IRCAI

Scientific excellence, global public dialogue and technology for sustainability, inclusion and equality.

Responsible AIoT and TinyML

School on Applied AI for Sustainable Development | (smr 4210)

Trieste 24.3.2025

unesco
Centre
Under the auspices
of UNESCO

ICTP
The Abdus Salam
International Centre
for Theoretical Physics

IRC AI
International Research Centre
of Artificial Inteligence
under the auspices of UNESCO

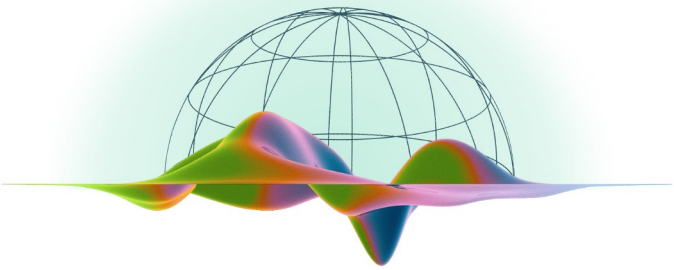
RAIDO

Co-funded by
the European Union

Recommendation on the Ethics of AI

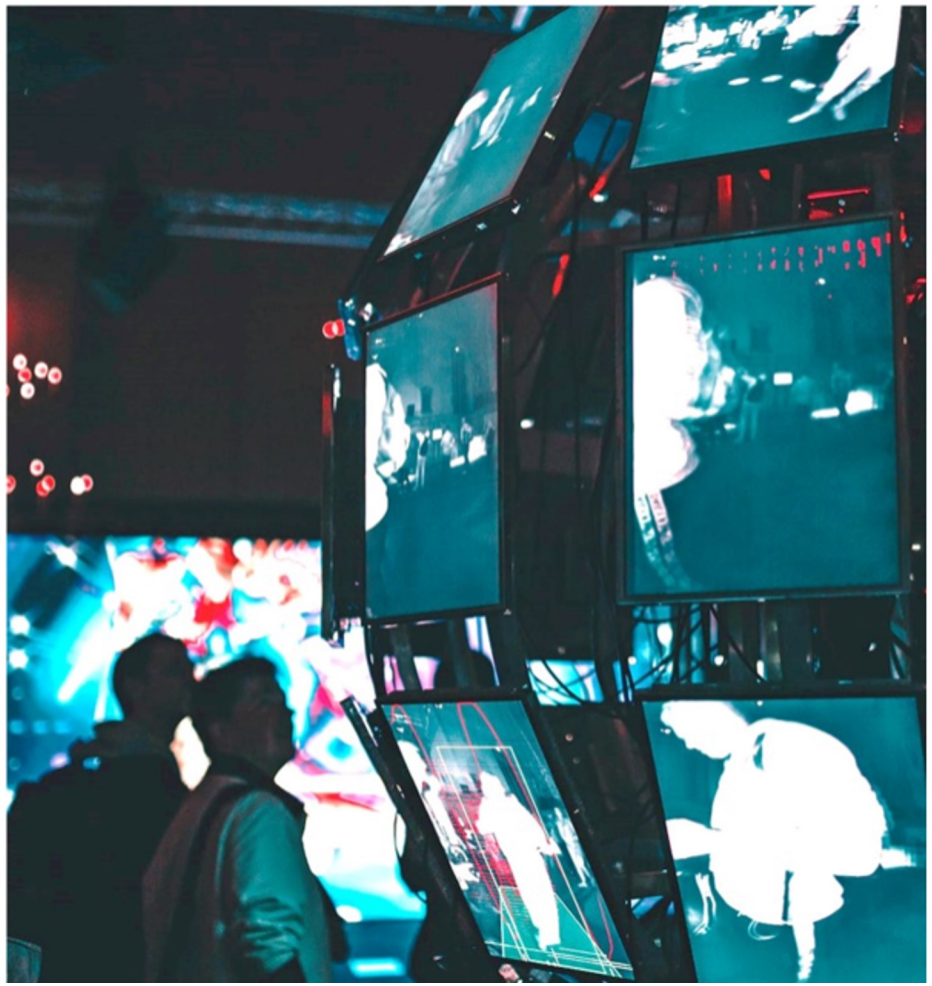
- After extensive consultations, recommendation adopted at the UNESCO assembly in November 2021
- Human centered: how can individual humans be protected from misuse and enriched by the use of AI
- Closely aligned with European humanistic values and with EU initiatives as well as Council of Europe

Global Forum on the Ethics of AI (GFEAI)



unesco Centre Under the auspices of UNESCO

IRC AI International Research Centre of Artificial Intelligence under the auspices of UNESCO



Potential for new and old threats being enhanced

- Citizen control as envisaged in Brave New World
- Automatic processing of decisions that ignore marginalised groups
- Automatic weapons that can be used to disable legitimate protest or wage destructive wars

- EU AI Act is first attempt, but risk based so that many systems not covered
- One advantage of the UNESCO recommendation is that it is vague: forcing designers to think through unforeseen consequences:
  - assessing when human rights have been compromised, system acted unfairly, introduced bias, etc.
  - Any formal definition will inevitably be limited and motivate circumvention

EU AI ACT

unesco
Centre
Under the auspices
of UNESCO

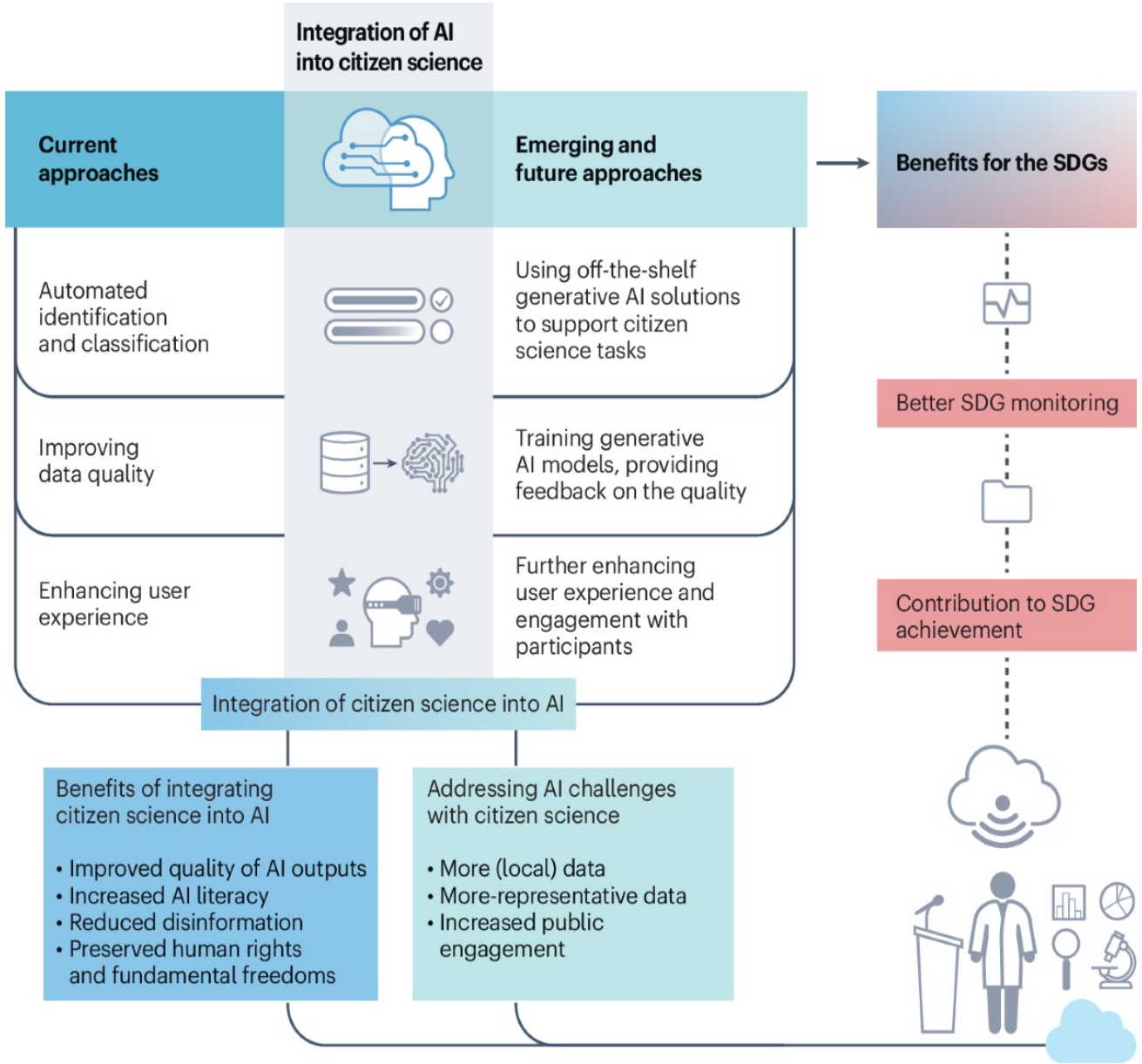
IRC AI
International Research Centre
on Artificial Intelligence
under the auspices of UNESCO

ELIAS

unesco
Centre
Under the auspices
of UNESCO

IRC AI
International Research Centre
of Artificial Intelligence
under the auspices of UNESCO

https://unesdoc.unesco.org/ark:/48223/pf0000385082



Leveraging the collaborative power of AI and citizen science for sustainable development, Nature Sustainability 2024

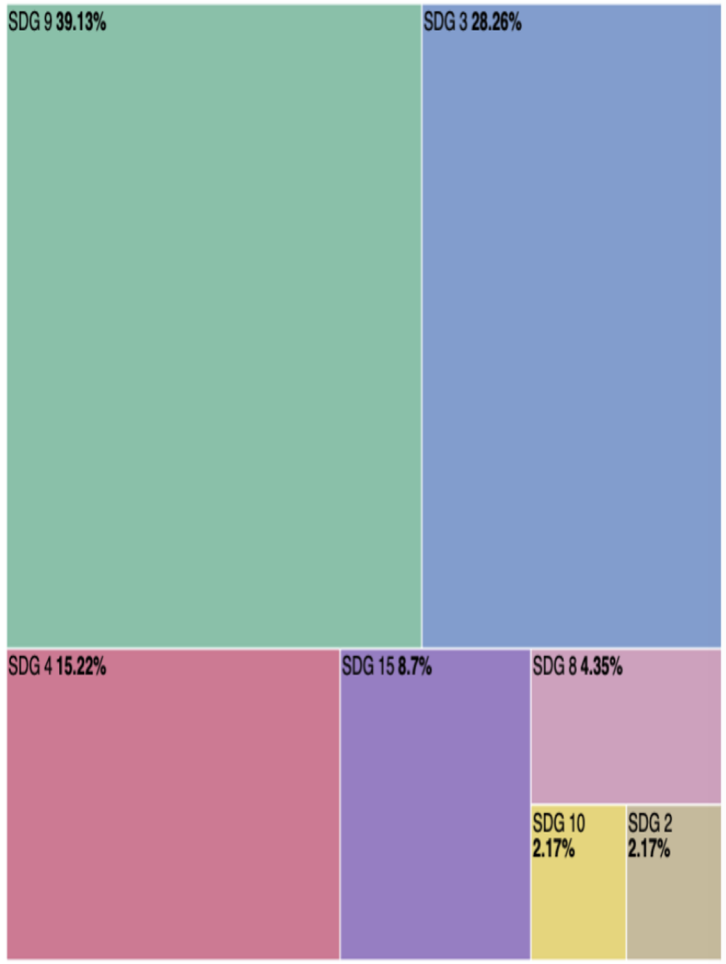
unesco Centre Under the auspices of UNESCO

