Sustainable Accessible Artificial Intelligence (Sustainable zug.KI) DFKI technical report

DANIEL SONNTAG, HANNES KATH, CHRISTOPH A. JOHNS, and THIAGO S. GOUVÊA, German Research Center for Artificial Intelligence (DFKI) and University of Oldenburg, Germany

The 'Sustainable zug.KI' project, funded by the Lower Saxony Ministry of Science and Culture (2025–2028), aims to develop robust methods to quantify the diverse effects of explainable artificial intelligence (XAI) on sustainability across ecological, economic, and societal dimensions. By incorporating XAI techniques into machine learning, knowledge representation, and intelligent user interfaces, the project strives to facilitate the creation of more sustainable systems. An essential goal is to ensure that the complete AI lifecycle—ranging from development and deployment to usage and reuse—is scrutinised from a sustainability perspective. In response to the escalating demand for AI systems that maintain both transparency and trustworthiness, the project is primarily concentrated on resource-limited areas, such as ecological monitoring. By leveraging interactive machine learning (IML), Sustainable zug.KI intends to generate tools that enable domain specialists, including ecologists, to actively engage in the machine learning process. Moreover, the project will devise efficient methods for transferring AI knowledge, thereby fostering interdisciplinary cooperation and enhancing the practical implementation of AI in sustainability initiatives. Sustainable zug.KI aspires to further both foundational research and real-world applications, contributing to global efforts to fulfil sustainability goals through responsible and explainable AI.

1 INTRODUCTION

The term **Accessible AI** refers to current research topics in basic and applied AI research. This includes the development of approaches that enhance the transparency of decision-making processes, facilitate human comprehension of AI-generated outputs, and ensure the reliability and ethical alignment of AI-based solutions.

From the perspective of AI research, these topics have a key requirement: AI systems must be explainable, interpretable, and trustworthy [Marcinkevics and Vogt 2023]. While these concepts partially overlap, they contribute to an integrated perspective, particularly in the context of combining symbolic representations with complex neural architectures. The differentiation of these terms in the literature reveals diverse approaches. Interpretability primarily concerns the understanding of AI models, including their comprehensibility and traceability [Marcinkevics and Vogt 2023; Rudin 2019]. Explainability, in contrast, focuses on generating explanations for the conclusions or actions of such systems. Interpretability is often regarded as a passive characteristic of a system, meaning that its operations can be understood by humans either through direct introspection or by producing explanations [Biran and Cotton 2017]. This property is particularly associated with inherently interpretable models [Rudin 2019], which allow users to directly trace how inputs map to outputs, offering full transparency, such as decision trees, linear/logistic regression, or rule lists. Interpretable models that are not fully transparent include support vector machines or Bayesian networks, which are interpretable (post-hoc) because of the concept of their graphical structure (a directed acyclic graph) that explicitly shows the probabilistic relationships and conditional independencies between variables, respectively.

Explainability, on the other hand, is considered a more active characteristic, aimed at making a system's functioning clear and easy to understand [Arrieta et al. 2020]. This is often achieved through post-hoc methods [Rudin 2019] and it requires approaches from human-computer interaction to offer human-understandable justifications for specific end-users.

Authors' Contact Information: Daniel Sonntag, daniel.sonntag@dfki.de; Hannes Kath, hannes.kath@dfki.de; Christoph A. Johns, christoph_albert.johns@dfki.de; Thiago S. Gouvêa, thiago.gouvea@dfki.de, German Research Center for Artificial Intelligence (DFKI) and University of Oldenburg, Oldenburg & Saarbrücken, Germany.

Ultimately, these aspects contribute to the trustworthiness of AI systems [Floridi 2019; Wing 2021], which encompasses explainability, interpretability, and transparency. Trustworthiness ensures that users can gain insights into system modelling and procedures, quantify uncertainties, and verify AI-driven processes.

In addition to the fundamental research questions arising from AI development, there is a crucial need to address the **dissemination of AI knowledge** across society, industry, and policymaking. To facilitate informed discussions on AI governance, enable well-founded decisions regarding its profitable application in products and processes, and establish new regulatory frameworks where necessary, significant and continuous efforts are required. These efforts must focus on transferring AI knowledge to the business sector, fostering dialogue with society, and enhancing AI-related education, training, and professional development. Such initiatives are essential to ensure that stakeholders across various domains possess the necessary expertise to engage with AI technologies responsibly and effectively.

Both dimensions—research and dissemination—are encapsulated here under the term **Accessible AI**. AI methods must inherently allow for explainability, interpretability, and trustworthiness from a scientific and technical standpoint. At the same time, the transfer of AI knowledge must ensure that AI is effectively communicated to society, industry, and policy.

This project, Sustainable zug.KI, complements the objectives of the IML (Interactive Machine Learning) research area by incorporating sustainability aspects. In the IML research area, the foundations of intelligent algorithms and user interfaces are developed, enabling machine learning through human interaction, either via natural dialogue or learning from observational data. The primary focus is on methodological research, with key areas of transfer including applications in ecology, medicine, and industry.[1]

In a world of increasingly scarce resources, unreliable energy supply, and increasingly fragile supply chains, particularly for hardware, universities, research institutions, and industry will continue to provide more powerful AI systems. However, this also means that in the future, the consumption of resources must be justified on a much larger scale. This is accompanied by corresponding legislation at the EU level. The EU aims to achieve climate neutrality by 2050, creating an economy with net-zero greenhouse gas emissions. This goal is at the heart of the European Green Deal [European Union 2020] and, thanks to the European Climate Law [European Union 2021], it has become a legally binding objective.

However, AI institutes are currently not in a position to adequately meet these requirements. At present, AI practitioners lack systematic tools to justify the use of AI in general, the decision against using AI, or the selection of a particular AI system for a given problem based on a solid cost-benefit analysis within the context of sustainability. This has become a pressing issue since Large Language Models (LLMs) became mainstream AI technologies in 2025. Furthermore, there is a lack of awareness regarding the significant resource consumption. Estimates suggest that the training phase of an LLM consumes as much energy as 500 German households do in a year (estimates based on [Patterson et al. 2021]). Data centres and network infrastructure already account for a non-negligible share of global electricity demand and associated emissions, underscoring the environmental relevance of AI and cloud-based computing.

The challenge lies in the development of methods that comprehensively quantify the impacts of AI technologies on sustainability throughout their entire lifecycle, encompassing the phases of development, deployment, use, and, where applicable, reuse. This requires a holistic approach to measuring resource consumption and assessing the impacts of AI across ecological, economic, and societal dimensions.

[1] See iml.dfki.de for example.

The project **Sustainable zug.KI** aims to develop robust methods for quantifying the diverse impacts of explainable artificial intelligence (XAI) on sustainability across ecological, economic, and societal dimensions. In doing so, this project contributes significantly to achieving the Sustainable Development Goals (SDGs) [United Nations 2015]. We focus on **Computational Sustainability (CS)** and **Explainable AI (XAI)**. Specifically, we leverage XAI methods in the fields of machine learning, knowledge representation, and intelligent user interfaces to contribute to more sustainable systems (**XAI for CS**) and to the development of more sustainable XAI systems (**CS for XAI**). To make XAI technologies more sustainable, the entire lifecycle of development, deployment, usage, and reuse, among other factors, must be taken into account. Currently, there are no established methods to measure and quantify the impacts of XAI throughout the lifecycle in the aforementioned dimensions.

