Towards a comprehensive assessment framework for the implementation of sustainable AI systems

Sustainable AI Conference

June 15th

Friederike Rohde, IÖW
Marcus Voß, TU Berlin (DAI-Labor)

Supported by:

Federal Ministry for the Environment, Nature Conservation and Nuclear Safety

based on a decision of the German Bundestag

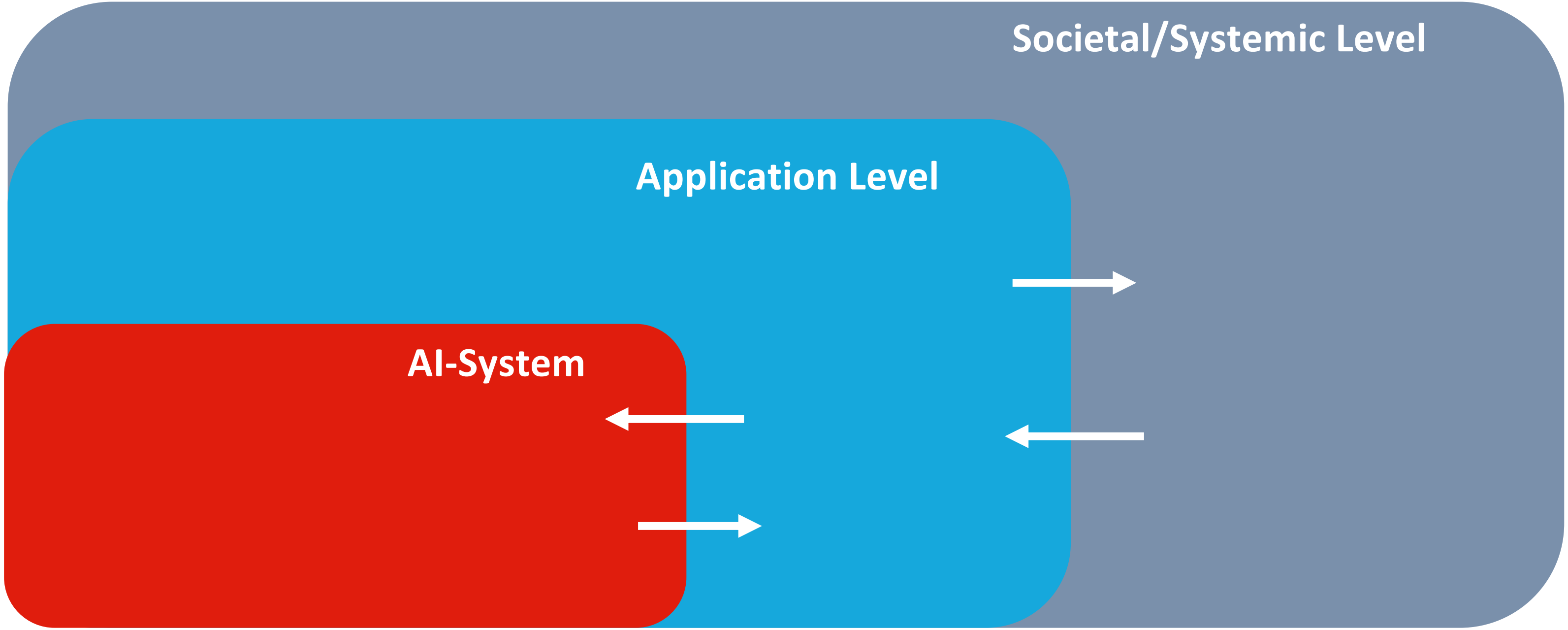
Why do we need sustainable AI systems?