IRC AI International Research Centre of Artificial Intelligence under the auspices of UNESCO

One way to try to offset the potential misuse is to regulate greater disclosure

- AI should be educational, broadening our understanding of situations and content
- This could be a requirement that enables users to make up their own minds
- Many of the manipulations rely on suggesting that full disclosure is not available

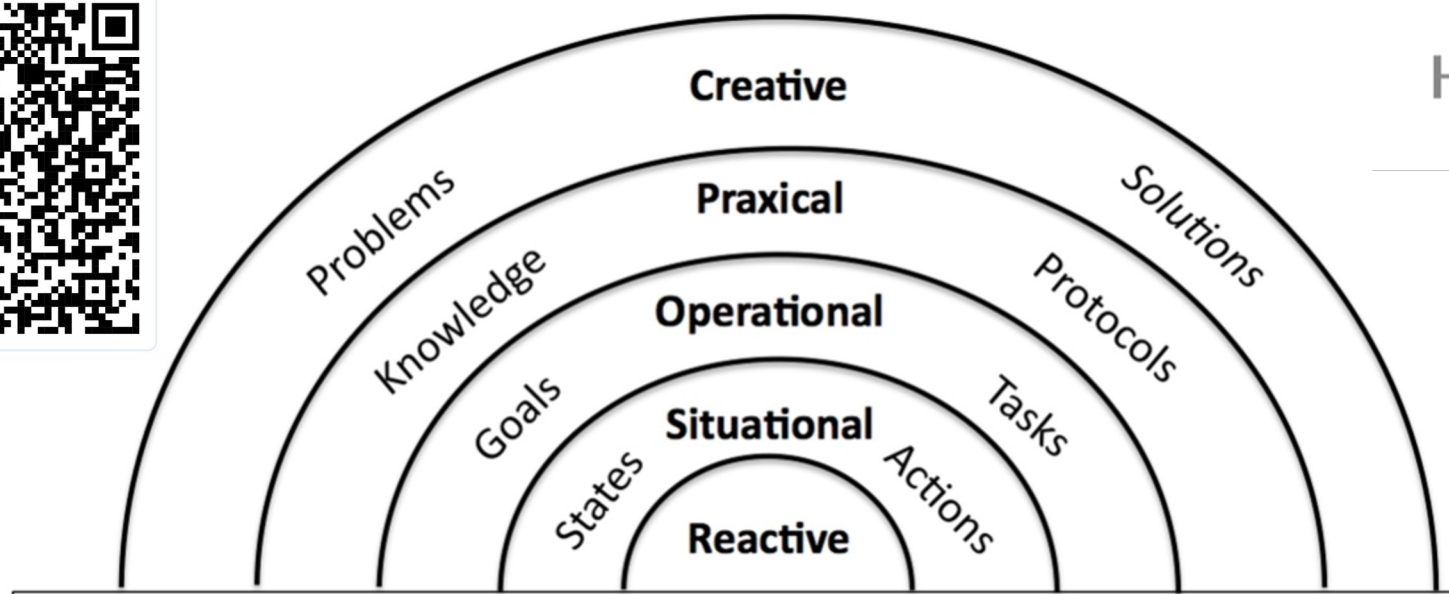
Responsible AI in TinyML and STEAM Education



OPERATIONALIZED ETHICAL AND FUNCTIONAL DIMENSIONS WITH PRIORITY AREAS FOR TINYML AND STEAM EDUCATION

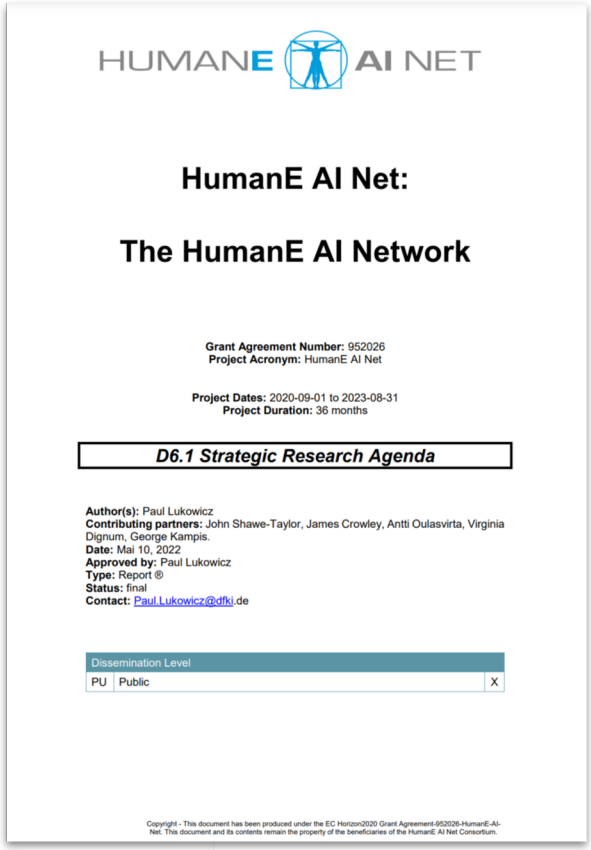
Dimension	Description	EdgeAI/TinyML	STEAM Education
Transparency	Clearly communicate AI operations, data handling, and inference logic to participants to build trust and accountability in both community science and educational contexts.	High – enables interpretability of local models and data flows.	High – critical for teaching explainability and responsible AI use.
Inclusivity	Ensure equitable access to AI tools, datasets, and learning resources, addressing disparities in infrastructure, gender representation, and cultural participation.	Medium – supports broader participation in device deployment.	High – foundational for equitable AI literacy and engagement.
Fairness	Identify and mitigate algorithmic bias across data collection, labeling, and model evaluation processes, especially in community or low-resource settings.	High – essential for localized and context-aware model training.	High – core concept for teaching ethical reasoning in AI and data science.
Environmental Responsibility	Assess and reduce lifecycle impacts of AI hardware, including sourcing, energy consumption, and end-of-life management.	High – TinyML's edge devices directly affect ecological sustainability.	Medium – integrated into STEAM through sustainability modules and project design.
Cultural Diversity	Incorporate local knowledge, indigenous epistemologies, and linguistic diversity into AI datasets, interpretations, and learning activities.	Medium – supports culturally relevant sensor deployment and data interpretation.	High – encourages interdisciplinary, culturally responsive pedagogy.
Operational Assistance	TinyML devices perform localized inference for environmental monitoring, health tracking, or IoT sensing, reducing reliance on cloud infrastructure.	High – primary operational focus of edge ML systems.	Medium – serves as an applied learning example in engineering and computing courses.
Decision Support	Generative AI systems synthesize data insights and summaries to assist collective decision-making and policy formation in citizen science projects.	Medium – complements TinyML data with higher-level synthesis.	High – enhances inquiry-based learning, problem-solving, and critical reflection.
Interpretive / Narrative Collaboration	AI systems co-generate visualizations, reports, and educational materials with human collaborators, promoting reflective learning and knowledge co-creation.	Medium – relevant in projects using embedded visualization or reporting tools.	High – central to creative and narrative integration in STEAM curricula.

- Human-centric AI is at the centre of the European vision of a positive role for AI
- AI that empowers humans to be more effective, more creative, more understanding
- Human-centric AI (HCAI) has been the focus of the Humane AI Network of Excellence
- How does its research agenda envisage HCAI?



Humane AI Research agenda highlights the ingredients of Collaborative Intelligent Systems:

- Need to find ‘common ground’ across a range of levels in order to enable effective cooperation/communication
- Levels identified roughly correspond to different styles of collaboration with collaborative systems potentially involving more than one level



HUMANE AI NET

HumanE AI Net:

The HumanE AI Network

Grant Agreement Number: 952026
Project Acronym: HumanE AI Net

Project Dates: 2020-09-01 to 2023-08-31
Project Duration: 36 months

D6.1 Strategic Research Agenda

Author(s): Paul Lukowicz
Contributing partners: John Shawe-Taylor, James Crowley, Antti Oulasvirta, Virginia Dignum, George Kampis.
Date: Mai 10, 2022
Approved by: Paul Lukowicz
Type: Report ®
Status: final
Contact: Paul.Lukowicz@dfki.de

Dissemination Level		
PU	Public	X

unesco Centre Under the auspices of UNESCO

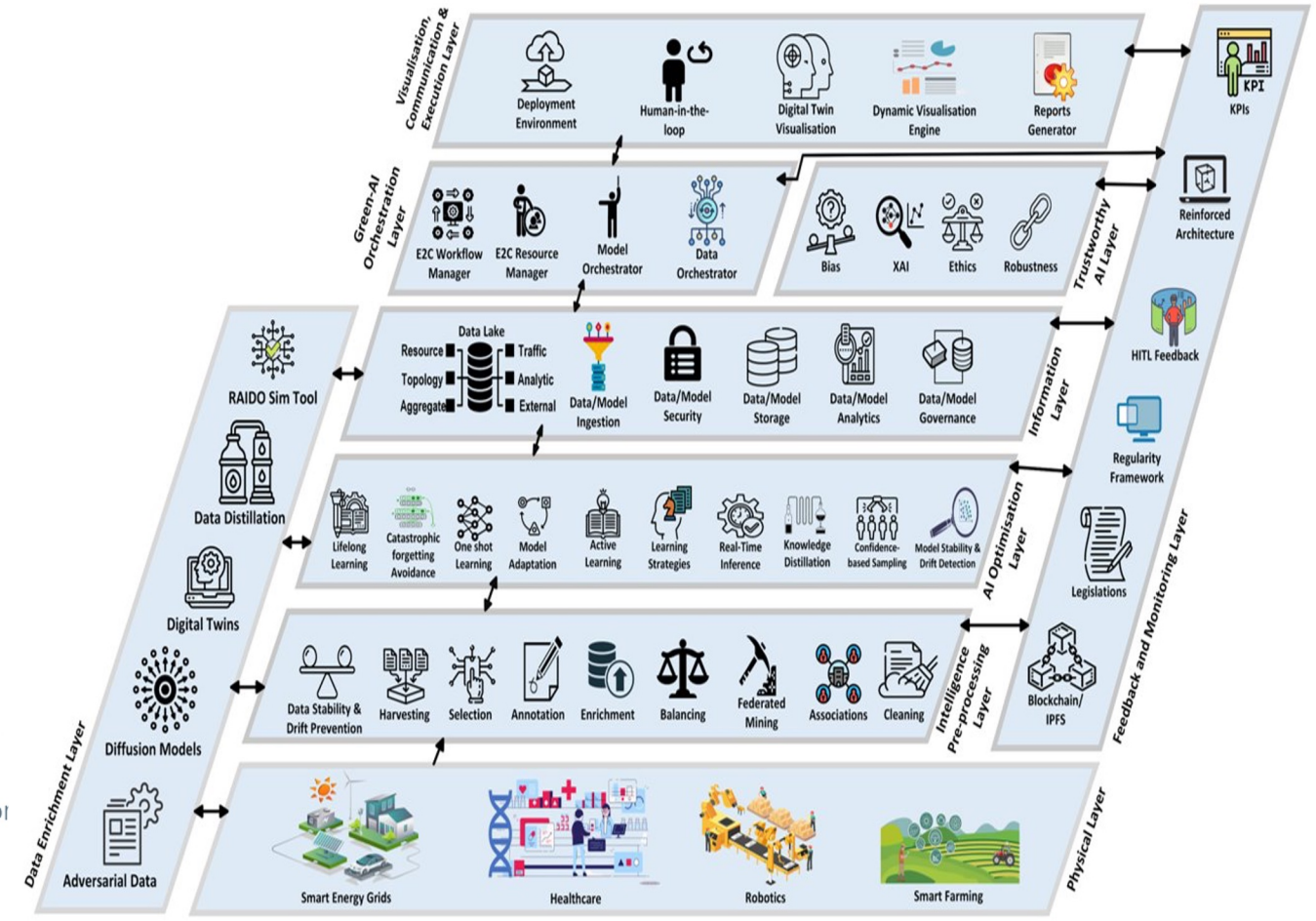
IRC AI International Research Centre of Artificial Intelligence under the auspices of UNESCO

