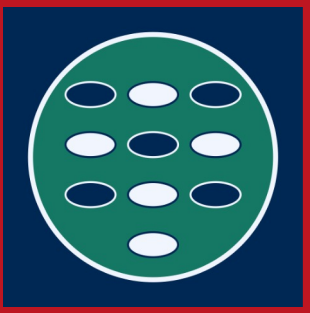
Noregs miljø- og biovitskaplege universitet

Institutt for datavitskap



Digitalisering på Ås

DAT390 Data science seminar

4 Research impact and ethics

4.4 Ethical constraints on the research process

4.5 AI-related recommendations on ethics

ALTAI categories of AI-related ethics issues

The following seven aspects have been identified by the High-Level Expert Group on Artificial Intelligence within its **Assessment List for Trustworthy Artificial Intelligence** (ALTAI)[1,2]:

1) Human agency and oversight

2) Technical robustness and safety

3) Privacy and data governance

4) Transparency

5) Diversity, non-discrimination and fairness

6) Environmental and societal well-being

7) Accountability

[1]EC Directorate-General for Communications Networks, *Assessment List for Trustworthy Artificial Intelligence (ALTAI)*, Brussels: EC, ISBN 978-92-76-20009-3, doi:10.2759/002360, **2020**

[2]https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

ALTAI #1: Human agency and oversight

The following seven aspects have been identified by the High-Level Expert Group on Artificial Intelligence within its **Assessment List for Trustworthy Artificial Intelligence** (ALTAI):

1) **Human agency and oversight**

2) Technical robustness and safety

3) Privacy and data governance

- Informed decisions
- Human-in-the-loop
- Human-in-command approaches

European AI Act proposal: "To address the **opacity** that may make certain AI systems incomprehensible to or too complex for natural persons, a certain degree of transparency should be required for high-risk AI systems.[1] Users should be able to interpret the system output and use it appropriately. High-risk AI systems should therefore be accompanied by **relevant documentation**".

[1]Systems with "high risk" include all "safety components" related to "water, gas, heating, and electricity."

ALTAI #2: Technical robustness and safety

The following seven aspects have been identified by the High-Level Expert Group on Artificial Intelligence within its **Assessment List for Trustworthy Artificial Intelligence** (ALTAI)[1]:

1) Human agency and oversight

2) Technical robustness and safety

3) Privacy and data governance

- Resilient and secure
- Safe, ensuring a fall back plan
- Accurate, reliable and reproducible

«Could the AI system have adversarial, critical or damaging effects? […]

Is the AI system certified for cybersecurity (*e.g.* the certification scheme created by the **Cybersecurity Act in Europe**)[2] or is it compliant with specific security standards?»[1]

[1]EC Directorate-General for Communications Networks, *Assessment List for Trustworthy Artificial Intelligence (ALTAI)*, Brussels: EC, ISBN 978-92-76-20009-3, doi:10.2759/002360, **2020**

[2]https://ec.europa.eu/digital-single-market/en/eu-cybersecurity-act

ALTAI #3: Privacy and data governance

The following seven aspects have been identified by the High-Level Expert Group on Artificial Intelligence within its **Assessment List for Trustworthy Artificial Intelligence** (ALTAI)[1,2]:

1) Human agency and oversight

2) Technical robustness and safety

3) **Privacy and data governance**

Besides ensuring full respect for **privacy and data protection**, adequate data governance mechanisms must also be ensured, taking into account the **quality and integrity of the data**, and ensuring **legitimised access** to data.

7) Accountability

[1]EC Directorate-General for Communications Networks, *Assessment List for Trustworthy Artificial Intelligence (ALTAI)*, Brussels: EC, ISBN 978-92-76-20009-3, doi:10.2759/002360, **2020**

[2]https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

ALTAI #4: Transparency

The following seven aspects have been identified by the High-Level Expert Group on Artificial Intelligence within its **Assessment List for Trustworthy Artificial Intelligence** (ALTAI):

1) Human agency and oversight

2) Technical robustness and safety

3) Privacy and data governance

4) Transparency

Tendency: Data must become explainable-AI-ready (XAIR). **Making data trust-worthy through explanations** will increasingly become a legal requirement.

«Can you trace back which data was used by the AI system to make a certain decision(s) or recommendation(s)? […]
Do you continuously survey the users if they understand the decision(s)?»[1]
-> Humans need to be aware that they are interacting with an AI system, and must be informed of the system's capabilities and limitations.

[1]EC Directorate-General for Communications Networks, *Assessment List for Trustworthy Artificial Intelligence (ALTAI)*, Brussels: EC, ISBN 978-92-76-20009-3, doi:10.2759/002360, **2020**

ALTAI #5: Diversity, fairness, and #6: Well-being

Cognitive biases (*cf.* types of biases[1]) can be introduced at many points in the process. They can create epistemic injustice and put groups of people at a disadvantage.
- AI systems should be accessible to all, regardless of any disability.

CARE principles[2]

–Origin: Global Indigenous Data Alliance

–Uptake supported by the Research Data Alliance

–Orientation: Sovereignty and epistemic justice

5) Diversity, non-discrimination and fairness

6) Environmental and societal well-being

7) Accountability

Sustainable and environmentally friendly

See also NMBU's ethical guidelines[3], *pp.* 12 and 14.

[1]E. Dimara *et al.*, *IEEE Transact. Vis. Comp. Graph.* **26**: 1413, doi:10.1109/tvcg.2018.2872577, **2020**.
[2]S. Russo Carroll *et al.*, *Sci. Data* **8**: 108, doi:10.1038/s41597-021-00892-0, **2021**.
[3]https://www.nmbu.no/en/research/forskningsetikk

ALTAI #7: Accountability

The following seven aspects have been identified by the High-Level Expert Group on Artificial Intelligence within its **Assessment List for Trustworthy Artificial Intelligence** (ALTAI)[1,2]:

1) Human agency and oversight

2) Technical robustness and safety

«Did you ensure that the AI system can be audited by independent third parties? […] Did you foresee any kind of external guidance or third-party auditing processes to oversee ethical concerns and accountability measures?»[1]

«Did you establish a process for third parties […] to report […] vulnerabilities»?[1]

7) Accountability

Assessment of algorithms, data and design processes

[1]EC Directorate-General for Communications Networks, *Assessment List for Trustworthy Artificial Intelligence (ALTAI)*, Brussels: EC, ISBN 978-92-76-20009-3, doi:10.2759/002360, **2020**

[2]https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment