







# 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?

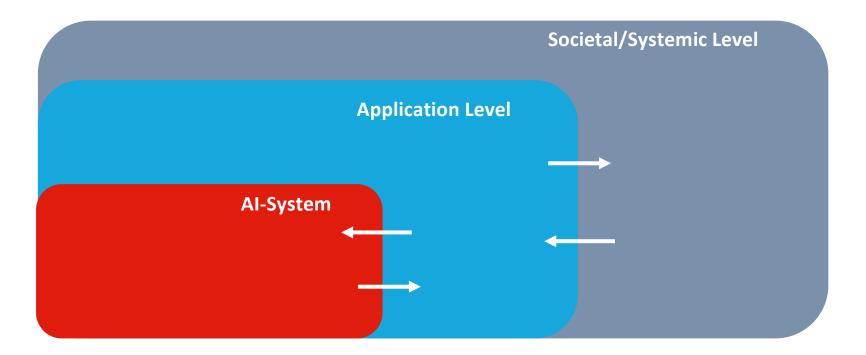
- What part of the AI-Universe do we refer to?
  - Machine Learning & Deep Learning
- AI based-systems are used in wide range of fields (finance, health, industry, jurisdiction, e-commerce) and have impacts on society, our planet, the way we work and on how we make sense of the world
- develop, implement, 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, retuning, 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	Ensure operation within the planetary	work in progress
(Vallance et al. 2011)	boundaries, the environmental limits	Overarching definition
social cohesion, working and living conditions, institutions, social infrastructures (Littig and Grießler, 2005)	within which humanity can safely operate (Rockström et al. 2009, Steffen et al. 2015)	Economic activities that ensure the satifactions of needs of current generations without compromising the ability
cultural aspects and societal norms and values		for future generations to satisfy their needs

### Systematisation of the impacts



# Preliminary set of sustainability criteria

Social	Environmental	Economic
criteria	criteria	criteria
<ul> <li>Transparency &amp; Traceability</li> <li>Fairness &amp; Justice</li> <li>Harm avoidance</li> <li>Non-discrimination</li> <li></li> <li>Respect for Human Autonomy &amp; Freedom</li> <li>Human Oversight</li> <li></li> <li>Participatory/inclusive design</li> </ul>	<ul> <li>Energy consumption</li> <li>CO2 and greenhouse gas emissions</li> <li>Resource consumption, waste and recycling</li> <li>Ecological impact of changing production and consumption</li> </ul>	<ul> <li>Market power/market concentration</li> <li>Labor market effects &amp; working conditions</li> <li>Exploiting innovation potential</li> <li>Distribution effect in targeted markets</li> <li></li> </ul>