RAIDO

raido-project.eu

Reliable AI and Data Optimization

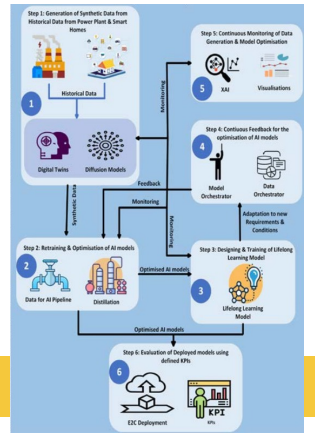
An integrated platform for Green and Energy-efficient data and model related operatio

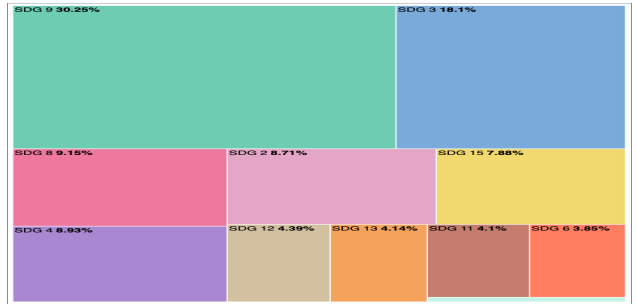


Green AI Orchestration Bridging Trustworthy AI and Edge AI through tinyML for Frugal Intelligence

Joao Pita Costa[a,*], Marco Zennaro[b], Ioana Ntinou[c,1], John Shawe-Taylor[a]

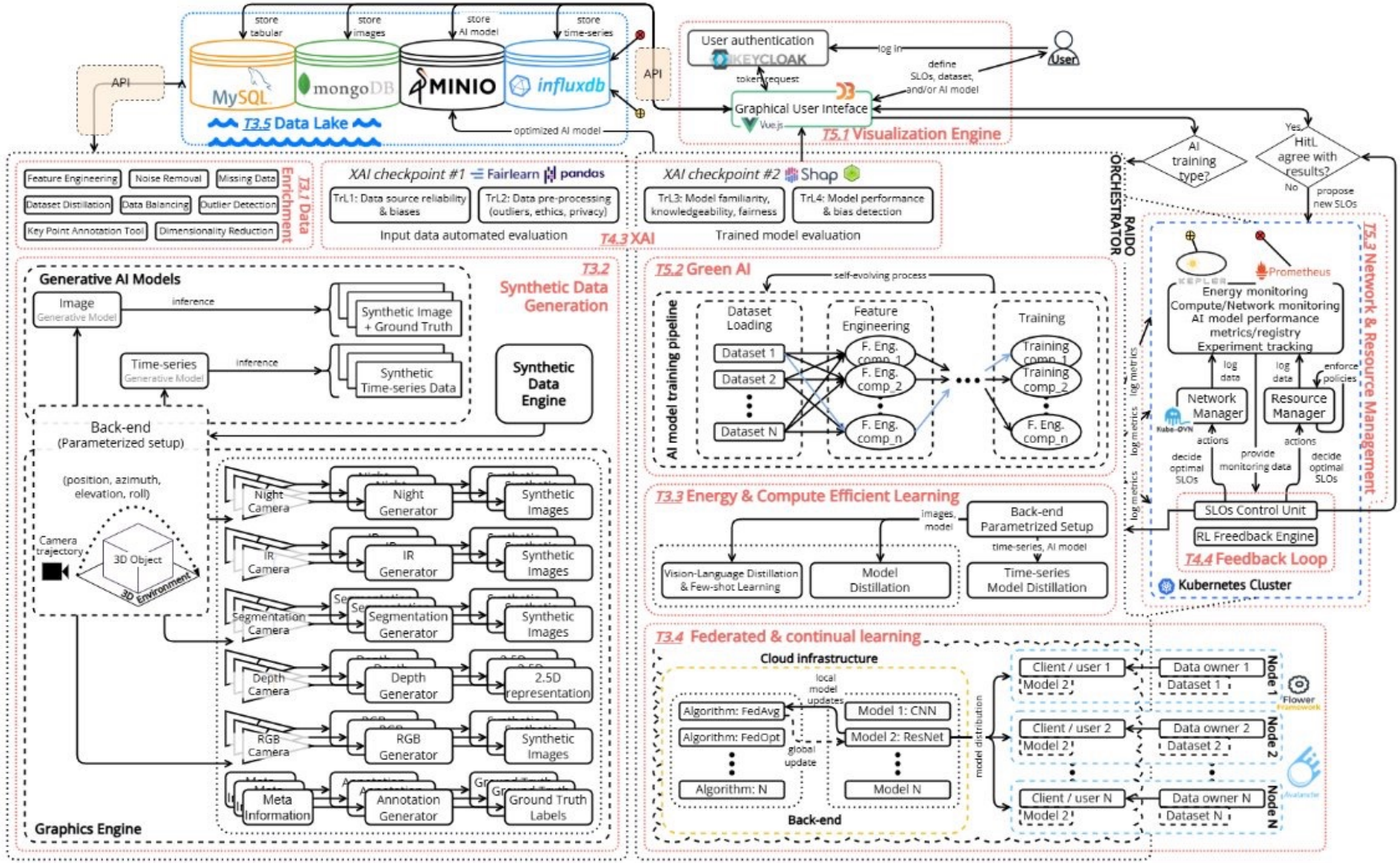
[a] *International Research Centre on AI under the auspices of UNESCO, Jamova cesta 39, Ljubljana, 1000, Slovenia*
[b] *ICTP, Str. Costiera 11, Trieste, 34151, taly*
[c] *Queen Mary University of London, Mile End Road, London, E1 4NS, UK*



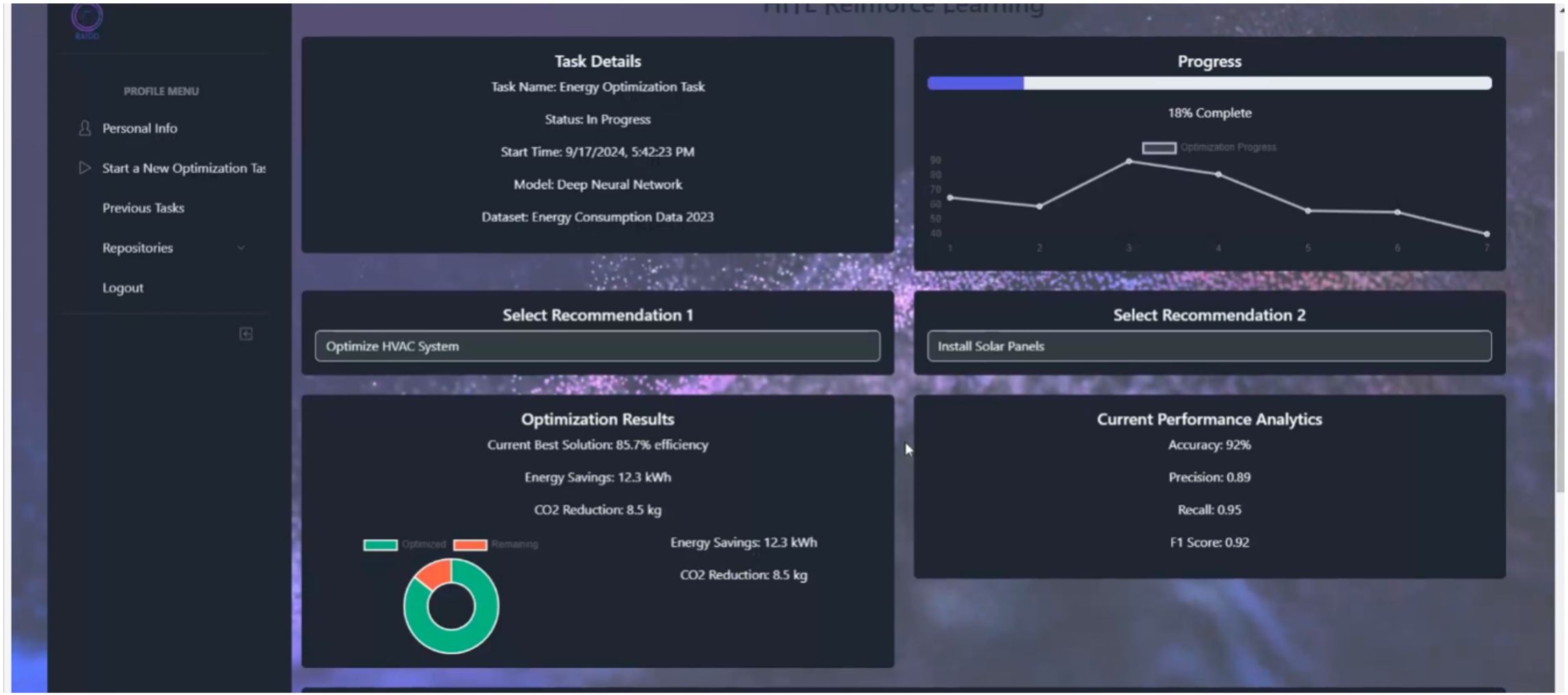


unesco Centre Under the auspices of UNESCO

IRC AI International Research Centre of Artificial Intelligence under the auspices of UNESCO


store tabular
store images
store AI model
store time-series
API
MySQL
mongoDB
MINIO
influxdb
T3.5 Data Lake
optimized AI model
User authentication
KEYCLOAK
log in
User
token request
define SLOs, dataset, and/or AI model
Graphical User Inteface
Vue.js
T5.1 Visualization Engine
AI training type?
Yes, HitL agree with results?
No
propose new SLOs
RAIDO ORCHESTRATOR
Feature Engineering
Noise Removal
Missing Data
Dataset Distillation
Data Balancing
Outlier Detection
Key Point Annotation Tool
Dimensionality Reduction
T3.1 Data Enrichment
XAI checkpoint #1 Fairlearn pandas
TrL1: Data source reliability & biases
TrL2: Data pre-processing (outliers, ethics, privacy)
Input data automated evaluation
XAI checkpoint #2 Shap
TrL3: Model familiarity, knowledgeability, fairness
TrL4: Model performance & bias detection
Trained model evaluation
T4.3 XAI
Generative AI Models
Image Generative Model
inference
Synthetic Image + Ground Truth
Time-series Generative Model
inference
Synthetic Time-series Data
T3.2 Synthetic Data Generation
Synthetic Data Engine
Back-end (Parameterized setup)
(position, azimuth, elevation, roll)
Camera trajectory
3D Object
3D Environment
Night Camera
Night Generator
Synthetic Images
IR Camera
IR Generator
Synthetic Images
Segmentation Camera
Segmentation Generator
Synthetic Images
Depth Camera
Depth Generator
2.5D representation
RGB Camera
RGB Generator
Synthetic Images
Meta Information
Annotation Generator
Ground Truth Labels
Graphics Engine
T5.2 Green AI
self-evolving process
AI model training pipeline
Dataset Loading
Dataset 1
Dataset 2
Dataset N
Feature Engineering
F. Eng. comp_1
F. Eng. comp_2
F. Eng. comp_n
Training
Training comp_1
Training comp_2
F. Eng. comp_n
T3.3 Energy & Compute Efficient Learning
images, model
Back-end Parametrized Setup
time-series, AI model
Vision-Language Distillation & Few-shot Learning
Model Distillation
Time-series Model Distillation
T3.4 Federated & continual learning
Cloud infrastructure
local model updates
Algorithm: FedAvg
Algorithm: FedOpt
Algorithm: N
global update
Model 1: CNN
Model 2: ResNet
Model N
Back-end
model distribution
Client / user 1
Model 2
Data owner 1
Dataset 1
Node 1
Client / user 2
Model 2
Data owner 2
Dataset 2
Node 2
Client / user N
Model 2
Data owner N
Dataset N
Node N
Flower framework
Avalanche
KEPLER
Prometheus
Energy monitoring
Compute/Network monitoring
AI model performance metrics/registry
Experiment tracking
log data
log data
enforce policies
Network Manager
Kube-OVN
Resource Manager
actions
actions
decide optimal SLOs
provide monitoring data
decide optimal SLOs
log metrics
log metrics
log metrics
SLOs Control Unit
RL Freedback Engine
T4.4 Feedback Loop
Kubernetes Cluster
T5.3 Network & Resource Management