- What part of the AI-Universe do we refer to?
  - Machine Learning & Deep Learning
- AI based-systems are used in wide range of fields (finance, health, industry, jurisdiction, e-commerce) and have impacts on society, our planet, the way we work and on how we make sense of the world
- develop, implement, and use AI in a way that minimizes negative social, ecological and economic impacts *(Rohde et al. 2021)*
- consider the entire lifecycle of AI (design, training, development, validation, re-tuning, implementation and use of AI) (*van Wynsberghe 2021)*

Definition of Sustainability

Social sustainability	**Environmental sustainability**	**Economic sustainability**
Ensure basic needs and inter- and intra-generational equity (Vallance et al. 2011) social cohesion, working and living conditions, institutions, social infrastructures (Littig and Grießler, 2005) cultural aspects and societal norms and values…	Ensure operation within the planetary boundaries, the environmental limits within which humanity can safely operate (Rockström et al. 2009, Steffen et al. 2015)	….work in progress…. Overarching definition Economic activities that ensure the satifactions of needs of current generations without compromising the ability for future generations to satisfy their needs

Systematisation of the impacts



Preliminary set of sustainability criteria

Social criteria

- Transparency & Traceability
- Fairness & Justice
- Harm avoidance
- Non-discrimination
- .....
- Respect for Human Autonomy & Freedom
- Human Oversight
- ...
- Participatory/inclusive design

Environmental criteria

- Energy consumption
- CO2 and greenhouse gas emissions
- Resource consumption, waste and recycling
- Ecological impact of changing production and consumption

Economic criteria

- Market power/market concentration
- Labor market effects & working conditions
- Exploiting innovation potential
- Distribution effect in targeted markets
- ...

Criteria sheets

- Literature review / taking stock of current discourses and approaches
- Criteria sheets for a common definition
- Entail a description of AI-specificity
- Current & possible assessment approaches and/or indicators and possible operationalization methods

Criterion	Non-discrimination
Relevance	AI-based systems can discriminate against people based on age, gender, or skin color, because, among other things, the data used to train the AI contains a bias and reproduces societal prejudices.
AI-specifity	The way in which the data is used for decision-making is specific to AI. This is because in a deep learning algorithm (compared to a "simple" algorithm), decisions are made on the basis of previous decisions without this decision-making being directly influenced by the development team. Thus, the system independently optimizes its decision making. This makes it more difficult to identify causes for discriminatory decisions.
Impact directions and dimensions	**Algorithmic system :** Discrimination by algorithmic decision systems has inherent effects that relate primarily to how the algorithm makes decisions and what data basis is used for the decisions. **Application level :** The extent of the impact of discrimination depends on the area of application and is particularly relevant in areas where AI makes decisions about humans (e.g. AI for applicant selection/HR, AI in the legal system, etc.). **Societal level:** ??
What would be an assessment approach/ How could this be operationalised?	• Cases of discrimination with this AI system have been reported (Yes/No) • Preliminary tests are being conducted (Yes/No) • Information exists about the composition of the development team (Yes/No)
Indicators: What could be used to identify improvements?	• The developing or applying company has responded to reported cases of discrimination and made adjustments (Yes/No)