2 PRELIMINARY WORK

Research in the Applied Artificial Intelligence (AAI) department[2] and DFKI [3] focuses on the application and adaptation of artificial intelligence techniques to fields such as industry[4] and medicine[5].

The application aspects relevant to the research primarily involve the utilisation of systems based on machine learning and intelligent user interfaces. The main focal points encompass multimodal input and output, as well as environment and state detection, sensor data processing, and issues related to real-time capabilities and interactivity when learning from extremely large or extremely small datasets. Moreover, aspects of reliability, including trust in AI and explainable AI, are deemed significant.

Sustainability topics and engagement have a long-standing tradition in the teaching, research, and operations at the Carl von Ossietzky University of Oldenburg. The focus on sustainability in this project evolves around tangible AI-related questions.[6]

In addition to **sustainability** itself, the socio-technical aspect of explainability plays a central role in the context of predictions: **Interactive Machine Learning and Explainable AI, two sides of the same coin**. Overall, the emphasis on XAI can be illustrated as the intersection of machine learning, end-user explanations and human-computer interaction (figure 1):

Recent research contributions of AAI and DFKI IML can be identified in these subfields as follows:

Machine Learning. This primarily includes the so-called image captioning task and other aspects of (medical) image processing, as well as sensor data and video processing, and theoretical topics such as self-supervised learning [Barz and Sonntag 2019; Biswas et al. 2021, 2019; Nguyen et al. 2020, 2022a, 2023, 2022b; Nunnari et al. 2020, 2021a,b].

Human-Computer Interaction. With regard to multimodal multi-sensor interaction, the focus is primarily on work in eye tracking, attention recognition, and multimodal demonstrators of complete systems, along with 'Human Factor' topics [Barz et al. 2018, 2023, 2021; Barz and Sonntag 2021; Barz et al. 2020; Bittner et al. 2025; Gouvea et al. 2023; Kopácsi et al. 2023, 2024, 2026; Le et al. 2025; Prange and Sonntag 2021b; Sonntag and Profitlich 2019].

End User Explanations (XAI) / User Modelling. Published works primarily focus on passive and active user modeling during stylus interaction, a publicly available software toolbox for XAI, and user interfaces for specific

[2] https://uol.de/aai
[3] https://dfki.de/iml
[4] https://smartfactories.dfki.de/
[5] https://ai-in-medicine.dfki.de/
[6] The sustainability research profile in Oldenburg is being enhanced through doctoral programs, supplementary projects, and involvement in international initiatives. The Sustainable zug.KI project lends support to these programs.

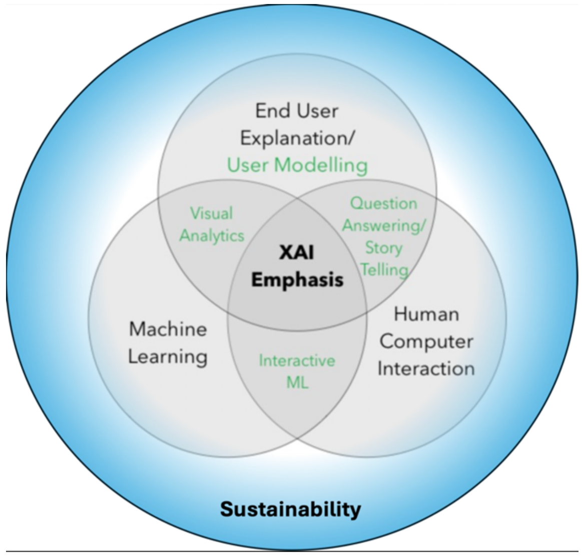

Fig. 1. Explainable AI sits at the intersection of machine learning, end-user explanation/user modelling and human-computer interaction.

XAI end-users, as well as NLP narratives for the natural language generation of domain-specific explanations for the decisions of an ML system [Alam et al. 2025; Biswas et al. 2020; Hartmann et al. 2022; Kadir et al. 2023; Leist et al. 2025; Nunnari et al. 2021a; Prange and Sonntag 2019, 2021a, 2022]

2.1 Computational Sustainability & Technology

The research field Computational Sustainability & Technology[7] combines theoretical foundational research with practical international application projects. Computational Sustainability (CS) is the scientific domain that aims to harmonise societal, economic, and ecological resources using methods from computer science and artificial intelligence. AI models, such as machine learning models, enhance computational sustainability models. Research can make significant contributions to addressing key sustainability challenges (AI for CS). Questions related to computational sustainability also enrich AI research as a whole, not only by raising issues related to vagueness or uncertainty, thus presenting new AI challenges, but also by providing a framework of requirements for resource-bound computations (CS for AI).

A central pillar of CST research is interactive, AI-supported **wildlife monitoring**. Passive acoustic monitoring (PAM) has emerged as a powerful, non-invasive approach for biodiversity observation, enabling continuous recording of environmental soundscapes and providing insights into animal behaviour, species richness, and ecosystem health [Ross et al. 2023; Sugai et al. 2019]. While PAM allows large volumes of ecological data to be collected with minimal disturbance, its widespread adoption is limited by the effort required to analyse and annotate massive, largely unlabelled audio datasets. AI—and in particular interactive machine learning (IML)—offers a means to address this bottleneck by combining scalable machine learning models with expert knowledge, explainable representations, and iterative human feedback.

[7] https://cst.dfki.de

Against this backdrop, the first international CST project of the IML research area, Amber[8] (Interactive Machine Learning Solutions for Acoustic Monitoring of Animal Wildlife in Biosphere Reserves), has established a methodological and organisational foundation for human-in-the-loop biodiversity monitoring using PAM. Building on this foundation, both research aspects and practical application aspects, including prototype development in applied contexts, will be promoted through ongoing doctoral projects within Sustainable zug.KI (personnel resources, material resources, investments). These activities draw on and extend results obtained under real-world conditions, including the validation of interactive learning workflows during the XPRIZE Rainforest competition, where our contribution as part of the Brazilian Team achieved a Top-3 placement[9]. Within the CST roadmap, this work represents a step toward sustaining, scaling, and disseminating interactive AI methods for long-term biodiversity monitoring across biosphere reserves and related conservation contexts [Gouvea et al. 2023].

Beyond wildlife acoustics, CST pursues a second flagship application in **coral morphology analysis**. Coral reefs are among the most diverse and ecologically significant marine ecosystems, yet understanding how environmental factors influence coral growth remains challenging. Laboratory-based experiments—where parameters such as temperature, light, and water chemistry can be systematically controlled—play a crucial role in isolating causal relationships relevant to climate change impacts and associated biodiversity loss. However, the analysis of laboratory-generated photographic and volumetric coral data remains labour-intensive, limiting experimental throughput and reproducibility.

To address this bottleneck, we develop interactive, cloud-based analysis workflows that combine computer vision methods with expert-driven refinement for two- and three-dimensional laboratory coral imagery. Pre-trained segmentation and detection models provide scalable initial analyses, while interactive annotation and adaptive learning enable researchers to correct errors, refine model behaviour, and iteratively improve performance on specialised datasets. These methods support the systematic study of growth patterns and morphological responses under controlled environmental conditions, thereby accelerating experimental research on coral resilience and vulnerability. Beyond marine ecology, this work also serves as a methodological bridge to medical image analysis: coral morphology data and research questions share structural similarities with medical imaging, while regulatory constraints are substantially lighter. As such, this domain provides a valuable testbed for interactive machine learning methods that can later inform medical AI research under stricter regulatory regimes.