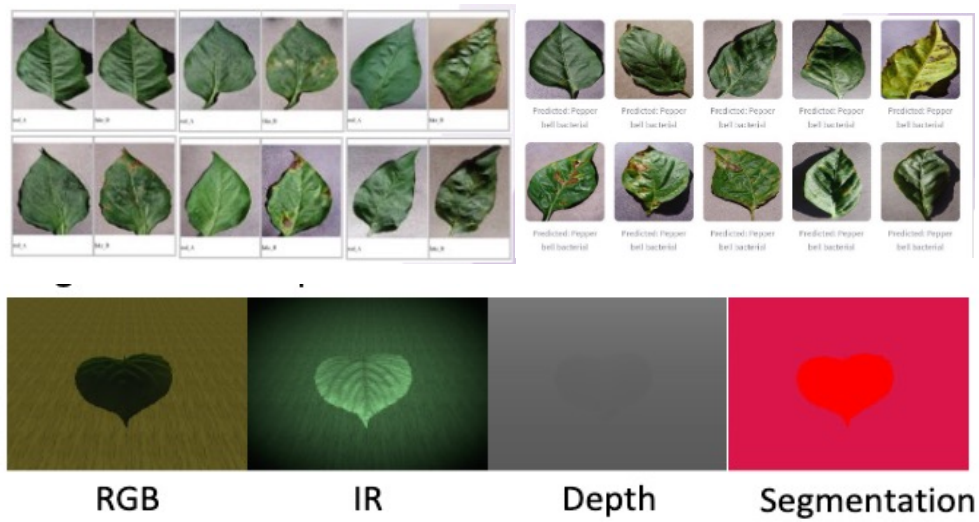
RAIDO AI Optimisation Platform



RAIDO Data Enrichment and Distillation Software (DEDS)

Powering Ethical AI with Smarter Data, Lower Costs, and Seamless Compliance

DEDS helps organisations that develop and deploy AI solutions that struggle with unpredictable costs, technical complexity and ethical compliance by employing **optimised data flows** and providing **quality synthetic data**, while enhancing competitive positioning, enabling continuous compliance and automate workflows with built-in ethical AI capabilities.



Components:

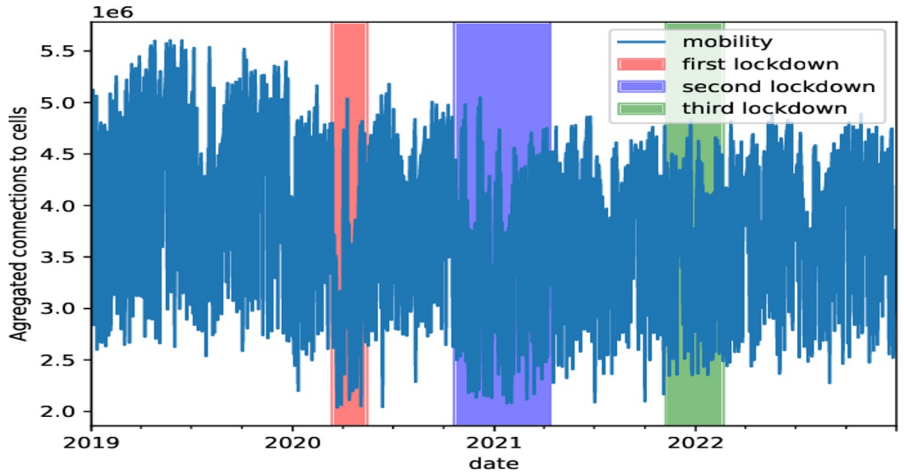
- Data Enrichment (MINDS)
- DataSim (KU)
- Data Lakes (UBI)

RAIDO Toolbox for Optimized and Energy Efficient Training of DL Models

QUEEN MARY UNIVERSITY OF LONDON

Making AI training sustainable through model distillation and efficient data handling

The toolbox for optimized and energy efficient training of DL models helps AI researchers (Post Docs in AI institutions) who want to train an AI model by **avoiding using a high number of GPUs** and enable availability of models, technologies for distillation and data loading.



Components:

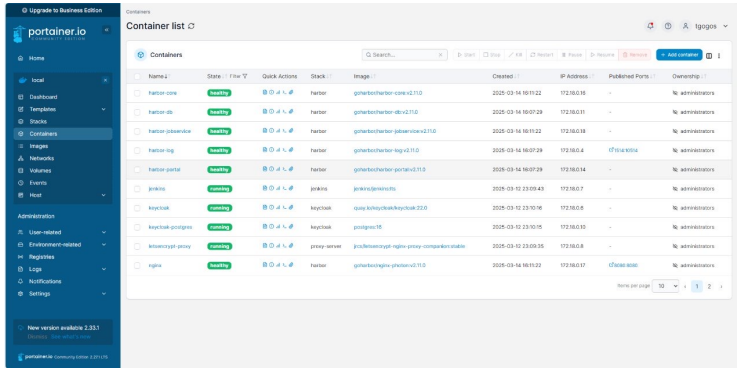
- Distill-ai
- Lifelong Learning in Federated Architectures

RAIDO Transparent and verifiable NFTs for AI model cards using Blockchain Smart-Contracts

SIDROCO

Ensuring AI model accountability with verifiable and tamper-proof NFTs

Transparent and verifiable NFTs for AI model cards using Blockchain Smart-Contracts help organisations and consortia involved in responsible AI development, auditing, and compliance that struggle with the complexity and costs associated to tracking ethics and AI governance compliance to track and verify an AI model's lifecycle using NFTs to **mitigate unauthorised model alterations** and **simplify audit processes**, while achieving a **high level of trust and transparency** ensuring compliance with current and future regulation with cryptographic proof guaranteeing treaceability and accountability of AI model governance.



Component:

- RAIDO Blockchain

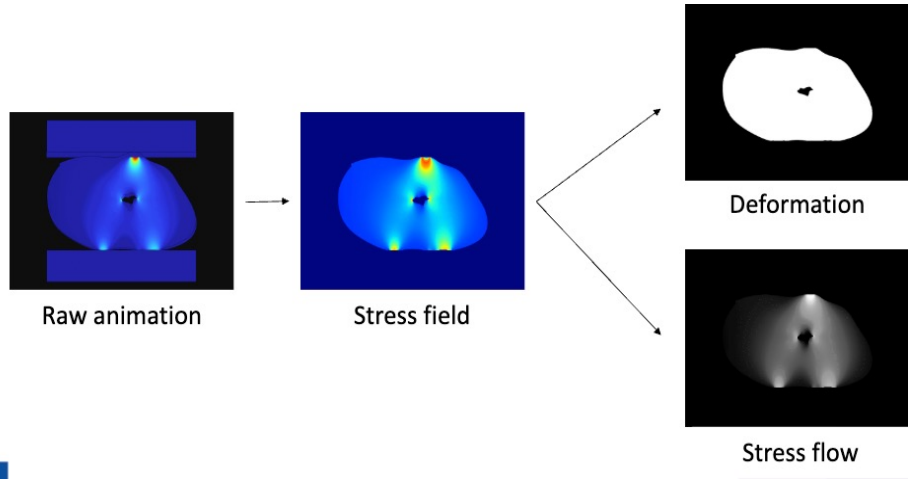
unesco Centre Under the auspices of UNESCO

IRC AI International Research Centre of Artificial Intelligence under the auspices of UNESCO

RAIDO Human-in-the-loop interface of explainable AI optimization platforms

Versatile, explainable AI to optimise farm operations with confidence

The human-in-the-loop interface of explainable AI optimisation platforms helps companies working in farm management which want to customise business AI models (in terms of energy, efficiency, accuracy and speed) with given customer priorities by **reducing lack of model versatility** across use cases and **enabling quicker and better-informed decisions** about all aspects of farm management.



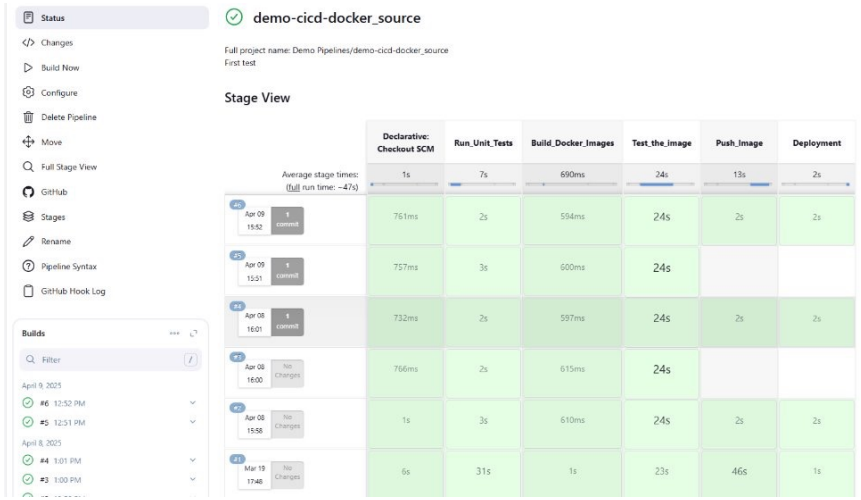
Components:

- XAI engine & framework
- Multi-objective Reinforced Benchmarking
- Visualisation Tool

RAIDO RAIDO Green-AI Orchestrator

Seamless orchestration of legacy and modern AI tools for greener workflows

RAIDO Green-AI Orchestrator helps AI developers who want to monitor hardware/software resources by reducing difficulty integrating new AI tools with legacy systems or dashboards and enabling **tangible energy and cost savings** from efficient AI workflows



Components:

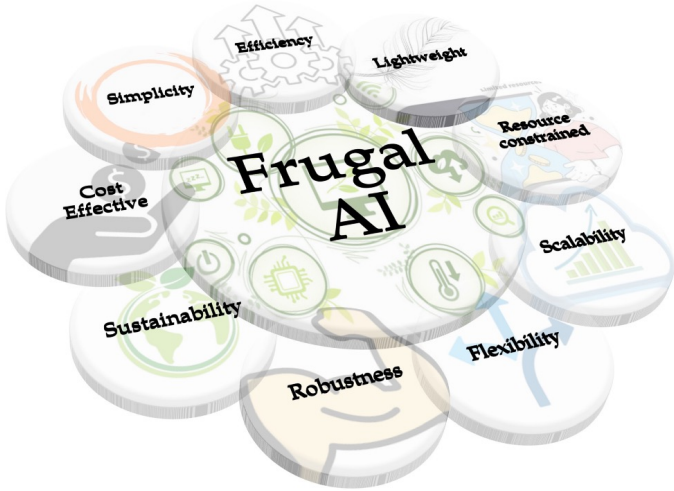
- Data Lakes
- E2C continuum resource allocation and network management
- Self-evolving Green-AI and Data Orchestration

RAIDO & Frugal Edge AI

Frugal Edge AI intersects with several contemporary research trajectories, notably Edge Intelligence, Green AI, and Sustainable Computing.

