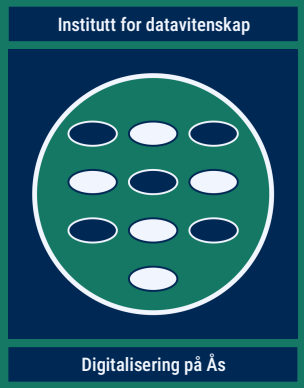
Norges miljø- og biovitenskapelige universitet



DAT390
Data science seminar

4 Research impact and ethics

4.3 Ethical constraints on the research process
4.4 AI-related recommendations on ethics

Norwegian University of Life Sciences

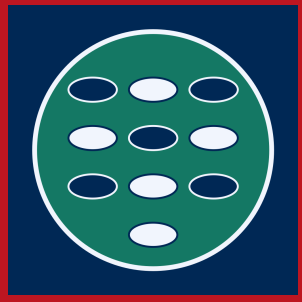
Highlight talk schedule for today

Monday, 27th November 2023

					Peer feedback from …	
15.10 – 15.14	#1	Razieh Kaveh		15.14 – 15.18	#1	Julie Overrein Suhail Rauf
		Classification confidence visualization of ANNs with adversarial robustness				
15.20 – 15.24	#2	Trishaban Jegatheeswaran		15.24 – 15.28	#2	Eljar Alihosseinzadeh Aydin Baris Ustun
		Application of semi-supervised learning in the process industry				
15.30 – 15.34	#3	Avnik Orbelians		15.34 – 15.38	#3	Mahrin Tasfe Bastian Undheim Øian
		Evaluating digital tools for welfare assessments of Atlantic salmon				
15.40 – 15.44	#4	Petter Bøe Hørtvedt		15.44 – 15.48	#4	Asim Rasheed Joel Yacob Teklemariam
		Laser disdrometers: Instrument characteristics and uncertainties				
15.50 – 15.54	#5	David Ajaegbu		15.54 – 15.58	#5	Awo Arab Sougata Bhattacharya
		Data-driven approach for the prediction of power flexibility				

Noregs miljø- og biovitskaplege universitet

Institutt for datavitskap



Digitalisering på Ås

4 Impact and ethics

4.3 Ethical constraints

DAT390

27. november 2023

Discussion: Whose responsibility is it?

Norwegian University of Life Sciences

Who needs to ensure compliance with research ethics guidelines, principles, and good practices in a master thesis research project?

Who is responsible if this fails and unethical practices have been followed?

- The person immediately carrying out the research?
  (That is, in this case, the master student.)

- The supervisors?
  (Here, that is particularly the main supervisor of the master thesis.)

- The institution?
  (Forskningsutvalg, forskningsetisk utvalg, prorector for research, *etc.*)

What does the law say?

Who is responsible?

- The person immediately carrying out the research?
  The researcher always has the primary responsibility. (FEK's summary.)
- The supervisors?
  Should usually also be involved as researchers.
  Additionally, they are a link between the researcher and the institution.
- The institution?
  «Forskningsinstitusjoner skal sikre at forskningen ved institusjonen skjer i henhold til anerkjente forskningsetiske normer. Institusjonen har ansvaret for:
  - a. nødvendig opplæring av kandidater og ansatte i anerkjente forskningsetiske normer og
  - b. at alle som utfører eller deltar i forskningen, er kjent med anerkjente forskningsetiske normer.» (Forskningsetikkloven §5)

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?")

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

7) Environment, health, and safety

8) Artificial Intelligence

Categories of research ethics issues

Which of these can most plausibly become relevant to DAT390 students?

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
→ see Assessment List for Trustworthy Artificial Intelligence (ALTAI)

Processing of personal data

Law and ethics are separate issues, but ethics can be backed up by the law.

Concerning personal data, we need to comply with GDPR, and therefore:

- You need to make sure that there is a line of responsibility connecting your work to the **Data Protection Officer** (DPO) of the organization.
- Your work may require a **Data Protection Impact Assessment**[1] (DPIA) …
  - «if you're using new technologies»,
  - «data […] used to make automated decisions about people»,
  - «if you're tracking people's location or behavior», «monitoring a publicly accessible place» or «processing children's data», *etc.*[1]
- You need **freely given**, **specific**, **informed**, and **unambiguous consent**.
  - Be aware that this can *introduce an additional bias* into your study!
  - Be aware of simultaneous requirements from NMBU's RDM policy.