A third CST focus area addresses AI-supported **sustainability reporting**, with a particular emphasis on Life Cycle Assessment (LCA). Increasing regulatory demands, such as those emerging from European sustainability frameworks, require organizations to produce transparent, consistent, and continuously updated sustainability assessments. CST research explores how interactive knowledge systems, ontology-based representations, and agentic AI can support the integration of Life Cycle Inventories, environmental impact metrics, and qualitative sustainability criteria into coherent reporting workflows. By embedding LCA as a core analytical backbone, these tools aim to transform sustainability reporting from a periodic compliance task into a dynamic, data-driven decision-support process. From a CS-for-AI perspective, this line of work also provides a concrete framework for assessing the environmental costs and benefits of AI systems themselves across their lifecycle.

Across all three pillars, CST places strong emphasis on transfer, networking, and the provision of prototypes on-site. The existing international network spans research and practice across multiple ecological and geographical contexts. These collaborations ensure that methodological advances in interactive machine learning and explainable AI are continuously grounded in real-world sustainability challenges and informed by domain expertise. Through this network,

[8]https://cst.dfki.de/amber
[9]https://iml.dfki.de/news/iml-among-the-winners-of-the-10-million-xprize-rainforest-contest/

CST research emphasises iterative validation in applied settings, long-term engagement with stakeholders, and the deployment of prototypes in operational environments, thereby strengthening both scientific relevance and transfer potential.

3 SPECIFIC RESEARCH TOPICS

Regarding the ongoing dissemination of AI, a particular DFKI project should be highlighted, as it sets the transfer framework for the work in Sustainable zug.KI: XAINES, explaining AI with narratives (BMBF, 2020-2024). In the XAINES project [Hartmann et al. 2022], not only is explainability ensured, but explanations (narratives) are also provided. The central question is whether the AI system can explain in a sentence why it made a certain decision or if it needs to explain it interactively to the user.

The ongoing doctoral projects focus on making AI systems accessible, trustworthy, and effective for sustainability-oriented domains. All theses align with one of the three CST flagship areas: AI-supported wildlife monitoring, coral morphology analysis, and sustainability reporting.

DP1: Interactive Active Learning for Acoustic Wildlife Monitoring. This thesis investigates active learning strategies for large-scale passive acoustic monitoring, aiming to minimize annotation effort while maximizing ecological insight. By identifying the most informative audio segments for expert review, the work develops interactive learning loops that adapt model behaviour based on sparse, high-value feedback. The research contributes methods for uncertainty-aware sampling and expert-in-the-loop model refinement, and explainable representations tailored to ecological decision-making [Kath et al. 2024a, 2025, 2026, 2024c].

DP2: Visual Interaction and Explainable Representations in Bioacoustics. This thesis investigates how visual analytics and interactive interfaces can support the interpretation of large, high-dimensional soundscape datasets in bioacoustic research. It focuses on the design of interactive analysis pipelines that allow domain experts to flexibly explore and compare different acoustic representations, ranging from hand-crafted acoustic indices—familiar, interpretable, and inexpensive to compute—to learned audio embeddings that capture more complex structure. Rather than integrating these representations into a single feature space, the work emphasises rapid prototyping and interactive pipeline building, enabling experts to swap representations, dimensionality reduction techniques, and clustering methods to suit specific research questions and data characteristics.

These pipelines are embedded in interactive multimodal interfaces that allow ecologists to explore representations, inspect clusters, and iteratively guide analysis decisions based on ecological expertise rather than purely model-driven criteria. In doing so, the approach complements active learning not only during the cold-start phase, when labels and target variables are uncertain, but also in later stages by placing experts in a more active role than that of a passive oracle responding to algorithmic queries. By treating representations and clustering methods as explorable components of an analysis pipeline—rather than as fixed preprocessing steps—the work supports explainability, expert-driven sensemaking, and reproducible early-stage analysis. The resulting designs will inform CST tools such as *YAPAT*[10] and the *EcoScape Analyzer*, and are being evaluated with domain experts to assess usability and effectiveness in real-world bioacoustic workflows [Kath et al. 2023, 2024b; Saghir et al. 2025].

DP3: Weakly Supervised Learning for Ecological Sound Analysis. This thesis addresses the scarcity of strongly annotated training data in bioacoustics by developing weakly supervised methods for sound event detection and temporal

[10] https://yapat.readthedocs.io/

localisation in recordings from curated sound collections. Rather than relying on interactive exploration or query-based annotation strategies, the focus is on algorithmic approaches that transfer knowledge from existing wildlife sound libraries and museum collections—where recordings are typically curated at the level of species presence but lack precise temporal annotation—to large-scale, unlabelled PAM datasets. By leveraging transfer learning and multiple instance learning, the work demonstrates how information encoded in curated sound collections can be used to derive fine-grained temporal structure in continuous field recordings, substantially reducing dependence on costly expert annotation. The resulting methods complement active learning and visual analytics workflows by providing automatically generated, temporally resolved labels that can serve as an initial training signal or downstream input for expert-driven refinement [Troshani et al. 2024a,b].

DP4: Interactive Machine Learning for Coral Morphology Analysis. This doctoral project develops human-in-the-loop computer vision workflows for analysing two- and three-dimensional coral imagery. By combining automated segmentation with expert-driven correction and adaptive learning, the research improves efficiency and reproducibility in coral morphology studies. Methodologically, the thesis positions coral ecology as a bridge domain whose data characteristics closely resemble medical imaging, enabling transferable insights for interactive medical AI under lighter regulatory constraints.

DP5: AI-Supported Sustainability Reporting and Life Cycle Assessment. This thesis focuses on interactive AI systems for sustainability reporting, with LCA as a core analytical component. The research develops ontology-based knowledge representations and agentic AI workflows to integrate heterogeneous data sources, automate consistency checks, and support scenario analysis. By embedding human feedback into reporting pipelines, the work aims to improve transparency, traceability, and trust in AI-supported sustainability assessments, while also enabling evaluation of the environmental impacts of AI systems themselves.

The following list provides examples of additional specific research topics:

(1) Interactive learning of structured and hierarchical representations from sensor data, including disentangled and ontology-informed latent spaces.
(2) Use of learning systems and intelligent user interfaces for interactive learning incorporating XAI aspects and sustainability.
(3) Sensor data processing and issues of real-time capability and interactivity in learning with very large or very small datasets, including reliability and sustainability aspects.
(4) Adaptation of artificial intelligence methods in interactive machine learning under XAI and sustainability considerations, applied to industrial and medical applications (as secondary transfer domains).
(5) Sustainable learning systems with continuous adaptation mechanisms and feedback loops for explanations.
(6) IML application systems in the field of sustainability, particularly international biosphere reserves.
(7) Computational sustainability and technology transfer with a focus on trust aspects.
(8) Information extraction from medical texts and images with a focus on trust aspects (as secondary transfer domains).

4 WORK PLAN

WP1: Project Coordination and Cross-cutting Themes; **Outcome**: Coordination with the projects of technological subject matter experts. Public outreach through publications, presentations, and joint demonstrations of project results.

