4 Ethics and Sustainable AI

Based on ETHICS GUIDELINES FOR TRUSTWORTHY AI, HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE SET UP BY THE EUROPEAN COMMISSION

Table 6: Ethics

Fundamental rights	No relevance to AI applied to a hydro power engineering problem. No risk identified.
Human Agency and Oversight	Risks could materialise if AI systems run in a closed control loop. The architecture concept is built on always having humans/operators in the loop.
Technical Robustness and Safety	Potential risk includes that the RUL calculated has a larger than expected error margin. The architecture concept includes a human in the loop which is best practice in the hydro power industries. Any major deviation from expected behaviour should be recognized by the human operator.
Resilience to attack and security	System is a) Running in an isolated environment, neither having contact to control systems, nor being connected to the internet b) System is hardened against known attack pattern, authorization mechanisms are implemented
Fallback plan	Hydro power operations can run without RUL estimation
Accuracy	Accuracy of models have been assessed by human specialists

Respect for privacy and data Protection	No personal data is being processes. No risk identified.
Quality and integrity of data	Data is sourced from standard SCADA interfaces
Access to data	Access to data is based on standard APIs, i.e. SCADA and REST
Transparency	Energy is a highly regulated and audited industry. Any life system needs to provide tracking of who accessed the system, when. Authorisation and authentication are important to track in a life deployment. The LexaTexer platform provides functions for authorisation and authentication. However, these will be required only when the system goes into life operations. For the pilot we will focus on getting the technology working to model and predict RUL.
transparency and explainability of the deployed AI	(1) We work on hybrid AI, e.g. combining operational knowledge from the operator with real world measured data points (2) We test and validate the performance of white-box and black-box AI. White-Box AI tends to deliver less accurate results but delivers to some extent an easier access to its inner working, so that again to some extent the functional dependencies can be validated, black-box AI tends to deliver more precise models, however makes it difficult to reach some level of explainability (3) We implement both types of AI to make an informed decision on what to use (4) Additionally we'll have internal as well as external experts to assess the results, particularly in fringe regions of the parameter space.
Diversity, Non-discrimination and Fairness	We don't see the relevance of these questions regarding AI applied to a hydro power engineering problem. We don't hire based on physical features and the above named

	requirements are part of the German Constitution.
Accountability	Energy is a highly regulated and audited industry. Any life system needs to provide tracking of who accessed the system, when. Authorisation and authentication are important to track in a life deployment. The LexaTexer platform provides functions for authorisation and authentication. However, these will be required only when the system goes into life operations. For the pilot we will focus on getting the technology working to model and predict RUL.
Project Management and data access	Data quality and availability is of utmost importance to the success of the pilot. We do have access to one data provider as well as to operational staff in a hydro power plant. We are working on getting access to a second hydro power plant to back up access.