[1]https://gdpr.eu/data-protection-impact-assessment-template/?cn-reloaded=1
(see also the DPIA template file)

Processing of personal data

Law and ethics are separate issues, but ethics can be backed up by the law.

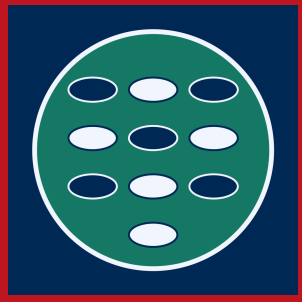
Concerning personal data, we need to comply with GDPR, and therefore:

- You need to make sure that there is a line of responsibility connecting your work to the **Data Protection Officer** (DPO) of the organization.
- Your work may require a **Data Protection Impact Assessment** (DPIA) …
  - «if you're using new technologies»,
  - «data […] used to make automated decisions about people»,
  - «if you're tracking people's location or behavior», «monitoring a publicly accessible place» or «processing children's data», *etc.*
- You need **freely given**, **specific**, **informed**, and **unambiguous consent**.
  - Be aware that this can *introduce an additional bias* into your study!
  - Be aware of simultaneous requirements from NMBU's RDM policy.

SIKT – Kunnskapssektorens tjenesteleverandør must be notified about this.

Noregs miljø- og biovitskaplege universitet

Institutt for datavitskap



Digitalisering på Ås

4 Impact and ethics

4.3 Ethical constraints

4.4 AI ethics considerations

DAT390

27. november 2023

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 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):[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 (we just discussed it)
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]

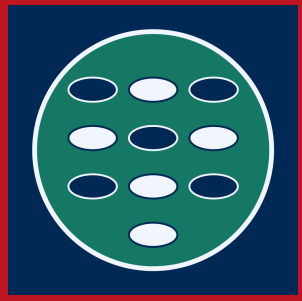
6) Environmental and societal well-being

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

27. november 2023

Example #1

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

Classification confidence visualization of artificial neural networks with adversarial robustness

Abstract: Classification confidence visualization [...] helps us understand how sure a computer is about its predictions [...] using visual tools like heatmaps to show which parts of the input data are crucial for [...] decisions [...]. We propose a bidirectional learning approach for neural networks that [...] can significantly improve the robustness of neural networks to noise and adversarial attacks, and it outperforms [...] deep belief networks and autoencoders. We also introduce hybrid adversarial networks [...] that combine supervised and unsupervised learning [...].

Example #2

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

Application of semi-supervised learning in the process industry

Abstract: This report examines the application of semi-supervised learning (SSL) in the process industry, focusing on […] quality control and operational efficiency in environments with large, partially labelled datasets. The study begins with a literature review on [...] particularly […] industrial settings. [...] The report discusses […] a semi-supervised learning model in a simulated process industry setting, comparing its performance against traditional methods. It also examines pseudo-labelling, a technique central to semi-supervised learning in data-limited scenarios.

Example #3

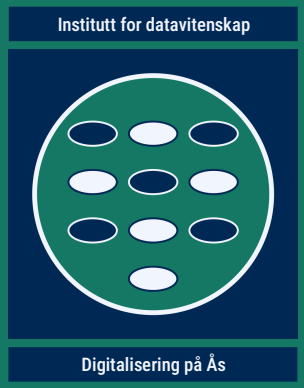
Norwegian University of Life Sciences

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

Evaluating digital tools for welfare assessments of Atlantic salmon

Abstract: [...] study [...] of hyperspectral imaging in aquaculture, an integration of spectroscopy and imaging [...] surpassing the capabilities of traditional colour-based systems. The heightened spectral resolution is leveraged to assess the welfare of Atlantic salmon [...] to compare fish welfare assessments, encompassing human scores, machine learning-based lesion quantification in colour images, and hemorrhaging quantification through spectral analysis. Results are […] in progress.

Norges miljø- og biovitenskapelige universitet



DAT390
Data science seminar

4 Research impact and ethics

4.3 Ethical constraints on the research process
4.4 AI-related recommendations on ethics