WP2: Requirements Analysis; **Outcome**: Specification of requirements and bottlenecks through structured analysis of the (HCI) workflows of sustainability domain experts. Research on state-of-the-art AI systems and programs aimed at enhancing workflow efficiency, with a focus on robust methods for quantifying the diverse impacts of explainable artificial intelligence (XAI) on sustainability across ecological, economic, and societal dimensions.

WP3: Implementation and Evaluation of a Visualisation Component for Relevant Data and Decision Processes of ML Models to Increase Transparency and Sustainability; **Outcome**: A component for the iterative development and testing of programs and algorithms to improve workflow efficiency and ML training processes, in collaboration with domain experts. This component will be intuitive and usable without programming expertise, providing interpretable results for domain experts. Additionally, a visual tool for data acquisition, data annotation, and data quality enhancement will be implemented.

WP4: Implementation and Evaluation of an XAI Interface that Actively Integrates Experts into the Learning Process and Explains Decision-Making Aspects of Sustainability; **Outcome**: A component to integrate elements that explain the results of the AI system to the user while simultaneously giving the user decision-making authority over the system. This will include new methods for measuring and quantifying the impacts of XAI.

WP5: Implementation and Evaluation of an ML Component to Enhance ML Models with Explainability and Sustainability by Relating ML Predictions to Standardized Concepts and Guidelines from the Application Domains; **Outcome**: Component for analysing available datasets and model structures. Training of a large model for the generation of representations of, for example, passive acoustic audio data and other multimedia data such as images and videos, including reuse strategies, or cost-benefit justification.

WP6: Enabling Accessibility to the Applications; **Outcome:** Development of a robust demo application to enhance the efficiency of workflows for subject matter experts, which is intuitive and self-sufficient for individuals without programming knowledge, offers possibilities for expansion, and is well-documented; Evaluation of trustworthiness in a user study.

5 HUMAN-COMPUTER INTERACTION RESEARCH

The two angles to tackle sustainability issues in and with XAI—XAI for CS and CS for XAI—imply several tasks that fall under the umbrella term of human-computer interaction (HCI) research. To effectively support work that contributes to more sustainable use of scarce resources, the needs and practices of relevant user groups, such as ecologists, must be thoroughly understood. HCI research has developed various methods to facilitate such understanding, such as participatory and co-design, observation, and interviews [Hornbæk et al. 2025]. These methods help identify, design, and deliver appropriate XAI tools and measure their effect on domain experts' workflows. They thus ensure that the goal of the Sustainable zug.KI project—to deliver robust methods for quantifying the multifaceted impacts of XAI on sustainability—remains rooted in human values and domain experts' practices.

To structure these HCI research activities, we draw upon the established design process model by Hornbæk, Kristensson, and Oulasvirta [Hornbæk et al. 2025], which conceptualises HCI research as a process of moving from understanding the present (through user research and analysis of interaction) to creating a desired future (through design, engineering, and evaluation).

5.1 Exploration of the Problem Space

The initial activities in HCI research relate to understanding the present, primarily covering *User Research* and *Understanding People* [Hornbæk et al. 2025]. The goal of these activities is to develop a thorough understanding of the context in which the newly developed AI and XAI methods are to be used. This implies an exploration of the problem space without preconceived solutions.

XAI for CS. To effectively support domain experts such as ecologists with AI and XAI technology, their activities around sustainable practice must be carefully observed and analysed. It must be understood how they collect, manage and analyse data (e.g., passive acoustic audio data) and how they derive decisions from it. Methods from Participatory Design [Bødker et al. 2022], such as contextual inquiry or co-design workshops, as well as interview and observation studies [Hornbæk et al. 2025], can help foster such an understanding. For example, an observation and interview study into ecologists' data analysis and modelling practices to monitor wildlife with bioacoustic data—that is, which tools, motivations, norms, collaboration habits, *etc.* shape how they process and interpret bioacoustic data to identify diversity and prevalence of wildlife and how they make decisions around preservation on their basis—can shed light on the breakdowns, barriers and contradictions in those domain experts' current use of technology. Such studies are necessary to recognise and understand underserved needs for analysis and communication that can be supported with AI. They further ensure that AI and XAI methods developed in the project preserve and amplify the way in which domain experts exercise their expertise. Exemplary outcomes of such activities are design guidelines for later phases of the development and assessment process or, in the case of early-stage co-design activities, concepts and sketches for technological solutions. Understanding experts' established practices—as well as their initial perceptions of AI—through exploration activities is a requirement for designing AI systems that earn domain experts' trust and that facilitate their activities.

CS for XAI. We must understand the workflows and decision-making rationales of AI practitioners to empower them to make more informed decisions regarding scarce resource use: What factors currently drive their choice of model architecture or training regimen? Which role do resource consumption and sustainability play in their decision-making processes? A human-centred or participatory approach that involves AI experts, such as data scientists or ML engineers, in the design process can help reveal the practical barriers to adopting more sustainable practices. For instance, joint observation and interview studies could pinpoint difficulties MLOps and LLMOps engineers face in locating or interpreting existing resource consumption metrics in industry-standard dashboards and documentation (e.g., *W&B*[11], *MLflow*[12]) when deciding upon cloud resources for deployment or how long a model will remain active [Tamburri 2020]. Similarly, the Product Requirements Documents (PRDs) or *Jira* tickets written by AI product managers (e.g., a product manager deploying a chatbot for an e-commerce app) could be analysed regarding sustainability aspects, for example, how commonly requirements related to energy efficiency are stated. Cultural probes could be used to gather rich, unsolicited insights into the daily priorities, concerns, and collaborative practices of AI teams, revealing the latent norms that may deprioritise sustainability. For example, a 'cut corners' diary could be distributed to AI engineers working in a consultancy firm to capture micro-decisions made during coding and meetings that record and let the AI practitioners reflect on when and why their work negatively affected energy consumption. Finally, the tools that AI practitioners use themselves could be studied, such as training and evaluation dashboards like *W&B*[13] or *Comet ML*[14]. Such analysis can reveal how those tools integrate and represent sustainability aspects of ML development and deployment and thus reflect on and conversely shape sustainability practices in AI development. These methods can help create a shared understanding, for example, whether it is a lack of awareness, a lack of tools, or institutional pressure

[11] https://wandb.ai/
[12] https://mlflow.org/
[13] https://wandb.ai/
[14] https://www.comet.com/

that deprioritises sustainability efforts. User research thus provides the necessary grounding to design interventions that can reshape established AI and XAI workflows.

5.2 Iterative Development and Assessment

The second set of activities in HCI research relates to creating a desirable future, covering *Design*, *Engineering* and *Evaluation* [Hornbæk et al. 2025]. The goal of these activities is to develop concepts and designs that address the needs identified in the previous activities and to turn them into practical systems. This is an iterative process of building, refining, and evaluation which tightly connects the findings of the previous activities to utility for the domain or AI experts.