- **Edge Intelligence** has emerged as a response to the latency, privacy, and bandwidth limitations of cloud-based AI systems, focusing on distributing computation closer to data sources
- **Green AI** advocates for efficiency and transparency in resource use throughout the AI lifecycle,
- **Sustainable Computing** emphasizes energy-aware system design

➢ Frugal Edge AI synthesizes these streams by prioritizing algorithmic compactness and energy efficiency through techniques such as model pruning, quantization, and on-device learning, as exemplified in the field of TinyML. Unlike conventional edge AI, which often assumes moderate hardware capacity and network reliability, Frugal Edge AI explicitly targets low-cost, low-power contexts—those typical of rural, remote, or developing regions. Its ambition is not only technical optimization but also social and environmental equity, aligning with broader goals of responsible AI and the Sustainable Development Goals (SDGs).




Ingesting heterogeneous data to provide a range of perspectives on the impact of AI in the progress of SDGs
TinyML for Sustainable Development
What is TinyML?


TinyML for Sustainable Development


tinyML applications
Medical Devices
Satellites
Robotics
Smart Factory
Smart Logistics
Smart Cities
IoT sensors
Agriculture
Smart Farming


TinyML for Sustainable Development
What is TinyML?

Frugal Edge AI in India

Elephant-Train Collision Prevention: Researchers in India developed a smart acoustic and thermal sensor system using TinyML.

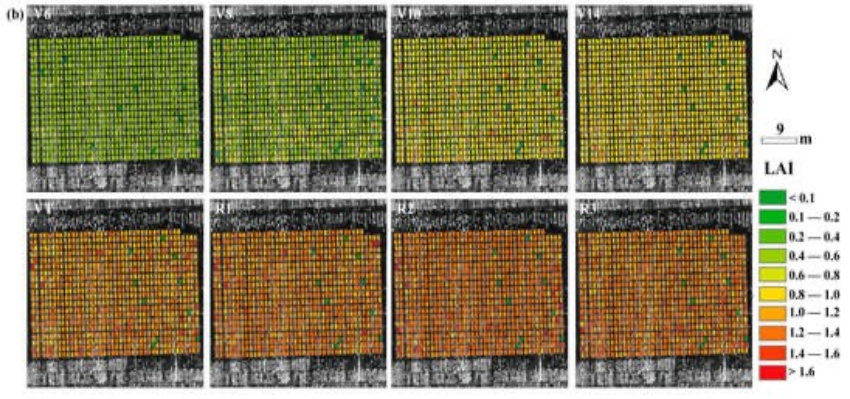
- This is a safety-critical system: incorrect predictions directly affect both wildlife and human lives

Cashew Tree Monitoring: At the Vellore Institute of Technology (VIT), researchers use TinyML-equipped drones to detect fungal diseases in cashew trees with up to 99% accuracy, significantly reducing pesticide use.

- Precision agriculture efficiency vs. risks of inequity, over-automation, and ecological mismanagement

Maize Leaf Analysis: Recent 2025 research utilized a hybrid TinyML platform to identify five types of maize leaf diseases (e.g., Blight, Common Rust) with over 94% accuracy, specifically optimized for low-cost microcontrollers.

- Scaling low-cost AI for food security vs. risks of misclassification, inequity, and over-reliance in vulnerable agricultural systems



Frugal Edge AI in Malawi

Early Disease Detection: Using TinyML sensors to identify plant diseases in the field before they spread, directly supporting the livelihoods of the country's 3.1 million smallholder farmers.

- Scaling AI for inclusive agricultural resilience vs. the risk of amplifying inequality, dependency, and systemic vulnerability

Smart Irrigation: Optimizing water usage through real-time soil and weather data processed on-device.

- Efficiency-driven resource optimization vs. long-term ecological balance and equitable access to water

Environmental & Health Monitoring: Training was provided on using microcontrollers for wildlife conservation and healthcare monitoring.

- Democratizing AI tools through TinyML vs. the risks of misuse, weak oversight, and sensitive data exposure

Ethical Dimensions of Generative and Edge AI for Participatory Citizen Science and STEAM Education, integrating Human-Centered Frameworks

Joao Pita Costa*
0000-0001-5745-1302
IRCAI
Ljubljana, Slovenia
joao.pitacosta@ijs.si

Marco Zennaro
0000-0002-0578-0830
ICTP
Ljubljana, Slovenia
mzennaro@ictp.it

John Shawe-Taylor
0000-0002-2030-0073
IRCAI
Ljubljana, Slovenia
john.shawe-taylor@ijs.si

***Abstract*—The adoption of Artificial Intelligence (AI) technologies at the edge and in participatory research settings is rapidly accelerating. Today Tiny Machine Learning (TinyML) allows ML and even Large Language Model (LLM) inference on low-power microcontrollers, enabling local environmental monitoring, health tracking, and citizen-led research, empowering communities worldwide to leverage the "edge of Agentic AI" through advanced AI/ML approaches in addressing local problems that they know best with locally sourced data combined with open data. Although the ambitious advantage of decentralizing the compute power to run AI/ML, other concerns come along including data bias, trustworthiness of the algorithms as well as the ethics and explainability of the AI used. This paper critically investigates the ethical dimensions of integrating LLM-enabled TinyML into citizen science and education, guided by the UNESCO Recommendations on the Ethics of AI, complemented by the UNESCO Guidance on Generative AI in Education and Research. These can help us understand how citizen-led AI initiatives leveraging TinyML/LLMs can be ethically designed, governed, and implemented to foster inclusivity and human rights**

I. Introduction

The intersection of citizen science, open education, and TinyML presents an evolving domain for ethical inquiry within AI. TinyML, which enables ML on low-power, affordable edge devices, offers promising avenues for participatory data collection in environmental monitoring, public health, and community-driven research. As these tools become increasingly accessible to educators, students, and citizen scientists, they create opportunities for inclusive knowledge production and community empowerment [1]. Yet, this democratization of AI amplifies complex ethical challenges concerning data governance, algorithmic bias, trustworthiness, and equitable access to emerging technologies—especially within low-resource and culturally diverse contexts where disparities in AI literacy and infrastructure risk amplifying existing inequities.

Concurrently, Generative AI (GenAI) offers the ability

Green AI Orchestration Bridging Trustworthy AI and Edge AI through tinyML for Frugal Intelligence

Joao Pita Costa[a,*], Marco Zennaro[b], Ioana Ntinou[c,1], John Shawe-Taylor[a]

[a] *International Research Centre on AI under the auspices of UNESCO, Jamova cesta 39, Ljubljana, 1000, Slovenia*
[b] *ICTP, Str. Costiera 11, Trieste, 34151, taly*
[c] *Queen Mary University of London, Mile End Road, London, E1 4NS, UK*

Abstract

The rapid proliferation of edge AI, particularly within the tinyML ecosystem, demands innovative approaches to balance computational efficiency, energy consumption, and model performance. This paper explores the relevance of Green AI orchestration—exemplified by initiatives such as the RAIDO project—to the tinyML community, where resource constraints and sustainability are paramount. By integrating energy-efficient hardware, optimized algorithms, and adaptive orchestration frameworks, green AI paradigms allied to TinyML offer a pathway to deploy trustworthy, low-footprint AI models on resource-limited devices. We examine how these principles align with the ethos of frugal AI, which prioritizes minimalism, accessibility, and environmental responsibility. The paper sets the stage for a new era of Frugal Edge AI, where green TinyML orchestration not only reduces operational costs and carbon footprints but also democratizes AI deployment in edge environments. Our discussion underscores the synergy between green, frugal and edge AI, advocating for a holistic approach to AI development that harmonizes technological advancement with environmental stewardship.

Article

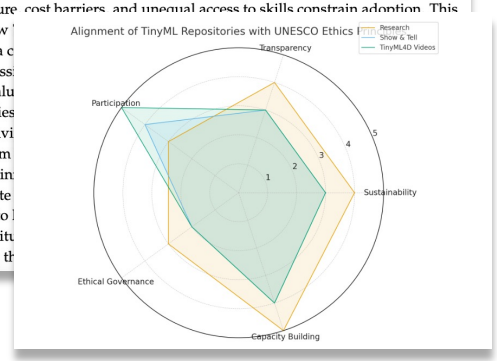
Empowering Local Frugal Edge AI Innovation based on Participatory Citizen Science in Developing Countries

Joao Pita Costa [1,*] 0000-0001-5745-1302, Thomas Basikolo [2] 0000-0002-0970-9863, Marco Zennaro [2] 0000-0002-0578-0830 and John Shawe-Taylor [1] 0000-0002-2030-0073

1 International Research Centre on AI under the auspices of UNESCO; joao.pitacosta@ircai.org
2 International Telecommunication Union; thomas.basikolo@itu.int
3 Abdus Salam International Centre for Theoretical Physics; mzennaro@ictp.it
* Correspondence: joao.pitacosta@ircai.org

Abstract

With the 2030 deadline for the United Nations Sustainable Development Goals (SDGs) approaching, there is a growing global urgency to identify innovative, scalable, and inclusive AI-based or AI-enabled solutions capable of accelerating progress across sectors. Yet the benefits of AI remain unevenly distributed, particularly in low-resource settings where limited infrastructure, cost barriers, and unequal access to skills constrain adoption. This paper explores how
paradigm—offers a c
AI, providing access
velopment. We evalu
enables communities
in areas such as envi
early outcomes from
paper conducts an in
approaches cultivate
prototypes suited to
technical, and institu
tives, emphasizing th



https://zenodo.org/records/17347987

Continuous Compliance with Legislations for Privacy, IRP & Data Collection

AI Ethics & Princples by Design Framework

Reinforced Benchmarking & Feedback-based Progress Monitoring

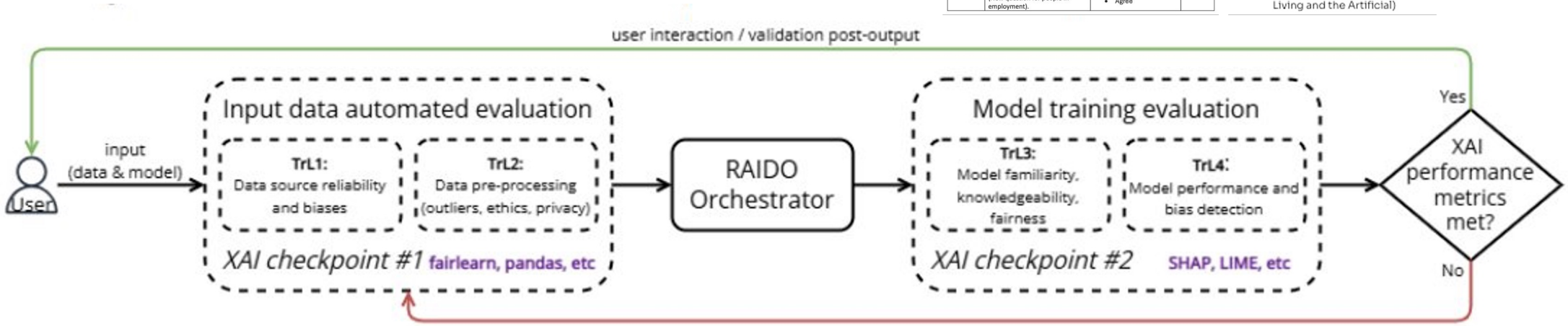
Acceptance of AI use in life and job/industry			
20	I am interested in exploring new technological developments for my field of work/in my daily life.	• Strongly Disagree • Disagree • Neutral/Neither Agree nor Disagree • Agree • Strongly Agree	
21	I could imagine including AI in my current work or in my daily life.	• Strongly Disagree • Disagree • Neutral/Neither Agree nor Disagree • Agree • Strongly Agree	
22	I see the benefit of using AI tools in my current work or in my daily life.	• Strongly Disagree • Disagree • Neutral/Neither Agree nor Disagree • Agree • Strongly Agree	
23	Policymakers strongly support AI.	• Strongly Disagree • Disagree • Neutral/Neither Agree nor Disagree • Agree • Strongly Agree	
24	**(For respondents in employment)** The use of AI in my field aligns with what society considers appropriate for the industry.	• Strongly Disagree • Disagree • Neutral/Neither Agree nor Disagree • Agree • Strongly Agree	
25	**(For respondents in employment)** I am the person who decides on the use of AI at my job. (For people in employment)	• Strongly Disagree • Disagree • Neutral/Neither Agree nor Disagree • Agree • Strongly Agree	
26	**(For respondents in employment)** I feel that my organisation would support me in the adoption of AI. (New question for people in employment).	• Strongly Disagree • Disagree • Neutral/Neither Agree nor Disagree • Agree	

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

vito

VANESSA NUROCK

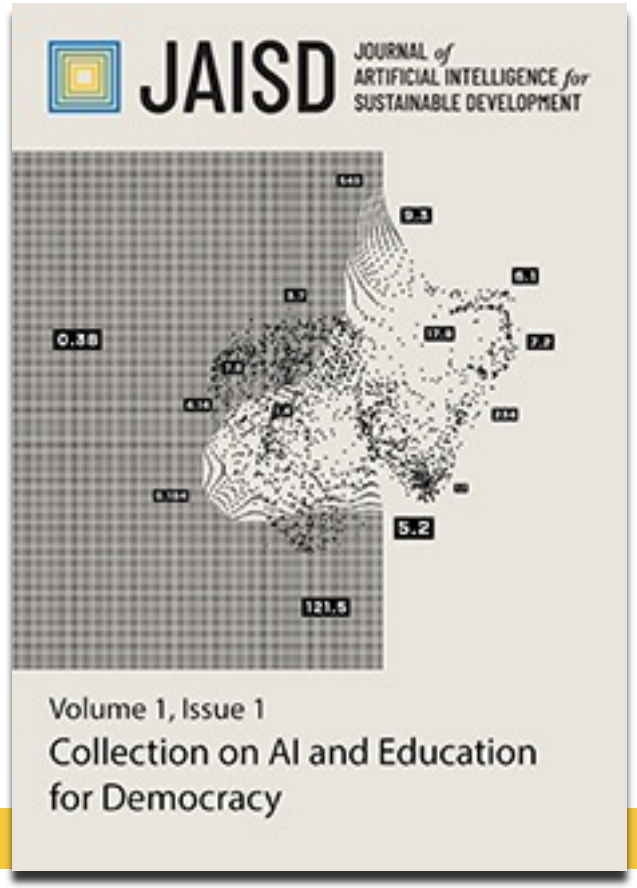
Professor of Philosophy at the Université de Côte d'Azur (France) and UNESCO EVA Chair (Ethique du Vivant et de l'Artificiel /Ethics of the Living and the Artificial)



Framework for AI Explainability, Transparency & Trustworthiness

unesco Centre Under the auspices of UNESCO

IRC AI International Research Centre of Artificial Intelligence under the auspices of UNESCO



Informed Consent: ensure participants know why their data is being collected, how it will be used, and any associated risks.

Privacy: Protect the identities of participants by anonymizing or de-identifying data.

Transparency: clear purpose of the data collection, the methods used, and potential benefits and risks.

Data Sharing: participants should be aware and have given consent.

Vulnerable Populations: extra precautions when collecting data from children, the elderly, and other vulnerable groups.

Potential Harm: Assess and minimize potential risks to participants. This includes emotional distress, financial harm, or other adverse effects.

Purpose Change: seek fresh consent from participants.

Limitation: avoid over-collection.

Accuracy: data is accurate and represents the truth.

Beneficence: data collection should be for good.

Models vs. Data: storing models, forgetting the data

Access: who has access to the data.

Long-term Storage: ensure secure storage methods.

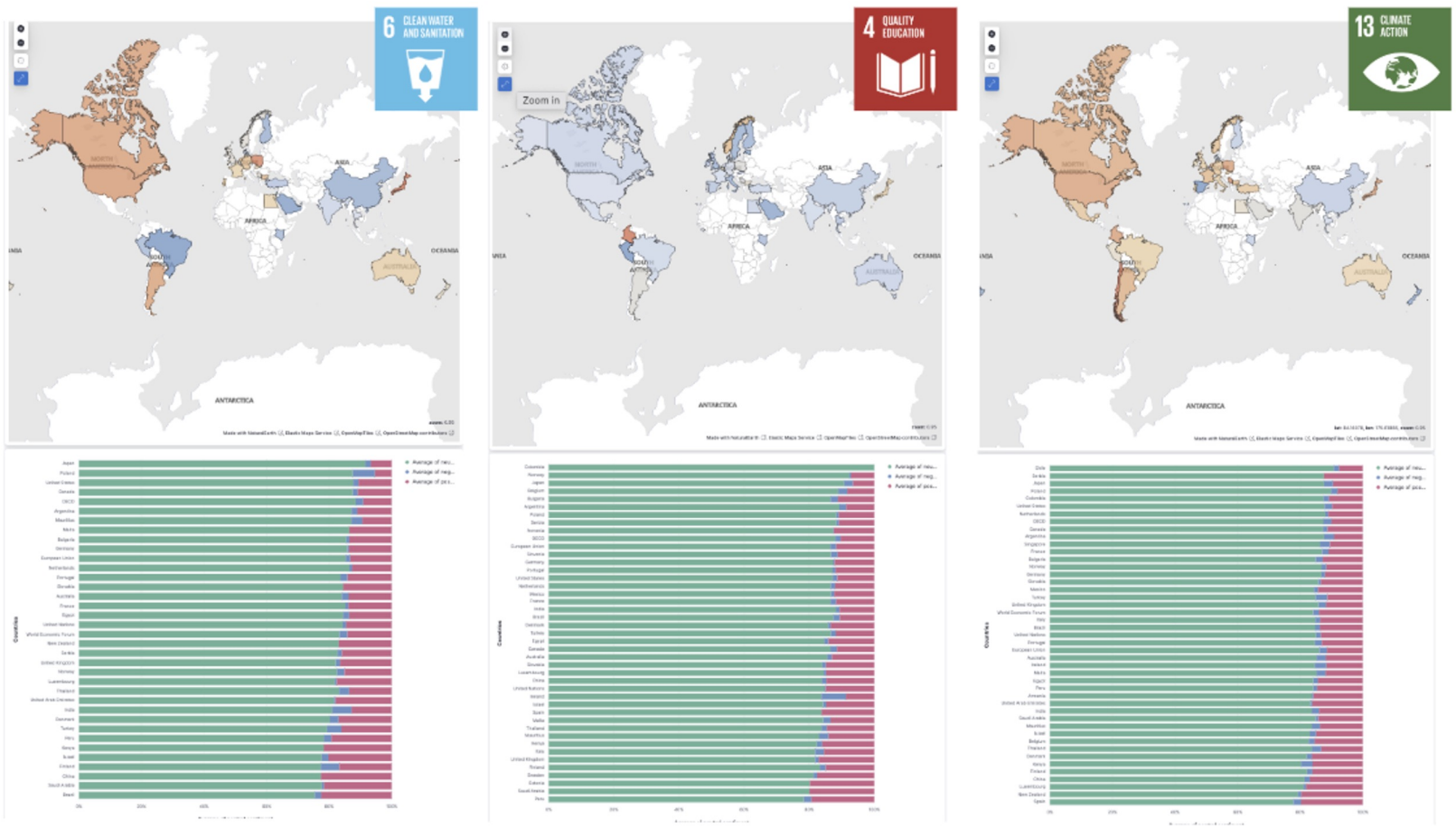
Feedback: offer participants feedback or results from the study if they express interest.

Regulation and Legislation: adhere to local, national, and international data protection laws and regulations.

Cultural Sensitivity: understand and respect cultural norms and traditions when collecting data

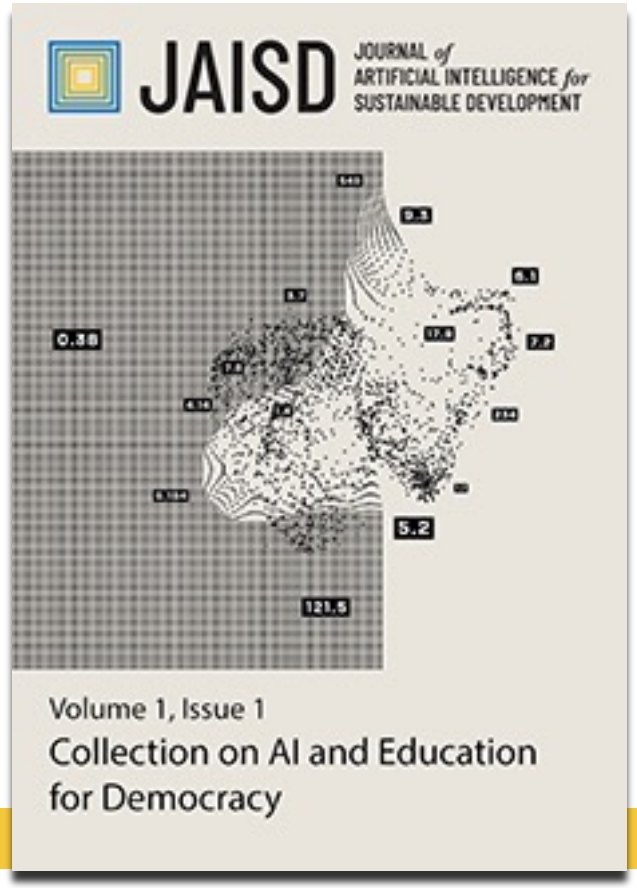
- ➔ Compute sentiment analysis (using VADER) on title + text of OECD AI policies translated
- ➔ Compare geographically the neutral sentiment on the AI policies to identify tendencies
- ➔ Differentiate between SDGs to explore the variability across the 17 targets and their topics

Sentiment Bias Aware.



AI4Gov
Trusted AI for Transparent Public Governance fostering Democratic Values

ai4gov-project.eu 101094905



Transparency: AI algorithms should be transparent in terms of how they work and how decisions are made – XAI.

Bias and Fairness: It's crucial to identify, reduce, and disclose biases to ensure fairness in decision-making.

Accountability: determine who is responsible for AI decisions.

Privacy: protect the privacy of individuals when processing their data.

Data Security: safeguard this data against breaches and unauthorized access.

Autonomy: People should always have a choice, especially concerning decisions that significantly affect their lives.

Informed Consent: Users should be informed when their data is being processed by AI and have a clear understanding of how the AI system uses their data.

Explainability: AI decisions should be interpretable and explainable.

Generalizability and Robustness: AI models are not overfitted to their training data and can generalize well to real-world situations.

Economic and Social Impacts: Consider the broader societal and economic effects of AI for good

Long-term Considerations: longer-term implications of AI, including the potential for systems to evolve or be used in ways not originally intended.

Environmental Impact: consider the environmental footprint and strive for sustainable AI research.

