

# The Data and Artificial Intelligence Group (DART)

Autumn 2023



**Big Data**



**AI**

# DART Teachers



# Three-pronged Focus of DART



Research

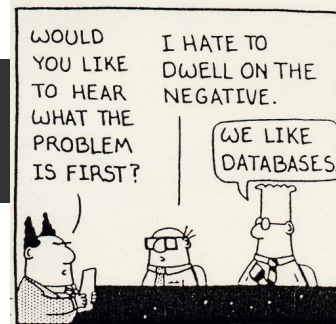


Education



Innovation

# Databases and Search



## Databases

- Algorithms for multicore and GPUs
- In-memory databases
- Handling complex datatypes: trajectories, graphs, etc.

- Parallel and distributed** DB systems.

- NoSQL/NewSQL DBs

- Big Data Frameworks (Hadoop, Spark...)

## Search

- Text-, person-, product-search

- Combining text and time

- Machine Learning for search

- Natural-Language Processing (NLP)

- Large Language Models (GPTs)

## Text / Web Mining

- Finding useful info in a large data corpus.

- Social media mining – increasing the value of user-generated data by:

  - Finding relationships between users

  - Opinion mining

  - Community detection

- Linking open data to structured knowledge bases (DBPedia etc.)

# Previous Projects and Master's Theses

- Determining creation time of documents based on contents
- **Handling big spatial data**
- DB tracking on remote processors and GPUs
- Automatic scaling and maintenance of a NoSQL database
- In-memory databases
- Spatio-textual trajectory mining
- Analyzing bank transactions
- Text mining on Twitter data
- Mining dynamic graphs
- Personalised, intelligent object search and recommendation
- Scalable search in knowledge bases
- Machine learning in cache replacement algorithms
- Online upgrade in NoSQL databases
- **Detection of fake news**

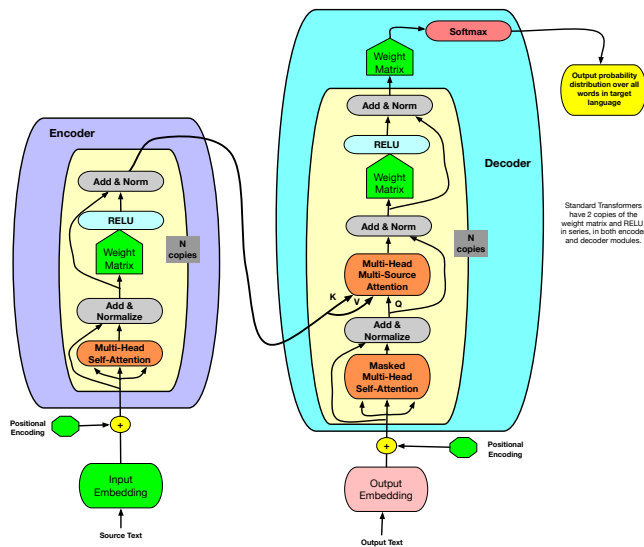


**DataBase Research:**  
**Structuring, Searching + Analyzing Data + ...**

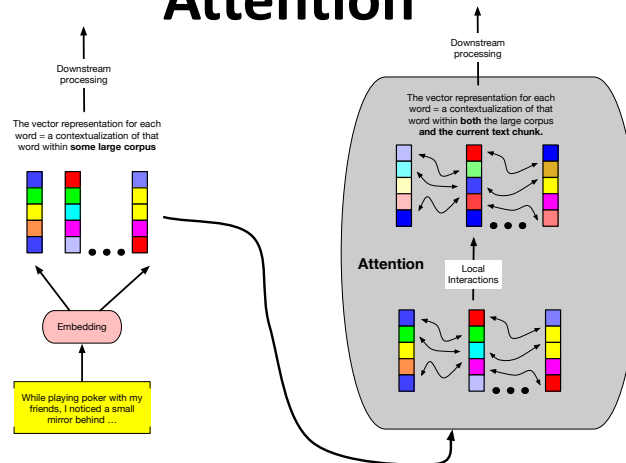
**AI Research:**  
**Abstracting + Generating Data + ...**