XAI for CS. A key notion for the design of AI systems that support sustainability activities is trustworthiness. Research must recognise those factors that contribute to a fair assessment of the AI's ability (i.e., performance) and integrity (i.e., adherence to the user goals) [Hoff and Bashir 2015] in the given application context. For example, it must identify those aspects that ecologists look out for to understand when and to what degree they can rely on the classification of bioacoustic signals into different species and those XAI techniques that are particularly effective at helping them diagnose system strengths and shortcomings in that regard. XAI systems must thus help users form an accurate model of the AI's capabilities, fostering what is known as *calibrated trust* [Zhang et al. 2020]: trusting the system when it is competent and distrusting it when it is not. To achieve this, AI systems must be reliable, but also transparent and support users in understanding their actions and limitations. XAI methods are central here. The interaction between a domain expert and an AI system should establish common ground between the user and the machine. It must ensure that the feedback provided by the human is correctly interpreted by the AI, and that the AI's suggestions are comprehensible to the human. This communication, including the vocabulary and sharing practices, is highly likely to be use case-specific. For example, managers of conservation areas employ different terminology and look for different metrics than ecology researchers or regulators to analyse passive acoustic audio data, although all need to regularly collaborate to make decisions that affect wildlife preservation. The development of XAI methods and tools for sustainability thus requires evaluation studies directly involving and engaging with the relevant user groups. It also requires HCI activities, for example, related to prototyping, to explore alternative solutions and develop guidelines and principles for effective XAI. For example, empirical experiments can be used to compare established XAI methods with new explanation mechanisms that more tightly integrate domain-specific knowledge, for example, by basing explanations on an ecology-specific ontology-informed embedding space. If one method establishes a more accurate understanding of a system's capabilities (e.g., of a system's confidence range or error rate), it should be preferred. Similarly, to assess the practical utility and understand the nuances of the interaction, think-aloud protocols with domain experts, such as ecologists using an XAI system to label bioacoustic data for different animal species, could reveal mismatches between their mental models and the system's output. These activities can inform the iterative design of new interactive visualization and machine learning tools to support data and communication needs of sustainability practitioners.

CS for XAI. To support sustainability efforts in AI, sustainability must become a core component in the AI and XAI development workflow. HCI research efforts will be required to identify those interaction and visualisation patterns that effectively communicate the resource costs associated with different XAI methods. It can then help guide users toward more sustainable choices without hindering their work. The goal here should be to help users develop an intuition for the sustainability impact of their actions over time. This will empower them to make informed decisions about their

XAI needs and their ecological implications without prescribing one-size-fits-all methods. To achieve this, the interfaces that communicate such impact and help AI experts explore it (e.g., IDE plugins, MLOps/LLMOps dashboards) should provide salient, actionable feedback on resource consumption. At the same time, they must be designed in a way to empower AI practitioners to make informed choices regarding resource consumption while fitting into their established development and collaboration practices. HCI activities, such as prototyping or co-design workshops, can help quickly progress towards such system designs. For example, a participatory co-design workshop with Data Scientists and ML Engineers who regularly use XAI methods and libraries such as *SHAP* [Lundberg and Lee 2017] or *LIME* [Ribeiro et al. 2016] in their workflows could identify interaction and visualisation patterns that effectively communicate resource costs and fit with domain norms around ML development and deployment. Similarly, native prototyping could be used to develop new IDE plugins (e.g., based on existing energy profiling tools like *MANAi*[15]) to explore interaction design patterns such as tooltips, colour coding, or sidebars that can effectively convey resource consumption of XAI analyses to junior to mid-level AI developers who may not yet have developed a strong intuition for the environmental cost of model training. The consequent evaluation of such systems must consider usability, for example, using standardised task load and usability questionnaires such as NASA-TLX [Hart and Staveland 1988] and SUS [Brooke 1996]. But, more importantly, it must measure their effectiveness in causing a shift in developer attitudes and behaviour and practices towards more sustainable AI usage, which can, for example, be recognised through *in situ* deployment studies [Hornbæk et al. 2025].

5.3 Activities in the Work Packages

The general HCI activities outlined above map to the work packages from section 4 to yield specific HCI tasks to be completed as part of the Sustainable zug.KI project. Since *WP1: Project Coordination and Cross-cutting Themes* refers to general outreach and dissemination activities, we focus on *WP2* through *WP6* here.

5.3.1 WP2: Requirements Analysis. The goal of *WP2* is to identify breakdowns in how (X)AI systems are currently used by sustainability experts, such as ecologists, and to translate them into requirements for new (X)AI systems. These should support sustainability-related activities, such as the analysis of passive acoustic monitoring data, while adhering to established practices in the application domain.

Key activities in support of this goal include:

- Identification of stakeholder groups and profiles (e.g., ecologists, sustainability analysts, data scientists, policy makers) through interviews and field study methods, such as contextual inquiry, activity analysis, and artifact analysis,
- Analysis of AI lifecycle workflows—spanning data collection, annotation, model training, interpretation, and reporting—to identify breakdowns and cross-role hand-offs (e.g., between people with different expertise, such as between ecologists and data scientists), for example, through Hierarchical Task Analysis (HTA), and journey mapping,
- Definition of key evaluation criteria and performance indicators, for example, related to usability, cognitive load, efficiency, trust, or energy, data, and compute demands, to direct the development and evaluation, and

[15] https://github.com/aschuler84/MANAi

- Definition of design guidelines and structured requirements [Sonntag et al. 2010] for the development of new XAI and active learning methods, desktop, mobile, or extended-reality applications, or adaptive interfaces—including requirement analysis for 2D desktop/tablet vs. 3D VR/XR interfaces to identify when spatial data exploration or immersive review can be beneficial [Saghir 2025].

Within the doctoral projects outlined above, such activities serve to answer questions such as *Which forms of ecology knowledge do current active learning strategies fail to consider?* (DP1) or *How can representations and interactive visualisations best support collaborative sense-making around PAM data?* (DP2).

5.3.2 WP3: Visualisation of Data and Decision Processes in Machine Learning Models. The goal of *WP3* is to develop one (or multiple) software applications to let sustainability domain experts, such as ecologists, interactively explore, manipulate, and control their data as well as their machine learning models. This includes activities such as viewing and labelling data, and monitoring, inspecting, and directing model training and outputs. In contrast to standard data analysis and machine learning tools, the focus of the software tools developed in this project lies on the sustainability-related aspects of data curation and model training—both in terms of resource consumption (e.g., energy profiles; cf. *CS for XAI*) and mapping to vocabulary and concepts from the sustainability-related application domains (e.g., biophony/anthrophony metrics for PAM data; cf. *XAI for CS*, *WP5*).

Key HCI activities in support of these goals include:

- Interaction design of 2D or 3D visual analytics tools and dashboards for dataset overview, provenance, curation, and quality assurance, as well as review of model output and performance,
- Interaction design for iterative "inspect, correct, retrain" loops to facilitate interactive model training without programming expertise,
- Formative evaluation combining heuristic evaluation with cognitive walkthroughs to assess learnability for non-programmers performing core tasks, such as identifying data quality issues, labelling samples, and interpreting model outputs, and
- Summative evaluation to measure transparency effects.

Within the doctoral projects outlined above, these activities serve to answer questions such as *Which visual encoding strategies direct ecologists' attention to the most informative samples, for example, the labels most likely to be wrong?* (DP2) or *Can ecologists effectively (re-)configure the newly designed visual analytics tools for their practical analysis needs?* (DP2).

5.3.3 WP4: Expert-in-the-Loop XAI Interfaces. The goal of *WP4* is to develop methods and interfaces to let sustainability domain experts steer the behaviour of machine learning models within their application domains. This requires that the models' behaviour is first explained and made transparent to the domain experts. As a second goal, the ecological impact of those explanation mechanisms will be communicated to the practitioner to facilitate more sustainable model training and curation.