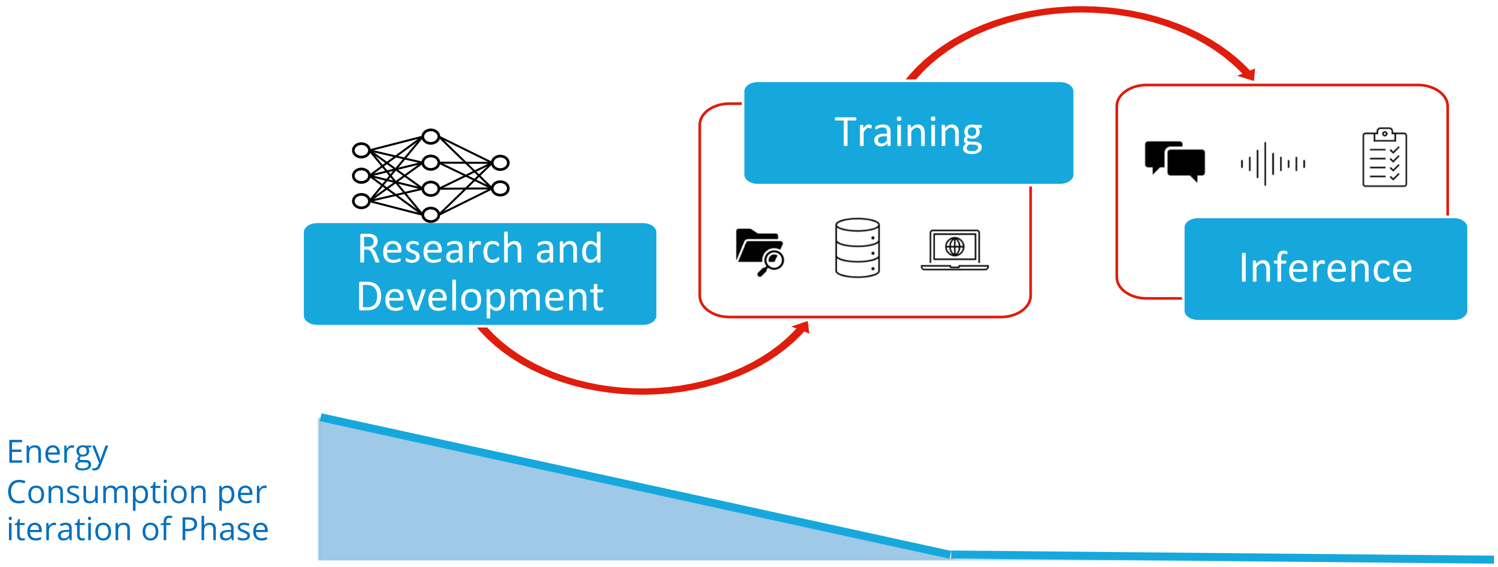
Preliminary assessment approach | exemplary criterion: Non-Discrimination

possible assessment criteria	Maturity Level	Assessment
▪ Diversity of the developer-team ▪ Implementation of pre-testing ▪ Systematic assessment of discrimination risk (per application context) ▪ Systematic monitoring of discrimination cases ▪	**Level 1: Sustainable**	● ● ●
The developing and/or applying organisation has taken systematic measures on its own initiative to minimise the potential for discrimination in advance. These measures include at least:	**Level 2: Proactive sustainability efforts**	● ●
Developing and/or applying organisation has responded to cases of discrimination and developed measures	**Level 3: Reactive sustainability efforts**	●
No indications of measures, preliminary tests or similar.	**Level 4: No sustainability efforts**	

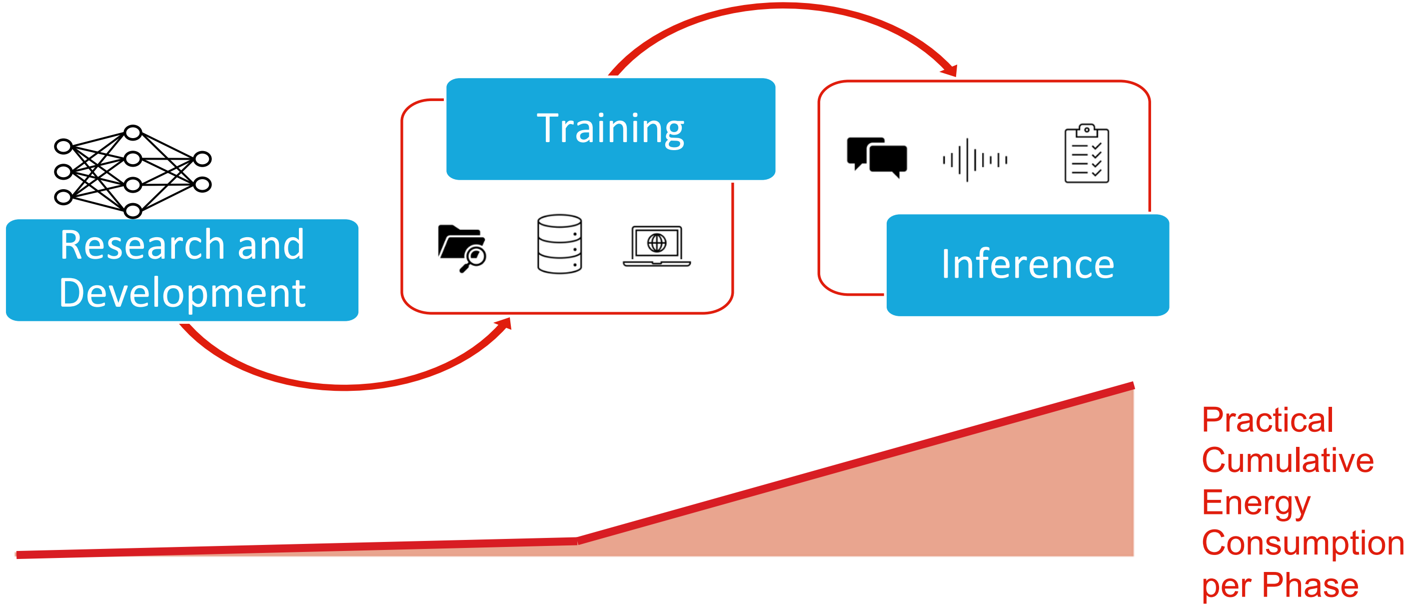
Preliminary assessment approach | exemplary criterion: CO_2-Emissions