#### **Criteria sheets**

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

		100			
	Al-based systems can discriminate against people based on age, gender, or skin color, because, among other things, the data used to train the Al contains a bias and reproduces societal prejudices.				
	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 "impire" 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.	ed in a variety of make predictions nings negative	1		
Impact directions and dimensions	Agentimike system : Discrimination by algorithmic decision systems has informed reflects that reliefs primarily to how the algorithm makes decisions and what data basis is used for the decisions. The present of the system of discrimination Applications that arise of applications and is particularly relevant meres where A manuels decisions about homores (e.g. Alfor applicant selection/H, Al in the legal particularly. Each and the selection/H, Al in the legal particularly.	and accountability systems is that effect"), mous behavior. decisions made	prease in L Hernandez, is area hand, to to reased energy print of ML , 2020). Yet the From 2012 to odem Al	ed CO2 emissions	
What would be an assessment approach/ How could this be operationalised?	Cases of discrimination with this AI system have been reported (tiss/No)     Preliminary tests are being conducted (Yts/No)     Information exists about the composition of the development team (Yts/No)	ed by decisions bility and liability responsibility ected by the	the use of awarear, even to widenced, for within one year rown, et al.,	umption in all life inference). In depend on the he intensity e data center or e CO2 intensity of Arnazon, Google	L
Indicators: What could be used to identify improvements?	<ul> <li>The developing or applying company has responded to reported cases of discrimination and made adjustments (Yes/No)</li> </ul>	uced results is No) onitored (Yes/ <mark>No</mark> )	ference. The imitself (e.g., to sut also to the transformed,	utrailty or even	
	damage (Yes/No) Indicators: What • The developing or applying company could be used to reported cases of discrimination and e identify (Yes/No) improvements2		ergy or example, Al the control of ate models and	e attributed a share sardware. a distributed over d inference. The rithm itself (e.g., to (), but also to the sd, transformed,	and market
	could this be training time.	r learning and pre-trained r (Viss/No) oplied to reduce model par		CO2 emissions are n, Al can also be e CO2 emissions to	logies in a stors will bring mies will thus
	Unnecessary     (Yes/No)	ary re-training and re-training "online" avoided.		i model life cycle. -neutral energy	ing Al models data, the algorithms, and ny of these
	could this be operationalised?	Compensation of			markets. ital companies re based on processing their
	indicators: What crucial be used to identify ingrowments?	<ul> <li>CO2 emissions due to energy consumption in t and development phase, in model training logi data management) and is inference and applic (algorithm and data management)</li> <li>Compensated CO2 quantities</li> </ul>		ing (algorithm and d application	mon et al, data pools in fata acquisition ge digital
		companie of their ap practices, effects for	s with data-driven b plications in the dat or when they use th users, maximise dat	usitess models, whe k about their data co eir market position t ta extraction and agg platforms into a det	n they keep users allection to create lock-in tregate data from

Preliminary assessment approach   exemplary criterion: Non-Discrimination			
possible assessment criteria	Maturity Level	Assessment	
<ul> <li>Diversity of the developer-team</li> <li>Implementation of pre-testing</li> <li>Systematic assessment of discrimination risk (per application context)</li> <li>Systematic monitoring of discrimination cases</li> <li></li> </ul>	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<sub>2</sub>-Emissions

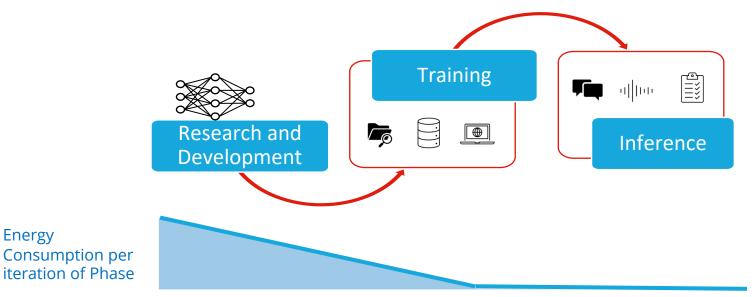


possible assessment criteria	Maturity Level	Assessment
<ul> <li>Indirect emissions (Scope 2 and Scope 3, i.e. including embodied emissions) are measured and actively reduced.</li> <li>Remaining direct and indirect emissions are offset.</li> </ul>	Level 1: Sustainable	
Direct emissions (Scope 1) are reduced 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	



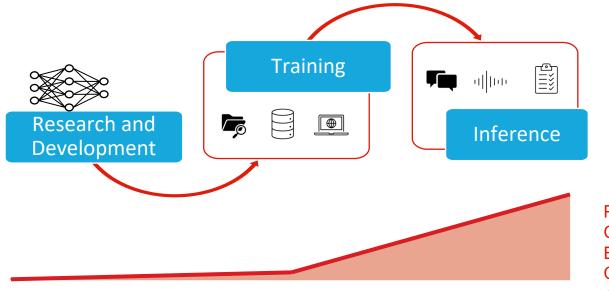
### **Energy Consumption in Machine** Learning Model Life-Cycle

Energy





### Energy Consumption in Machine Learning Model Life-Cycle



Practical Cumulative Energy Consumption per Phase



#### Outlook Exemplary assessment and validation through Case Studies (qualitative and quantitative analysis) Sustainability-Index Sustainable AI Reports Energy **Mobility Online-Shopping** ) E

- 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