Key HCI activities in support of these aims include:

- Design of an explanation-and-control user interface combining local explanations (instance-level) and global explanations (model-level), uncertainty estimates, and contrastive 'why/why-not' comparisons,
- design of user controls to override, constrain, approve, or request more XAI evidence, and to choose explanation granularity [Hartmann et al. 2022],

- Design of adaptive explanation interfaces supporting role- and expertise-dependent explanations (e.g., novice vs. expert, ecological vs. policy lens vs. other user modelling profiles), progressive disclosure (providing further details on-demand), and personalisation based on observed user behaviour and error patterns,
- interaction design for active learning using query strategies presented in human-understandable terms (e.g., 'label these instances reduces uncertainty/bias'),
- Interaction design to capture expert feedback, such as labels, rationales, corrections, or concept constraints, to facilitate interactive model training, and
- Interface design for contested decisions and accountability: provenance of suggestions (cf. data provenance as a data quality signal [Sonntag 2004]), audit trail, and traceable changes across iterations.

Within the doctoral projects outlined above, such activities serve to answer questions such as *How do interactive loops based on supported segmentation affect the accuracy and efficiency of coral morphology segmentation compared to purely manual annotation?* (DP4) or *How are ecology-informed active learning strategies perceived by ecologists acting as labellers? How do they affect training efficiency and expressiveness in labelling?* (DP1).

5.3.4 WP5: Ontology-based (X)AI. The goal of *WP5* is to develop a new large model to capture common data and data types in a sustainability-related application domain (e.g., a new embedding model of PAM data). The embedding space and consequent predictions of this model will directly relate to information and knowledge structures in the application domain, for example, to ontologies of ecological audio data for species identification.

Key HCI activities in support of this goal include:

- Design of concept-level validation and disagreement-resolution workflows,
- Design of interface representations and metaphors for standardised concepts, guidelines, and ontologies in a sustainability-related application domain, for example, in the form of concept browsers, glossaries, or links for PAM concepts in species identification,
- Interaction and interface design for multimodal data inspection, for example, to annotate data or communicate explanations around Life Cycle Assessment (LCA) data, and
- Empirical evaluation with sustainability domain experts (e.g., ecologists) regarding the correctness, actionability, usability, and comprehension of concept-linked explanations and outputs, for example, through semi-structured interviews.

Within the doctoral projects outlined above, such activities serve to answer questions such as *How are the ontology-informed embeddings perceived by ecologists? Are they perceived as more understandable, transparent, and correct?* (DP1) or *To what extent do the newly developed ontology-based embeddings align with the mental models and standardised taxonomies (e.g., ISO standards) used by LCA practitioners?* (DP5).

5.3.5 WP6: Accessible Applications. The goal of *WP6* is to make the technology and methods developed in this project accessible to the wider research community and to the sustainability practitioners. It also includes a second goal to evaluate the trustworthiness of the developed technology with the respective user groups.

Key HCI activities in support of this goal include:

- Design and implementation of a self-sufficient demo application, including onboarding, tutorials, guided workflows, and role-based access with simplified interaction modes for non-programmers and full-access modes for ML experts,
- Cross-platform interaction design (desktop/tablet; optional VR mode if adopted in earlier WPs 3-5),

- Integration of logging and structured user feedback mechanisms to enable continuous improvement, and report export functionality to support interdisciplinary collaboration,
- Documentation for domain experts in a sustainability-related application domain (e.g., ecologists), as well as wider public audiences (e.g., policy makers), and
- Summative evaluation of trustworthiness, including accurate trust calibration and error handling, and longitudinal studies or studies of repeated use on learning effects, adoption, and impact on sustainability practices.

Within the doctoral projects outlined above, such activities serve to answer questions such as *How are the newly developed active learning/embedding strategies most effectively distributed for easy off-the-shelf use by ecologists?* (DP1) or *How does the integration of LLM agents affect the perceived trustworthiness of AI-supported sustainability reports in practice?* (DP5).

6 CONCLUSION

Sustainable zug.KI establishes a research agenda at the intersection of explainable artificial intelligence, interactive machine learning, and sustainability. The project starts from the premise that AI systems should not only be accurate and useful, but also understandable, trustworthy, and justifiable with respect to their ecological, economic, and societal effects. To address this challenge, the project combines methodological work in machine learning, knowledge representation, and intelligent user interfaces with application-driven research in wildlife monitoring, coral morphology analysis, and AI-supported sustainability reporting.

A central contribution of the project is its dual perspective: using XAI to support sustainability-oriented domains and, conversely, making XAI systems themselves more sustainable across their full lifecycle. This includes the analysis of development, deployment, use, and reuse, as well as the creation of robust methods for measuring the impact of AI technologies on resources, workflows, and decision-making. Human-computer interaction plays a key role in this effort by ensuring that domain experts can meaningfully participate in model development, interpretation, and assessment, and by grounding technical advances in real practices and needs.

A particular strength of this new approach lies in its explicit human-centred orientation: domain experts are not treated merely as end users, but as active participants in iterative learning, explanation, validation, and decision-making processes. This emphasis is essential for ensuring that AI systems remain aligned with domain practices, support informed trust, and generate measurable value in real-world sustainability contexts.

The project contributes to a broader vision of AI that is not only technically advanced, but also transparent, participatory, and aligned with sustainability goals. By linking responsible AI development with real-world sustainability challenges, Sustainable zug.KI can help shape a more accountable and resource-aware future for AI research and application.

ACKNOWLEDGMENTS

This research is part of the Computational Sustainability & Technology project area[16], and has been supported by the Lower Saxony Ministry for Science and Culture (MWK), the Endowed Chair of Applied Artificial Intelligence, Oldenburg University, and DFKI.

[16] https://cst.dfki.de/

REFERENCES

Hasan Md Tusfiqur Alam, Devansh Srivastav, Md Abdul Kadir, and Daniel Sonntag. 2025. Towards Interpretable Radiology Report Generation via Concept Bottlenecks Using a Multi-agentic RAG. In *Advances in Information Retrieval*, Claudia Hauff, Craig Macdonald, Dietmar Jannach, Gabriella Kazai, Franco Maria Nardini, Fabio Pinelli, Fabrizio Silvestri, and Nicola Tonellotto (Eds.). Vol. 15574. Springer Nature Switzerland, Cham, 201–209. https://doi.org/10.1007/978-3-031-88714-7_18 Series Title: Lecture Notes in Computer Science.

A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58 (2020), 82–115.

M. Barz, F. Daiber, D. Sonntag, and A. Bulling. 2018. Error-aware gaze-based interfaces for robust mobile gaze interaction. In *ETRA*. 24:1–24:10. Best paper award.

M. Barz, O. Shahzad Bhatti, H. Alam, D. M. H. Nguyen, and D. Sonntag. 2023. Interactive Fixation-to-AOI Mapping for Mobile Eye Tracking Data based on Few-Shot Image Classification. In *IUI Companion*. Honorable mention.

M. Barz, O. Shahzad Bhatti, and D. Sonntag. 2021. Implicit Estimation of Paragraph Relevance From Eye Movements. *Frontiers in Computer Science* 3 (2021), 808507.

M. Barz and D. Sonntag. 2019. Incremental Improvement of a Question Answering System by Re-ranking Answer Candidates Using Machine Learning. In *IWSDS*. 367–379.

M. Barz and D. Sonntag. 2021. Automatic Visual Attention Detection for Mobile Eye Tracking Using Pre-Trained Computer Vision Models and Human Gaze. *Sensors* 21, 12 (2021), 4143.