possible assessment criteria	Maturity Level	Assessment
• Indirect emissions (Scope 2 and Scope 3, i.e. including embodied emissions) are measured and actively reduced. • Remaining direct and indirect emissions are offset.	Level 1: Sustainable	● ● ●
Direct emissions (Scope 1) are reduced through actively optimizing energy efficiency and using fully carbon neutral energy resources in model development, training and inference.	Level 2: Proactive sustainability efforts	● ●
Direct emissions (Scope 1) from model development, training and inference are measured and offset.	Level 3: Reactive sustainability efforts	●
No measures have been taken to measure, avoid or offset measures emissions.	Level 4: No sustainability efforts	

Energy Consumption in Machine Learning Model Life-Cycle



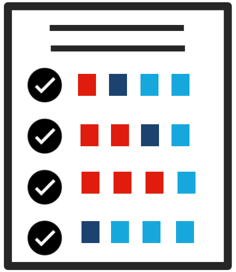
Energy Consumption in Machine Learning Model Life-Cycle



Outlook

Exemplary assessment and validation through ***Case Studies*** *(qualitative and quantitative analysis)*

Sustainability-Index



Sustainable AI Reports

Energy

Mobility

Online-Shopping

- Sustainability Index for Artificial Intelligence (Dashboard/App)
- **Policy Recommendations**
- **Developer Guidelines** for Sustainable AI

References

Kohl, J. L., van der Schoor, M. J., Syré, A. M., & Göhlich, D. (2020, May). Social Sustainability in the Development of Service Robots. In Proceedings of the Design Society: DESIGN Conference (Vol. 1, pp. 1949-1958). Cambridge University Press.

Littig, B. and Grießler, E. (2005), "Social sustainability: a catchword between political pragmatism and social theory", Sustainable Development, Vol. 8, No.1/2., pp. 65-79

Rohde, F., Gossen, M., Wagner, J., & Santarius, T. (2021). Sustainability challenges of Artificial Intelligence and Policy Implications. Ökologisches Wirtschaften-Fachzeitschrift, 36(O1), 36-40.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin III, F. S., Lambin, E., ... & Foley, J. (2009). Planetary boundaries: exploring the safe operating space for humanity. Ecology and society, 14(2).

Steffen, W., Richardson, K., Rockström, J., Cornell, S. E., Fetzer, I., Bennett, E. M., ... & Sörlin, S. (2015). Planetary boundaries: Guiding human development on a changing planet. Science, 347(6223).

van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. AI and Ethics, 1-6.

Vallance, S., Perkins, H. C., & Dixon, J. E. (2011). What is social sustainability? A clarification of concepts. Geoforum, 42(3), 342-348.

Thank you for your attention!

friederike.rohde@ioew.de

marcus.voss@dai-labor.de

sustain
Sustainability Index
for Artificial Intelligence

ALGORITHM
WATCH

i ö w
INSTITUT FÜR ÖKOLOGISCHE
WIRTSCHAFTSFORSCHUNG

DAI-Labor
TU Berlin