# Transformer

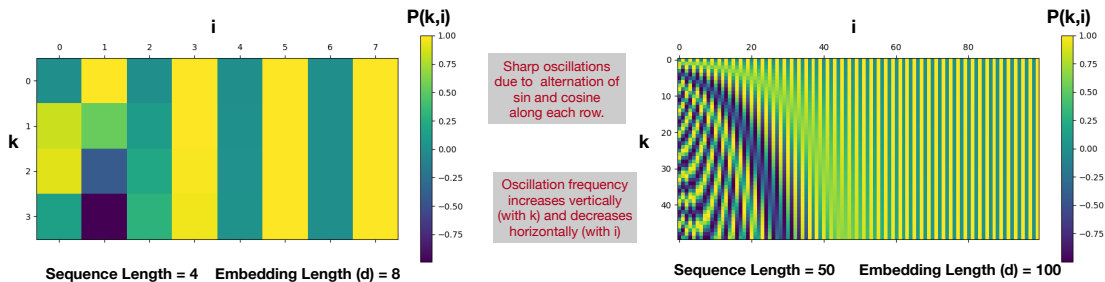
What is all this??



## Attention



## Embedding



**Answer:  
The guts of  
ChatGPT  
(as best  
we know)**





**Me:** Compose a 4-line poem that introduces VERY smart people to our Database and AI group.

**ChatGPT (3.5):**

In realms of wisdom, where brilliance does reside,  
Very smart minds, our AI group's pride,  
With knowledge vast and intellect ablaze,  
They join our database in myriad ways.

**Me:** Could you make it a little more dystopian?

**ChatGPT (3.5):**

In a world of shadows, where brilliance is concealed,  
The very smart we gather, their knowledge sealed,  
Within our database, their potential held tight,  
In this grim reality, their insights out of sight.

# Words from The Masters

*“... there is currently a war for the scarce AI talent...”*

*+ “AI is the new electricity.”*

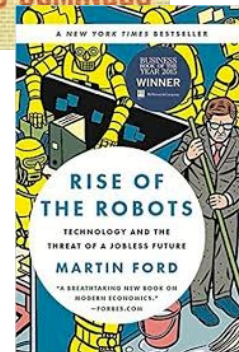
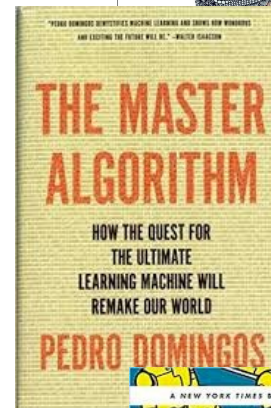
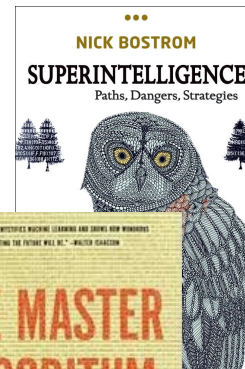
– **Andrew Ng**, Professor at Stanford

*“The rise of powerful AI will be either the best or the worst thing ever to happen to humanity. We do not know which.” – **Stephen Hawking**, famous physicist.*

*“If you don’t learn AI, you’re going to be a dinosaur within three years.” – **Mark Cuban**, famous US businessman.*

*“You won’t lose your job to an AI, but you might lose it to someone who knows AI.” - **Scott Galloway** (popular tech podcaster)*

*The big tech companies now compete to build  
**THE AI platform (Chatbot?) of the future.***



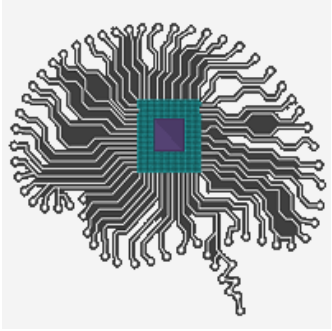
NTNU

Norwegian University of Science and Technology

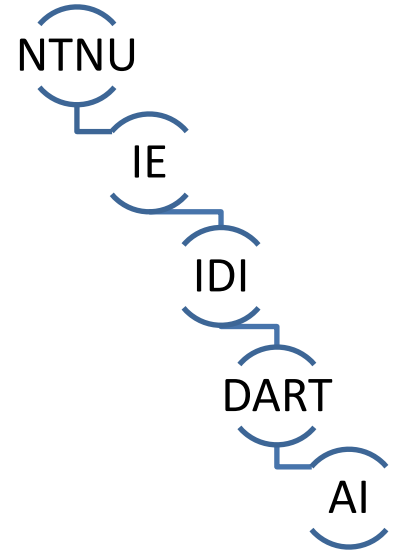
# Popular AI Topics at IDI

- Deep Learning (Neural Networks)
- Natural Language Processing / Text Analytics
- Bayesian Decision-Making
- Case-Based Reasoning
- Evolutionary Algorithms
- Multi-Agent Systems / Swarm Intelligence
  - What's in your toolkit?
  - Today, AI is everywhere, and anyone can use it, but you need **special competence** to **understand + develop** it.

