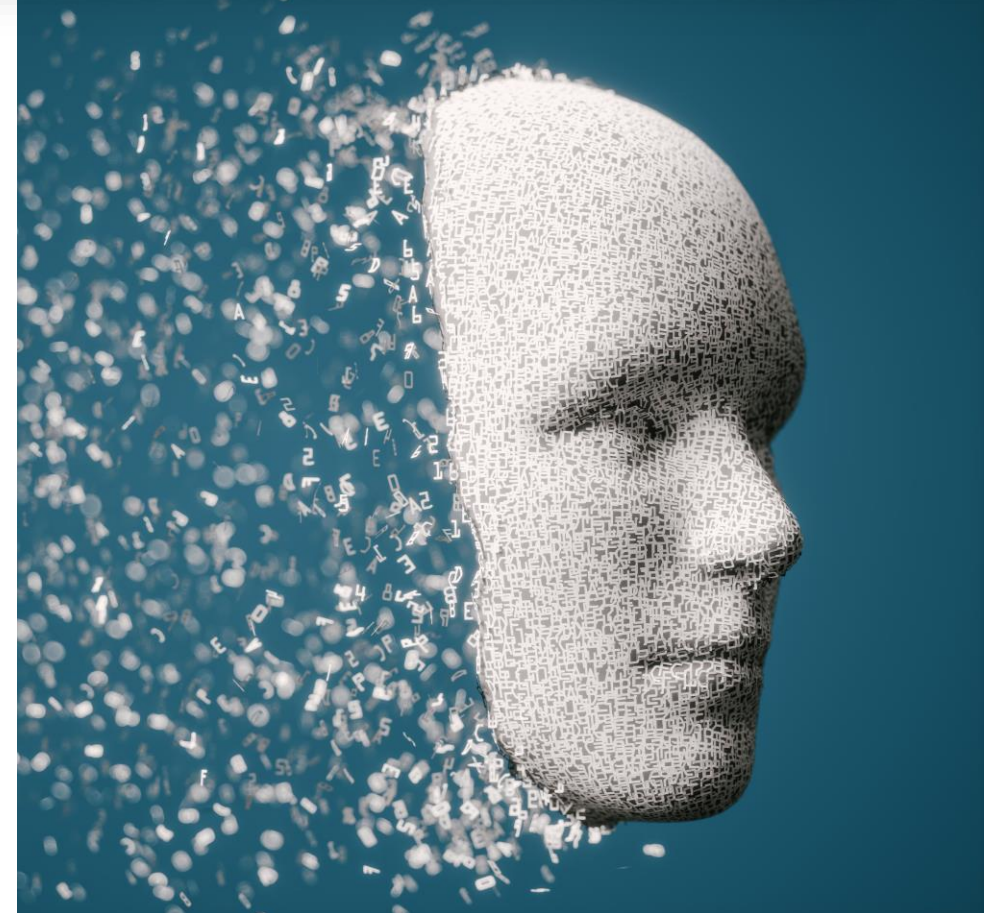
AI Regulation and Maritime Activities

- Issues regarding the territorial scope of application of the AI Act may arise

- AI is at the intersection of the AI Act and Law of the Sea

- No clarification as to its applicability to offshore activities

- Should apply throughout EEZ, but clarification would be welcome

- Interplay with any prospective IMO Convention must be clarified

THANK YOU!

Keep in touch: beatrice.schutte@ulapland.fi

Latest publication: Damage caused by autonomous ships: towards regulation for civil liability in EU waters? https://doi.org/10.4337/9781035321469.00018

LAPIN YLIOPISTO
UNIVERSITY OF LAPLAND
For the North – For the World