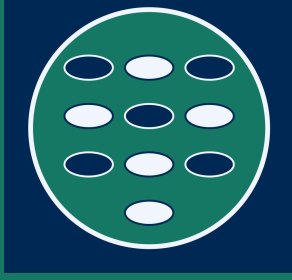
Norges miljø- og biovitenskapelige universitet

Institutt for datavitenskap



Digitalisering på Ås

DAT390

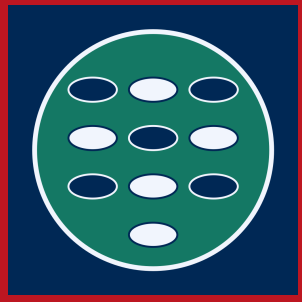
Data science seminar

4 Research ethics and impact

4.6 Ethical challenges specific to artificial intelligence

Noregs miljø- og biovitskaplege universitet

Institutt for datavitskap



Digitalisering på Ås

4 Ethics and impact

4.6 AI ethics challenges

DAT390

2nd December 2024

Categories of research ethics issues

List of ethics issues applicable to Horizon Europe research:

1) Human embryos and human embryonic stem cells

2) Humans ("Does this activity involve human participants?")
   → Special case: Clinical trials as defined by Regulation EU 536/2014

3) Human cells and tissues
   → Beyond embryonic cells/tissues which are covered under issue *no.* 1

4) Processing of personal data

5) Animals ("Does this activity involve animals?"), *cf.* NMBU's guidelines, *p.* 13*f.*

6) Activities carried out in other countries (for Horizon Europe: Outside the EU)

7) Environment, health, and safety

8) **Artificial Intelligence**

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]

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**

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

European AI Act: "To address concerns related to **opacity** and […] fulfil their obligations under this Regulation, **transparency** should be required for high-risk AI systems before they are placed on the market […]. High-risk AI systems should […] enable deployers to understand how the AI system works […]. High-risk AI systems should be accompanied by **appropriate information**".

[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):[2]

1) Human agency and oversight

2) Technical robustness and safety

3) Privacy and data governance

«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**)[1] or is it compliant with specific security standards?»[2]

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

[2]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 #4: Transparency

The following seven aspects have been identified by the High-Level Expert

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

1) Human agency and oversight
2) Technical robustness and safety
3) Privacy and data governance (note that you may need to do a DPIA)
4) **Transparency**
5) Diversity, non-discrimination and fairness

«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]

[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

The following seve
Group on Artificial
Artificial Intelligen

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.

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

See also NMBU's ethics guidelines, *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**.

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]

1) Human agency and oversight

«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]

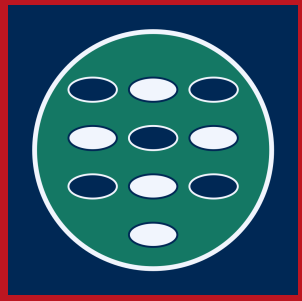
6) Environmental and societal well-being

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**

Noregs miljø- og biovitskaplege universitet

Institutt for datavitskap



Digitalisering på Ås

Examples / discussion

DAT390

2nd December 2024

Example #1

Could there be any ethics issues? What needs to be taken into account?

Hand-based microactivity recognition using wearable devices

Human Activity Recognition (HAR) has emerged as a necessary area of research in fields such as computer vision and robotics due to its wide-ranging applications in healthcare, surveillance, and human-robot interaction. […] My research focuses on methods and technical challenges associated with using wearable devices for recognizing fine-grained hand-based activities. […]

In my research, sensor data should be collected from participants wearing wearable devices. The main focus is on recording repetitive tasks such as typing and brief, less continuous movements such as reaching for something. However, I got the dataset from my advisor […] I am not sure if we need to do some experiments to collect more data.

Example #2

Could there be any ethics issues? What needs to be taken into account?

Privacy risk assessment in large-scale measurement datasets

This report explores the landscape of privacy risk assessment in large-scale Internet measurement datasets. It provides an in-depth literature review of current methods, including anonymization, differential privacy, and re-identification risk evaluation. The report lays the foundation for the proposed research methodology aimed at enhancing privacy preservation techniques while maintaining data utility. […]

The first step in this research involves the selection and collection of relevant large-scale measurement datasets. These datasets will be sourced from publicly available Internet traffic logs, network flow data, and anonymized user behavior records, ensuring com pliance with ethical and legal data use standards [1]. The raw data will undergo pre-processing to remove redundancies, clean erroneous records, and standardize formats, which is essential for minimizing noise that may interfere with subsequent analyses. […]

[1] A. Feldmann *et al.*, "The future of internet measurement: challenges and opportunities," *ACM Communications*, vol. **60**, pp. 82–89, **2022**.

Example #3

Could there be any ethics issues? What needs to be taken into account?

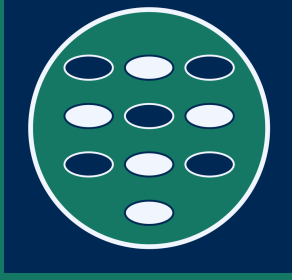
Semantic parsing for log analysis and anomaly detection

Semantic parsing is a technique in natural language processing (NLP) that converts unstructured text into a structured, machine-readable format. In the context of log analysis, semantic parsing focuses on extracting meaningful information from log mes sages, which are semi-structured text generated by software systems to record system events and activities. […]

The semantics miner aims to mine semantics at both the instance level and the message level. To achieve this, it solves two sub-problems: finding [concept-instance] pairs and classifying each token into a type. […] The parser uses a domain knowledge-assisted approach to resolve the implicit instance-level challenge of concepts and instances not coexisting in one log message.

Norges miljø- og biovitenskapelige universitet

Institutt for datavitenskap



Digitalisering på Ås

DAT390

Data science seminar

4 Research ethics and impact

4.6 Ethical challenges specific to artificial intelligence