# AI Research Goals



- Technological
  - Develop **smart** computer systems: those that behave **intelligently** (often in narrow domains), and which **learn** from experience.
- Scientific
  - Increased understanding of intelligence in both natural and artificial systems by designing computational models of sensing, reacting, reasoning and learning.

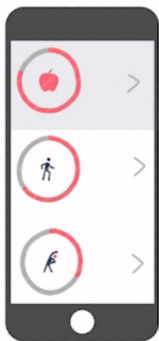


20+ ongoing externally-funded projects

# Artificial Intelligence in Healthcare

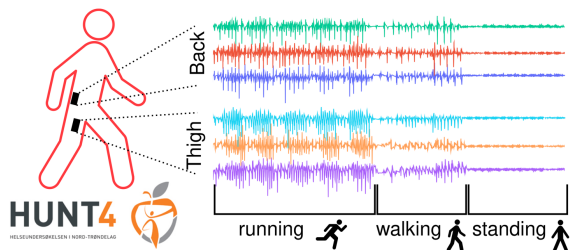
## Decision Support for Patients and Clinicians

- Developing AI-driven (CBR) applications for patients
- Developing co-decision making systems for GPs and Physiotherapists



## Time Series Data

- Human Activity Recognition from sensor data
- Classification models for real-time healthcare datasets



## Understanding Health Trajectories

- Role of sleep for Physical activity behaviour and mental health
- Analysis and prediction models of clinical interventions



# BIAS: Mitigating Diversity Biases of AI in the Labor Market

(EU-Horizon, Pinar Ozturk, 2022-26)



## Company context

- Business goals
- Hiring policy
- rules

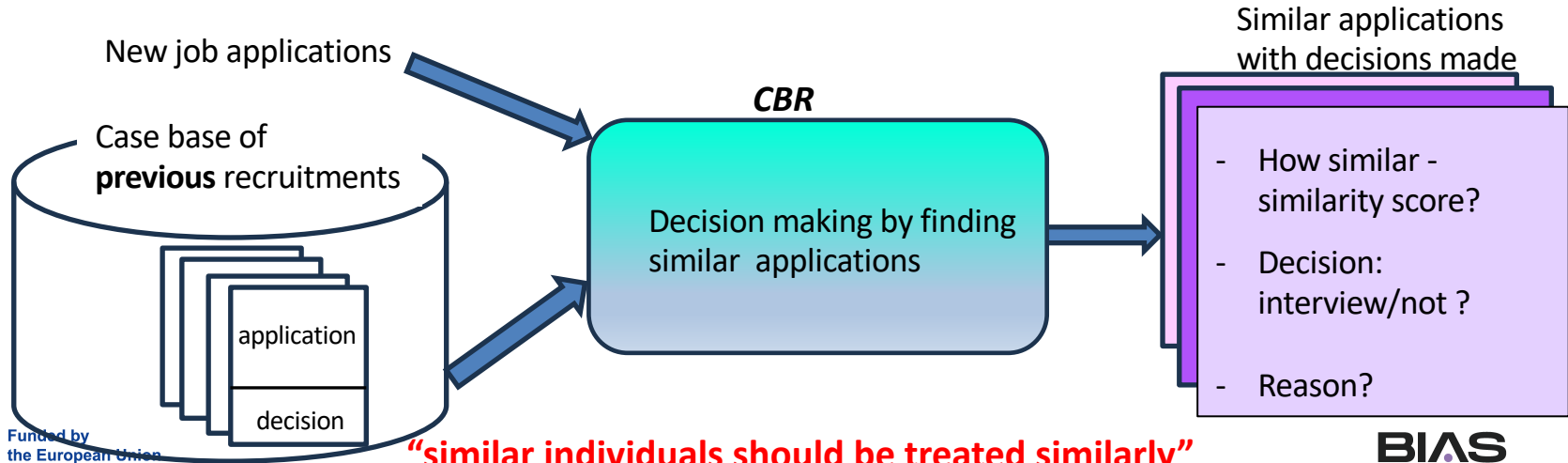
## Law/legal issues

Anti-discrimination  
law  
Employee rights  
...