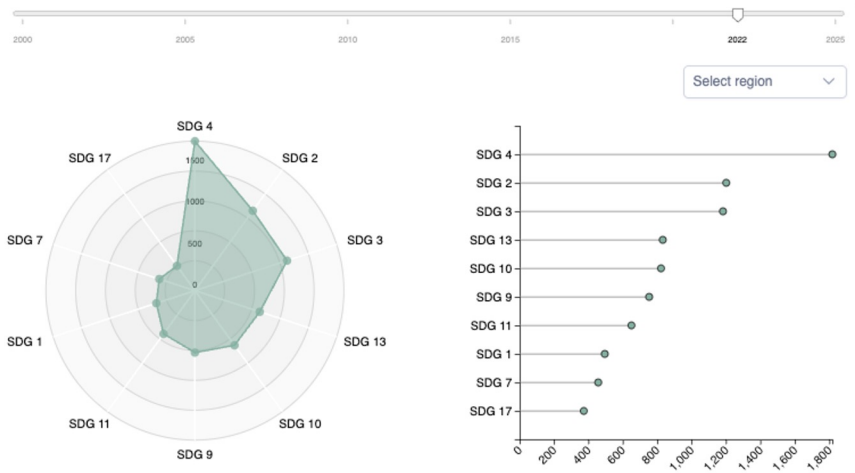
Continual Monitoring: AI systems, especially those deployed in dynamic environments, should be continuously monitored to ensure they are behaving as expected.

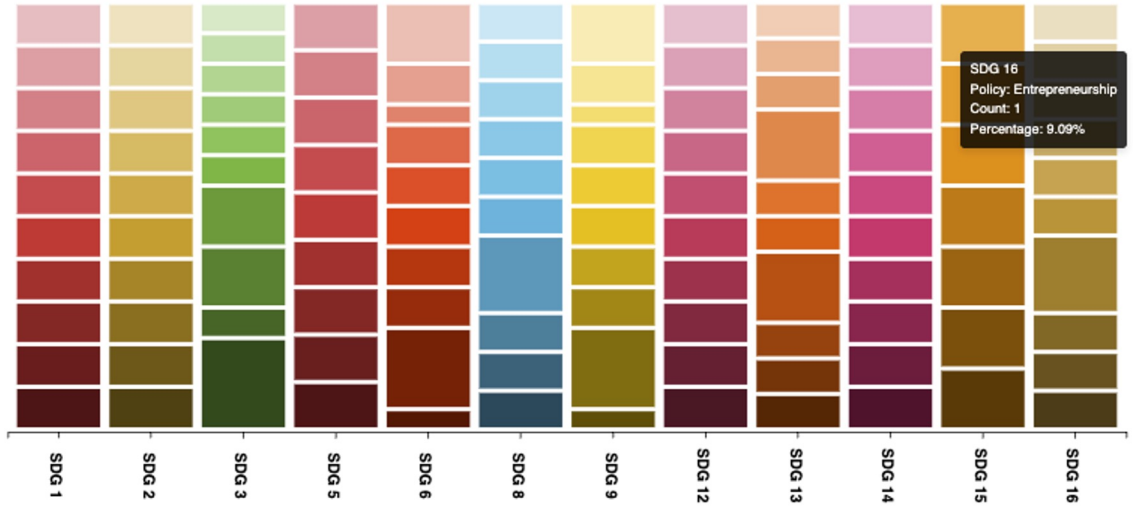
Stakeholder Participation: Involve relevant stakeholders, including those affected by AI systems, in the design, development, and deployment processes.

Regulation and Standards: Align AI practices with existing regulations and contribute to the development of ethical standards for AI.

- ➔ Evaluate the amount of SDG-classified topics across time
- ➔ Compare by geographic world regions to explore prioritisation allowing for comparison
- ➔ Encode the SDG topics based on wikidata concepts identified in the textual documents independently of their language

Coverage Bias Aware.





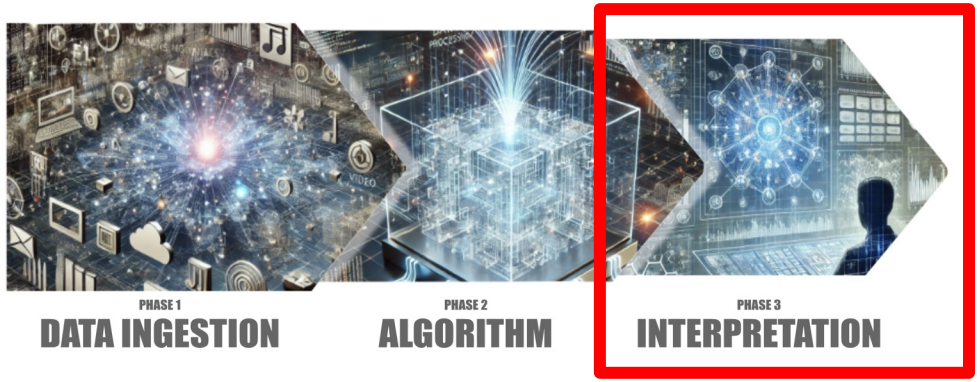
RAID

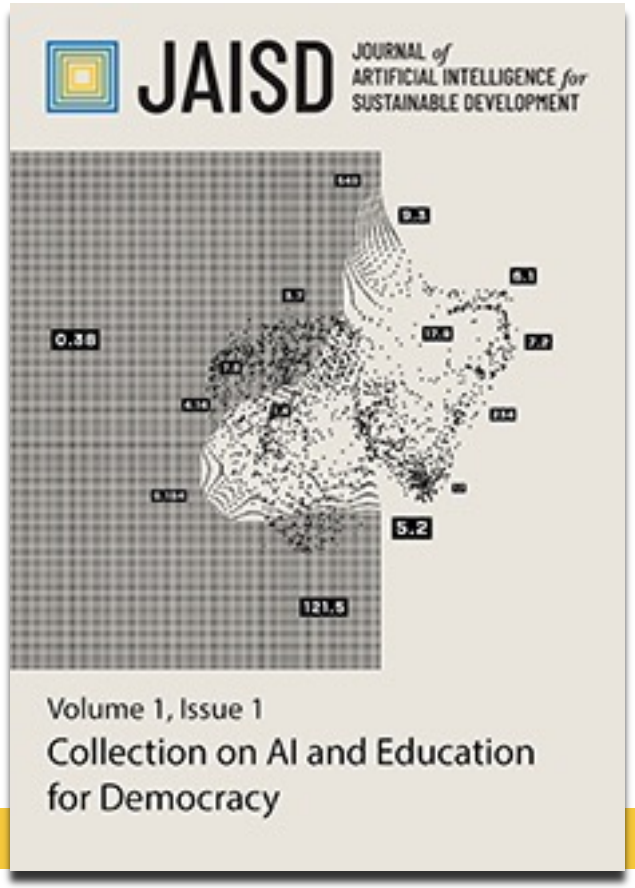
raido-project.eu

101135800

unesco Centre Under the auspices of UNESCO

IRC AI International Research Centre of Artificial Intelligence under the auspices of UNESCO





Avoiding Bias: confirmation bias (favoring information that aligns with one's existing beliefs) or availability bias (relying on immediate examples).

Transparency: clearly communicate the methods and tools used in the analysis so that others can understand and potentially reproduce the results.

Overreach: Do not make claims or conclusions that go beyond what the data actually supports.

Accuracy: Ensure that interpretations are based on accurate and robust findings. Avoid cherry-picking data or results to support a particular narrative.

Full Disclosure: Present any limitations, uncertainties, or assumptions associated with the analysis. This allows for a more nuanced understanding of the results.

Conflict of Interest: Be transparent about any potential conflicts of interest that might influence the interpretation or presentation of the results.

Sensitivity: Be aware of the broader social and cultural context when interpreting data.

Feedback Loop: Consider the potential feedback loop effects where the interpretation of data analysis can influence future data collection and results.

Stakeholder Consideration: Think about who is affected by the interpretation of the data and ensure that their interests and perspectives are considered.

Cultural Context: Especially in global or multicultural studies, understand the cultural nuances and be wary of interpreting data through only one cultural lens.

Ethical Implications: Consider the broader ethical implications of any conclusions drawn, especially if they may lead to actions or policies that impact people's lives.

Peer Review: Encourage peer review of findings and interpretations to ensure validity and mitigate potential biases.

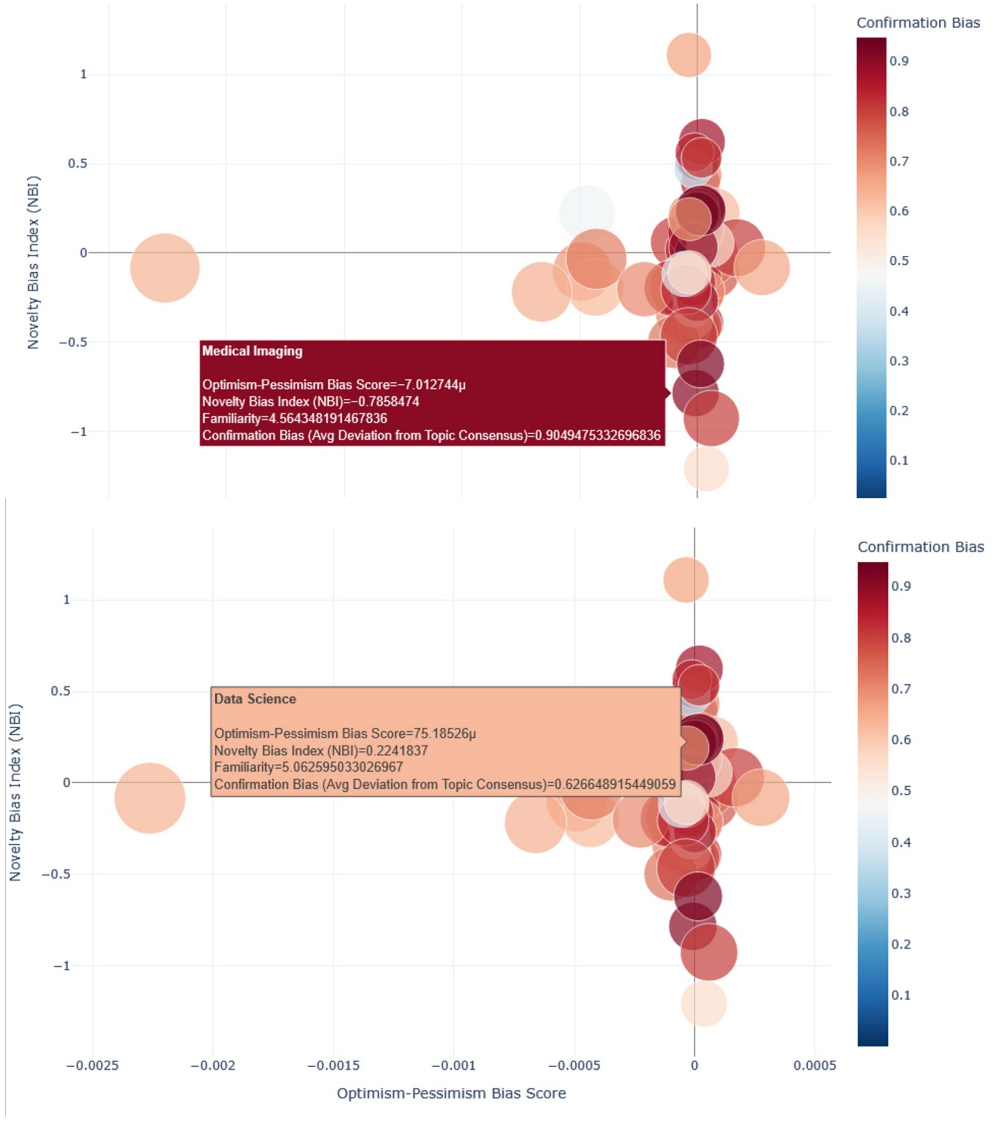
Translating to Policy: If data interpretation is used to inform policy or business decisions, ensure that the recommendations are ethically sound, justifiable, and in the best interest of those affected.