M. Barz, S. Stauden, and D. Sonntag. 2020. Visual Search Target Inference in Natural Interaction Settings with Machine Learning. In *ETRA*. 1:1–1:8.

Or Biran and Courtenay Cotton. 2017. Explanation and justification in machine learning: A survey. In *IJCAI-17 Workshop on Explainable AI (XAI)*, Vol. 8. 8–13.

R. Biswas, M. Barz, M. Hartmann, and D. Sonntag. 2021. Improving German Image Captions Using Machine Translation and Transfer Learning. In *SLSP*. 3–14.

R. Biswas, M. Barz, and D. Sonntag. 2020. Towards Explanatory Interactive Image Captioning Using Top-Down and Bottom-Up Features, Beam Search and Re-ranking. *Künstliche Intelligenz* 34, 4 (2020), 571–584.

R. Biswas, A. Mogadala, M. Barz, D. Sonntag, and D. Klakow. 2019. Automatic Judgement of Neural Network-Generated Image Captions. In *SLSP*. 261–272.

Sara-Jane Bittner, Robert Andreas Leist, László Kopácsi, Omair Shahzad Bhatti, Abdulrahman Mohamed Selim, Michael Barz, and Daniel Sonntag. 2025. Interactive Multimodal Photobook Co-Creation in Virtual Reality. In *Companion Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25 Companion)*. Association for Computing Machinery, New York, NY, USA, 146–151. https://doi.org/10.1145/3708557.3716355

John Brooke. 1996. SUS: A "quick and dirty" usability scale. In *Usability evaluation in industry*, Bernard Weerdmeester, Patrick W. Jordan, and Ian Lyall McClelland (Eds.). Taylor & Francis, London, 189–194.

Susanne Bødker, Christian Dindler, Ole S. Iversen, and Rachel C. Smith. 2022. *Participatory Design*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-031-02235-7

European Union. 2020. The European Green Deal. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en

European Union. 2021. European Climate Law. https://climate.ec.europa.eu/eu-action/european-climate-law_en

L. Floridi. 2019. Establishing the rules for building trustworthy AI. *Nature Machine Intelligence* 1, 6 (2019), 261–262.

T. Gouvea, H. Kath, I. Troshani, B. Lüers, and D. Sonntag. 2023. Interactive machine learning solutions for acoustic monitoring of animal wildlife in biosphere reserves. In *IJCAI*.

Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, Peter A. Hancock and Najmedin Meshkati (Eds.). Human Mental Workload, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

M. Hartmann, H. Du, N. Feldhus, I. Kruijff-Korbayová, and D. Sonntag. 2022. XAINES: Explaining AI with Narratives. *Künstliche Intelligenz* 36, 3 (2022), 287–296.

Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors* 57, 3 (2015), 407–434.

Kasper Hornbæk, Per Ola Kristensson, and Antti Oulasvirta. 2025. *Introduction to Human-Computer Interaction*. Oxford University Press.

M. A. Kadir, A. M. Selim, M. Barz, and D. Sonntag. 2023. A User Interface for Explaining Machine Learning Model Explanations. In *IUI Companion*. 59–63. Best demo award.

H. Kath, T. S. Gouvêa, and D. Sonntag. 2023. A Human-in-the-Loop Tool for Annotating Passive Acoustic Monitoring Datasets. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*. 7140–7144.

Hannes Kath, Thiago S. Gouvêa, and Daniel Sonntag. 2024a. Active Learning in Multi-label Classification of Bioacoustic Data. In *KI 2024: Advances in Artificial Intelligence*, Andreas Hotho and Sebastian Rudolph (Eds.). Springer Nature Switzerland, Cham, 114–127. https://doi.org/10.1007/978-3-031-70893-0_9

Hannes Kath, Thiago S. Gouvêa, and Daniel Sonntag. 2025. Speeding Up Bioacoustic Data Analysis: Fine-Tuning Deep Models with Active Learning for Efficient Wildlife Detection. In *Proceedings of the 2025 International Conference on Information Technology for Social Good*. ACM, Antwerp Belgium, 200–203. https://doi.org/10.1145/3748699.3749792

Hannes Kath, Thiago S. Gouvêa, and Daniel Sonntag. 2026. Intermediate-Task Transfer Learning for Bioacoustic Data. In *KI 2025: Advances in Artificial Intelligence*, Tanya Braun, Benjamin Paaßen, and Frieder Stolzenburg (Eds.). Vol. 15956. Springer Nature Switzerland, Cham, 47–61. https://doi.org/10.1007/978-3-032-02813-6_4 Series Title: Lecture Notes in Computer Science.

Hannes Kath, Patricia P. Serafini, Ivan B. Campos, Thiago S. Gouvêa, and Daniel Sonntag. 2024b. Demo: Enhancing Wildlife Acoustic Data Annotation Efficiency through Transfer and Active Learning. In *Proceedings of the Thirty-ThirdInternational Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, Jeju, South Korea, 8691–8695. https://doi.org/10.24963/ijcai.2024/1010 0 citations (Semantic Scholar/DOI) [2024-10-06].

Hannes Kath, Patricia P. Serafini, Ivan B. Campos, Thiago S. Gouvêa, and Daniel Sonntag. 2024c. Leveraging transfer learning and active learning for data annotation in passive acoustic monitoring of wildlife. *Ecological Informatics* 82 (Sept. 2024), 102710. https://doi.org/10.1016/j.ecoinf.2024.102710

László Kopácsi, Michael Barz, Omair Shahzad Bhatti, and Daniel Sonntag. 2023. IMETA: An Interactive Mobile Eye Tracking Annotation Method for Semi-automatic Fixation-to-AOI mapping. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 33–36. https://doi.org/10.1145/3581754.3584125

László Kopácsi, Panagiotis Karagiannis, Sotiris Makris, Johan Kildal, Andoni Rivera-Pinto, Judit Ruiz de Munain, Jesús Rosel, Maria Madarieta, Nikolaos Tseregkounis, Konstantina Salagianni, Panagiotis Aivaliotis, Michael Barz, and Daniel Sonntag. 2024. The MASTER XR Platform for Robotics Training in Manufacturing. In *Proceedings of the 30th ACM Symposium on Virtual Reality Software and Technology (VRST '24)*. Association for Computing Machinery, New York, NY, USA, 1–2. https://doi.org/10.1145/3641825.3689514

László Kopácsi, Albert Klimenko, Abdulrahman Mohamed Selim, Michael Barz, and Daniel Sonntag. 2026. Gaze-Based Menu Navigation in Virtual Reality: A Comparative Study of Layouts and Interaction Techniques. In *Human-Computer Interaction – INTERACT 2025*, Carmelo Ardito, Simone Diniz Junqueira Barbosa, Tayana Conte, André Freire, Isabela Gasparini, Philippe Palanque, and Raquel Prates (Eds.). Springer Nature Switzerland, Cham, 520–543. https://doi.org/10.1007/978-3-032-04999-5_31

Hoang H. Le, Duy M. H. Nguyen, Omair Shahzad Bhatti, László Kopácsi, Thinh P. Ngo, Binh T. Nguyen, Michael Barz, and Daniel Sonntag. 2025. I-MPN: inductive message passing network for efficient human-in-the-loop annotation of mobile eye tracking data. *Scientific Reports* 15, 1 (April 2025), 14192. https://doi.org/10.1038/s41598-025-94593-y

Robert Andreas Leist, Hans-Jürgen Profitlich, Tim Hunsicker, and Daniel Sonntag. 2025. Towards Trustable Intelligent Clinical Decision Support Systems: A User Study with Ophthalmologists. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*. ACM, Cagliari Italy, 1470–1484. https://doi.org/10.1145/3708359.3712136

Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. Curran Associates, Inc. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

Raimonds Marcinkevics and Joachim E. Vogt. 2023. Interpretable and explainable machine learning: A methods-centric overview with concrete examples. *WIREs Data Mining and Knowledge Discovery* (2023), e1493. https://doi.org/10.1002/widm.1493

D. M. H. Nguyen, A. O. Ezema, F. Nunnari, and D. Sonntag. 2020. A Visually Explainable Learning System for Skin Lesion Detection Using Multiscale Input with Attention U-Net. *KI* (2020), 313–319.