## Social/Philosophical issues

Justice  
Social-welfare  
...  
... .. ?

*Knowledge base*



**"similar individuals should be treated similarly"**

**BIAS**

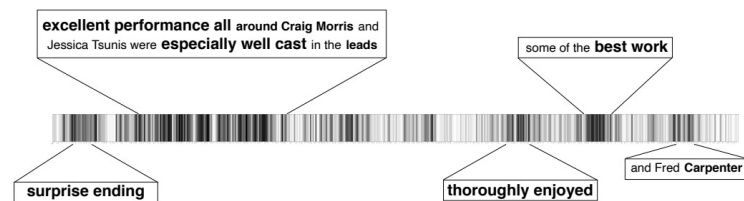
# ShuttleNet: Scalable Neural Models for Long Sequential Data

Norwegian Research Council project 287284

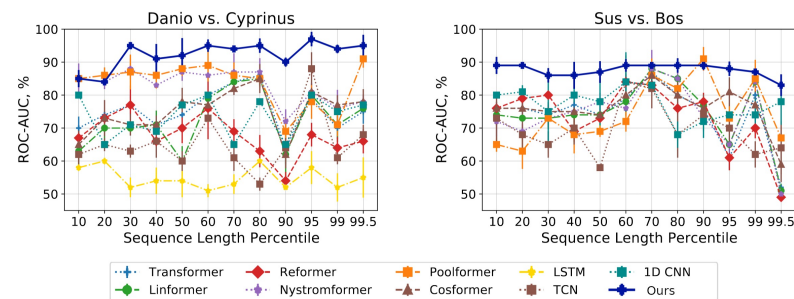
Research group: Zhirong Yang, Lei Cheng, Ruslan Khalitov, and Tong Yu

## Achievements

- Works accurately for sequence length **1,500,000** on a single machine (submitted to **ICLR2022**)
- Defeats **all** X-Formers in **all** Long Rang Arena benchmark tasks (published in **Neural Networks**)
- Surpasses DeepMind's Enformer by **15%** in accuracy for genetic variant classification (submitted to **Bioinformatics**)
- Wins in long DNA-based taxonomy classification (submitted to **ICLR2022**)
- Yields good attention maps (published in **CVPR2022**)
- Leads in long time series classification (accepted in **Neurocomputing**)



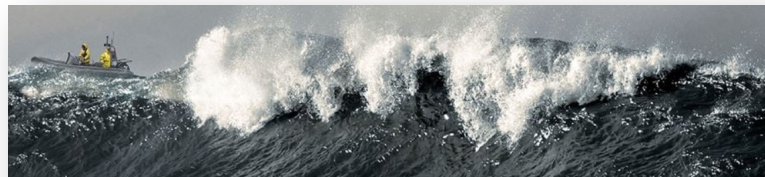
Attention vector visualization of a positive review in Text Classification.





# Exposed SFI (NFR)

- Coordinated by Sintef Ocean, ending 2023
- PI at IDI: Helge Langseth, collaboration across NTNU, Sintef, ++
- PhD student (IDI): Håkon Måløy, Sverre Herland
- Postdoc (IDI): Bjørn Magnus Mathisen
- Objective:
  - **Develop knowledge and technologies for exposed aquaculture operations**
  - **enabling a sustainable expansion of the fish farming industry**



## SFI Exposed

The EXPOSED aquaculture operations Centre will take advantage of Norway's strong position in the aquaculture, maritime and offshore sectors to enable safe and sustainable seafood production in exposed coastal and ocean areas.