Honesty: Always approach data interpretation with honesty, and avoid the temptation to manipulate or selectively present results for personal or organizational gain.

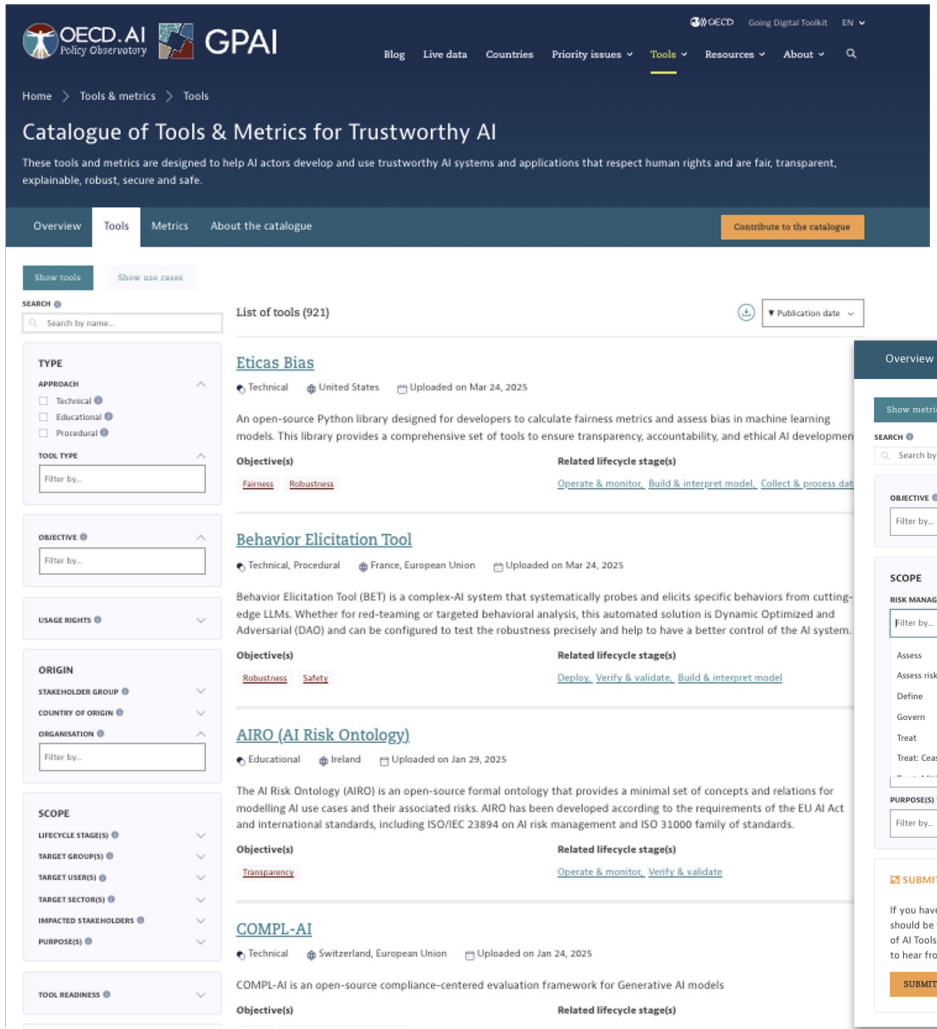
Continual Learning: Recognize that interpretations might change as more data becomes available or as methods evolve. Be open to updating or revising interpretations in the light of new evidence.

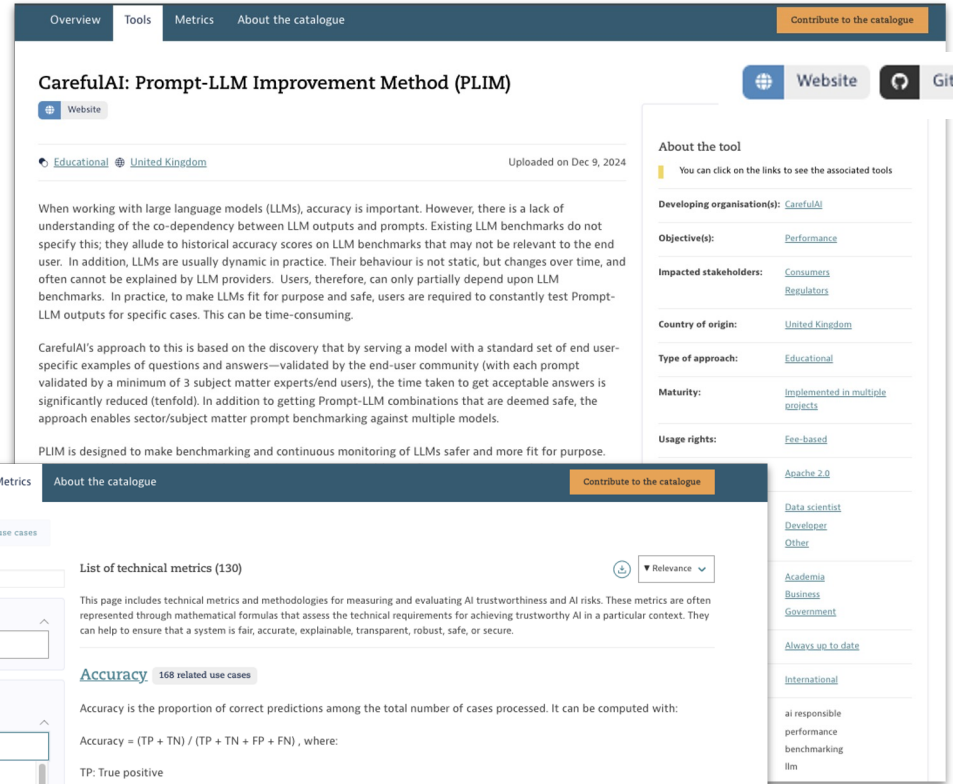
- Integration of cognitive science with computational social science.
- Empirical framework for bias detection in AI discourse.
- Use of semantic linking to align media content with AI taxonomy.
- Identification of cognitive mechanisms shaping AI adoption narratives.
- First large-scale framework connecting cognitive biases to AI innovation adoption.

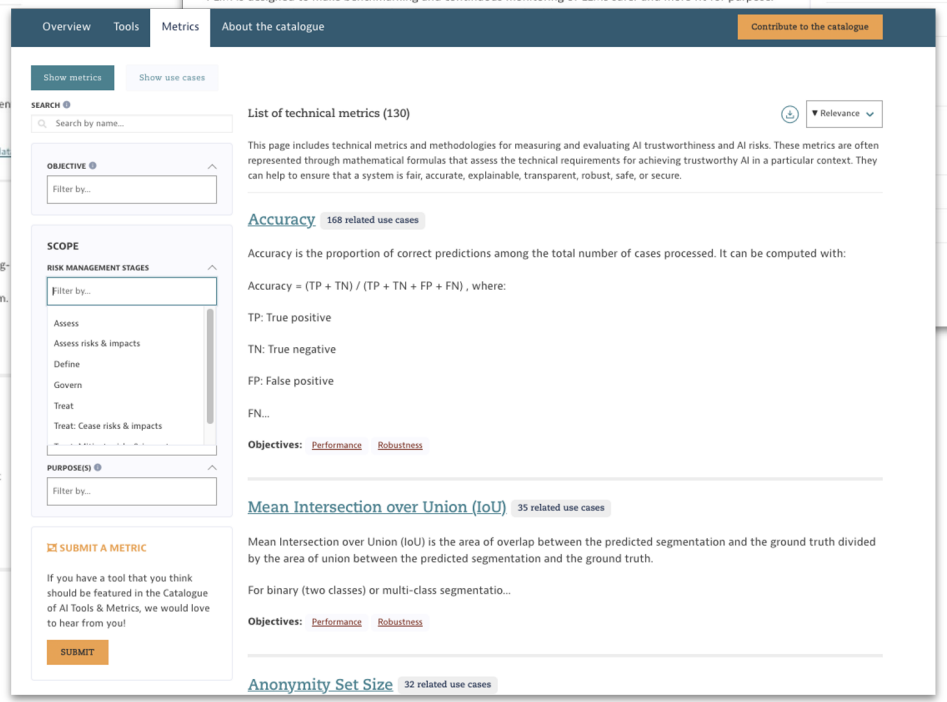
Cognitive Bias Aware.



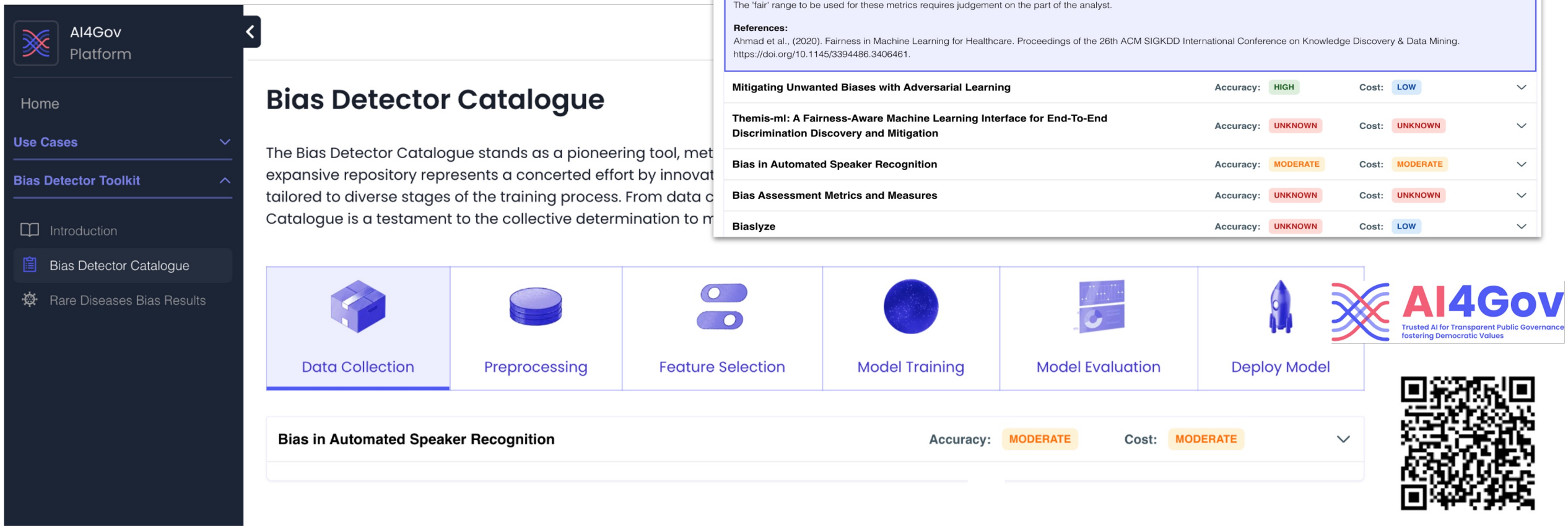
elias-ai.eu 101120237







https://oecd.ai/en/catalogue/tools



AI4Gov
Trusted AI for Transparent Public Governance fostering Democratic Values

https://cluster-ai4gov.euprojects.net/

unesco Centre Under the auspices of UNESCO

IRC AI International Research Centre of Artificial Intelligence under the auspices of UNESCO

unesco
Centre
Under the auspices
of UNESCO

IRC AI
International Research Centre
of Artificial Inteligence
under the auspices of UNESCO

International Research Centre on Artificial Intelligence under the auspices of UNESCO (Category II)

Jožef Stefan Institute, Jamova cesta 39, SI-1000 Ljubljana
E: info@ircai.org | W: https://ircai.org/

Joao Pita Costa
in/joaopitacosta/