D. M. H. Nguyen, R. Henschel, B. Rosenhahn, D. Sonntag, and P. Swoboda. 2022a. LMGP: Lifted Multicut Meets Geometry Projections for Multi-Camera Multi-Object Tracking. In *CVPR*. 8856–8865.

D. M. H. Nguyen, H. Nguyen, M. T. N. Truong, T. Cao, B. T. Nguyen, N. Ho, P. Swoboda, S. Albarqouni, P. Xie, and D. Sonntag. 2023. Joint Self-Supervised Image-Volume Representation Learning with Intra-Inter Contrastive Clustering. In *AAAI*.

D. M. H. Nguyen, T. T. Nguyen, H. Vu, Q. Pham, M.-D. Nguyen, B. T. Nguyen, and D. Sonntag. 2022b. TATL: Task agnostic transfer learning for skin attributes detection. *Medical Image Analysis* 78 (2022), 102359.

F. Nunnari, A. O. Ezema, and D. Sonntag. 2020. The Effects of Masking in Melanoma Image Classification with CNNs Towards International Standards for Image Preprocessing. *MobiHealth* (2020), 257–273.

F. Nunnari, A. O. Ezema, and D. Sonntag. 2021a. Crop It, but Not Too Much: The Effects of Masking on the Classification of Melanoma Images. *KI* (2021), 179–193.

F. Nunnari, M. A. Kadir, and D. Sonntag. 2021b. On the Overlap Between Grad-CAM Saliency Maps and Explainable Visual Features in Skin Cancer Images. In *CD-MAKE*. 241–253. Best paper award.

David Patterson, Joseph Gonzalez, Quoc Le, Chen Liang, Lluis-Miquel Munguia, Daniel Rothchild, David So, Maud Texier, and Jeff Dean. 2021. Carbon Emissions and Large Neural Network Training. *arXiv preprint arXiv:2104.10350* (2021). arXiv:2104.10350 [cs.LG] https://arxiv.org/abs/2104.10350

A. Prange and D. Sonntag. 2019. Modeling Cognitive Status through Automatic Scoring of a Digital Version of the Clock Drawing Test. In *ACM UMAP*. 70–77.

A. Prange and D. Sonntag. 2021a. Assessing Cognitive Test Performance Using Automatic Digital Pen Features Analysis. In *UMAP*. 33–43.

A. Prange and D. Sonntag. 2021b. A Demonstrator for Interactive Image Clustering and Fine-Tuning Neural Networks in Virtual Reality. *KI* (2021), 194–203.

A. Prange and D. Sonntag. 2022. Modeling Users' Cognitive Performance Using Digital Pen Features. *Frontiers in Artificial Intelligence* 5 (2022), 787179.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*. 1135–1144.

Samuel R. P.-J. Ross, Darren P. O'Connell, Jessica L. Deichmann, Camille Desjonquères, Amandine Gasc, Jennifer N. Phillips, Sarab S. Sethi, Connor M. Wood, and Zuzana Burivalova. 2023. Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions. *Functional Ecology* 37, 4 (2023), 959–975. https://doi.org/10.1111/1365-2435.14275 45 citations (Semantic Scholar/DOI) [2024-09-15] _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1365-2435.14275.

Cynthia Rudin. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1, 5 (2019), 206–215.

Rida Saghir. 2025. Flexible and Interpretable Soundscape Analysis for Biodiversity Assessment and Ecosystem Health for Domain Experts. In *Companion Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25 Companion)*. Association for Computing Machinery, New York, NY, USA, 218–221. https://doi.org/10.1145/3708557.3716144

Rida Saghir, Ivan B. Campos, Thiago S. Gouvêa, and Daniel Sonntag. 2025. EcoScape Analyzer: A Tool for Performing Soundscape Analysis With Flexible Pipeline for Biodiversity Assessment. In *Companion Proceedings of the 30th International Conference on Intelligent User Interfaces*. ACM, Cagliari Italy, 137–140. https://doi.org/10.1145/3708557.3716359

Daniel Sonntag. 2004. Assessing the quality of natural language text data. Gesellschaft für Informatik e.V., 259–263. https://dl.gi.de/handle/20.500.12116/28866

D. Sonntag and H.-J. Profitlich. 2019. An architecture of open-source tools to combine textual information extraction, faceted search and information visualisation. *Artificial Intelligence in Medicine* 93 (2019), 13–28.

Daniel Sonntag, Colette Weihrauch, Oliver Jacobs, and Daniel Porta. 2010. *THESEUS CTC-WP4 Usability Guidelines for Use Case Applications.* Technical Report. Bundesministerium für Wirtschaft und Technologie. https://www.dfki.de/web/forschung/projekte-publikationen/publikation/4788 Series: DFKI Technical Memos Volume: V 1.5.

Larissa Sayuri Moreira Sugai, Thiago Sanna Freire Silva, José Wagner Ribeiro, and Diego Llusia. 2019. Terrestrial Passive Acoustic Monitoring: Review and Perspectives. *BioScience* 69, 1 (Jan. 2019), 15–25. https://doi.org/10.1093/biosci/biy147 346 citations (Semantic Scholar/DOI) [2024-09-15].

Damian A. Tamburri. 2020. Sustainable MLOps: Trends and Challenges. In *2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*. 17–23. https://doi.org/10.1109/SYNASC51798.2020.00015

I. Troshani, T. S. Gouvêa, and D. Sonntag. 2024a. Leveraging Sound Collections for Animal Species Classification with Weakly Supervised Learning. In *3rd Annual AAAI Workshop on AI to Accelerate Science and Engineering AI2ASE*.

Ilira Troshani, Thiago S. Gouvêa, and Daniel Sonntag. 2024b. Leveraging Weakly Supervised and Multiple Instance Learning for Multi-label Classification of Passive Acoustic Monitoring Data. In *KI 2024: Advances in Artificial Intelligence*, Andreas Hotho and Sebastian Rudolph (Eds.). Springer Nature Switzerland, Cham, 260–272. https://doi.org/10.1007/978-3-031-70893-0_19

United Nations. 2015. Transforming our world: the 2030 Agenda for Sustainable Development. https://sdgs.un.org/goals

J. M. Wing. 2021. Trustworthy AI. *Commun. ACM* 64, 10 (2021), 64–71.

Yunfeng Zhang, Q. Vera Liao, and Rachel K. E. Bellamy. 2020. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 295–305. https://doi.org/10.1145/3351095.3372852