<https://www.sintef.no/en/ocean/initiatives/sfi-exposed/>

# AROS (NFR-funded project, 2020-2023)

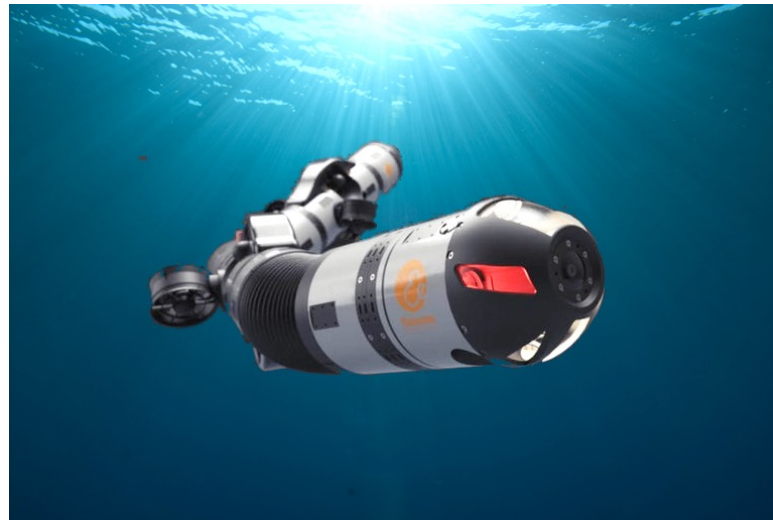


**Project goal: Perception, control and energy efficient operation of underwater snake robots**

- Contributors: K.Y. Pettersen, (lead), A. Stahl, T. Gravdahl (all ITK), M. Greco (IMT), R. Mester (IDI)

Work group Mester:

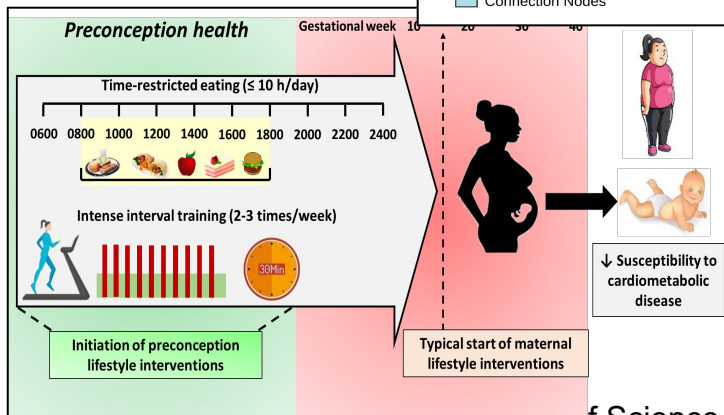
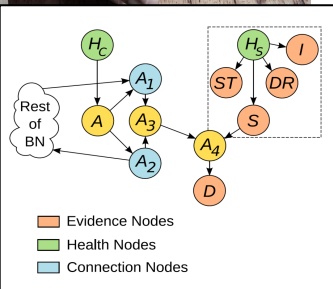
- visual perception and egomotion estimation
- Machine learning-based processing of (visual) sensor data from difficult viewing conditions



# Explainable Machine Learning for Management of Diabetes



Objective: To establish an optimal diet-exercise prevention strategy that commences *before* conception to improve maternal and infant health outcomes in women with increased risk of gestational diabetes



Objectives of project are to:

- identify, develop, and experiment with **explainable AI techniques**, focus on machine learning using **causal Bayesian networks** applied to diabetes data;
- complement an on-going randomised controlled trial at the Dept. of Circulation and Medical Imaging, *Before The Beginning* (BTB);
- **improve and develop explainable causal ML-models for prediction of diabetes and its progression of comorbidities** in order to provide patient-tailored treatment plans; and
- **discover correlation and causation in bio-data**, including clinical outcomes, from women with increased risk of GDM and possible comorbidities based on collected data from BTB.

**Ph.D. candidate: N.N.**

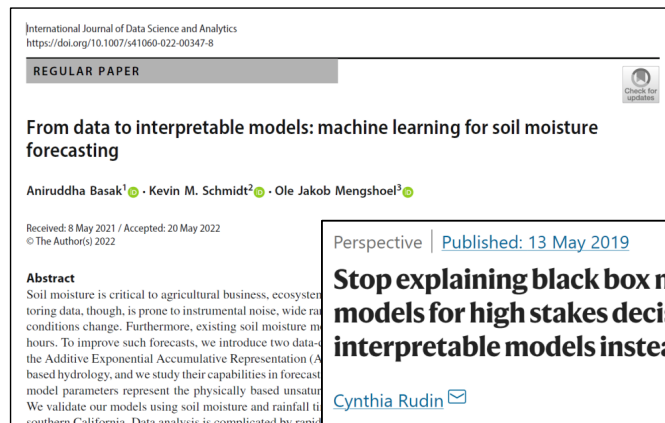
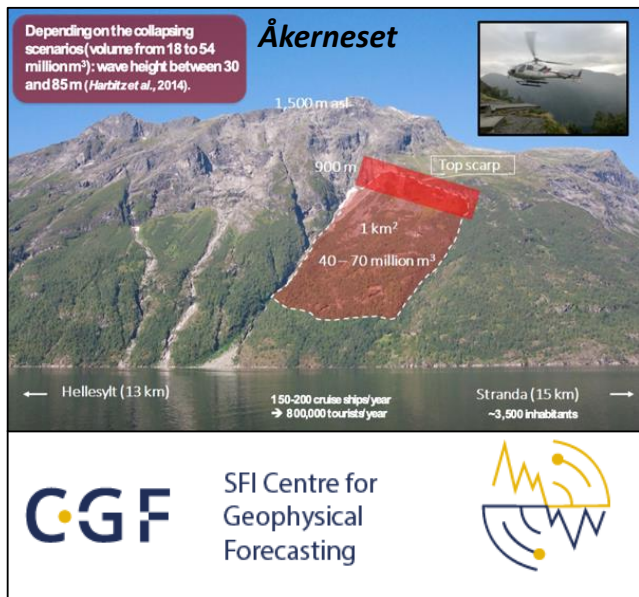
**Advisors:** Ole Jakob Mengshoel and Trine Moholdt (Department of Circulation and Medical Imaging)

**Time line:** Fall 2023 – Fall 2026 or Fall 2027

**Application deadline:** 15.1.2023, check [jobbnorge.no](http://jobbnorge.no)

**Funding:** NTNU

# Trustworthy, Explainable, Interpretable AI



## "Why Should I Trust You?" Explaining the Predictions of Any Classifier

Marco Tulio Ribeiro  
University of Washington  
Seattle, WA 98105, USA  
marcotcr@cs.uw.edu

Sameer Singh  
University of Washington  
Seattle, WA 98105, USA  
sameer@cs.uw.edu

Carlos Guestrin  
University of Washington  
Seattle, WA 98105, USA  
guestrin@cs.uw.edu

transportation ... earth science ... geophysics ... landslides ... human safety ... prediction ... forecasting ...  
classification ... black boxes ... explanation ... interpretability ... trustworthiness ... high stakes ... decisions



# Large Language Models (LLMs)

## Probability Distributions over Sequences of Tokens

Existing Models: ChatGPT (OpenAI),  
Bard (Google), LLAMA  
(Meta), BLOOM, ...

### Team

1 Professor  
3 PostDocs  
12 Research Assistants  
4+ External Collaborators

### Resources

IDUN (28 NVIDIA A100 80  
GB GPUs)  
500 GB Text  
→ 85% Norwegian (+SWE,  
DAN, GER, ENG)  
→ NCC (NB), Websites,  
News Articles, Social  
Media

### Outcome

23B GPT (summarization)  
23B GPT (generation)  
3B GPT (summarization)  
3B GPT (generation)  
3B LLAMA

### Motivation

- Reduce Reliance on Big Tech
- Reflect Norwegian Culture
- Facilitate Data Ownership

### Contact

Jon Atle Gulla  
Benjamin Kille

NorGPT (Jon Atle Gulla, Benjamin Kille)

and Technology

NorwAI

Norwegian Research Center  
for AI Innovation



# Norwegian Open AI Lab

- A center for research, innovation and education within Artificial Intelligence
- Established in 2017, as Telenor-NTNU AI Lab with SINTEF
- Ensures collaboration between academia, industry and public sector
- Hosted by the Department of Computer Science, NTNU
- Strong NTNU research team, with a core team of 12 professors, 21 associated professors & 75+ PhD students

## OUR PARTNERS



# NorwAI

## Center for Research-based Innovation

*Enhance technology transfer, internationalization and researcher training through long-term research conducted in close collaboration between research-performing companies and prominent research groups.*



NTNU



DNV



KONGSBERG



COGNITE

TrønderEnergi<sup>®</sup> ∞



Schibsted



Retriever



DNB

**2020 - 2028**







# AI and UN's Sustainability Goals



# AI and UN's Sustainability Goals



***The goal of science is not universal truth...but the gradual removal of prejudice (Niels Bohr).***

The Prejudice	The Science
Earth is center of the universe	Copernican Revolution
Humans are a special species	Darwinian Evolution
Superiority of human intelligence	AI