Julia PALMA
CeADAR
Programme Manager

julia.palma@ucd.ie

CeADAR
Ireland's Centre for AI

Building Trustworthy Systems from the Inside Out

CyclOps WINTER SCHOOL

Bolzano, 13 January 2026

CyclOps

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission as the granting authority. Neither the European Union nor the granting authority can be held responsible for them

Funded by the European Union

Julia Palma
julia.palma@ucd.ie

Julia Palma

Multidisciplinary academic background with focus on **European innovation projects, public policy analysis, public affairs, and policy development** (Universidad de Sevilla, LaSalle, LSE & Cambridge).

+**300 collaborative projects** integrating research, innovation, and policy impact across digital and AI domains.

Leads **policy-focused initiatives** at Ireland's Centre for AI, contributing to **EU and national reports, consultations, and recommendations**.

Former participant in the **European Commission's Digitising European Industry** WG policy initiative.

Represents Ireland in **European Health Data Space (EHDS)** and broader **EU data space policy initiatives**.

Former chair of **BDVA's Data i-Spaces subgroup.**

Current lead of **BDVA's etami (Ethical and Trustworthy Artificial and Machine Intelligence)** subgroup.

Member of the Steering Committee of the **Global AI Policy Research Network (GlobAIpol)**

Member of **EU and international policy groups**, including the **AI Act Regulatory Sandboxes Expert Group**, **Apply AI Alliance**, **EU Health Policy Platform**, and **AI Transparency Code of Practice Working Group**.

CeADAR
Ireland’s Centre for AI

- Not-for-profit established in 2013
- Funded by IDA Ireland and Enterprise Ireland
- Based in University College Dublin

UCD School of Computer Science
Scoil na Ríomheolaíochta UCD

Enterprise Ireland

IDA Ireland
Invest in extraordinary

TECHNOLOGY CENTRE
ENTERPRISE IRELAND
IDA IRELAND SUPPORTED

CeADAR's Mission

CeADAR supports companies and organisations across Ireland to explore, experiment, and build innovative AI solutions into processes and products, transforming productivity, competitiveness, digitalisation and sustainability.

CeADAR
Ireland's Centre for AI

EDIH

An Roinn Fiontar,
Turasóireachta agus Fostaíochta
Department of Enterprise,
Tourism and Employment

Enterprise
Ireland

IDA Ireland

UCD
DUBLIN

CeADAR

AIF IRL

EuroHPC
Joint Undertaking

EuroHPC AI Factory Antennas

EDIH
European
Digital Innovation
Hubs Network

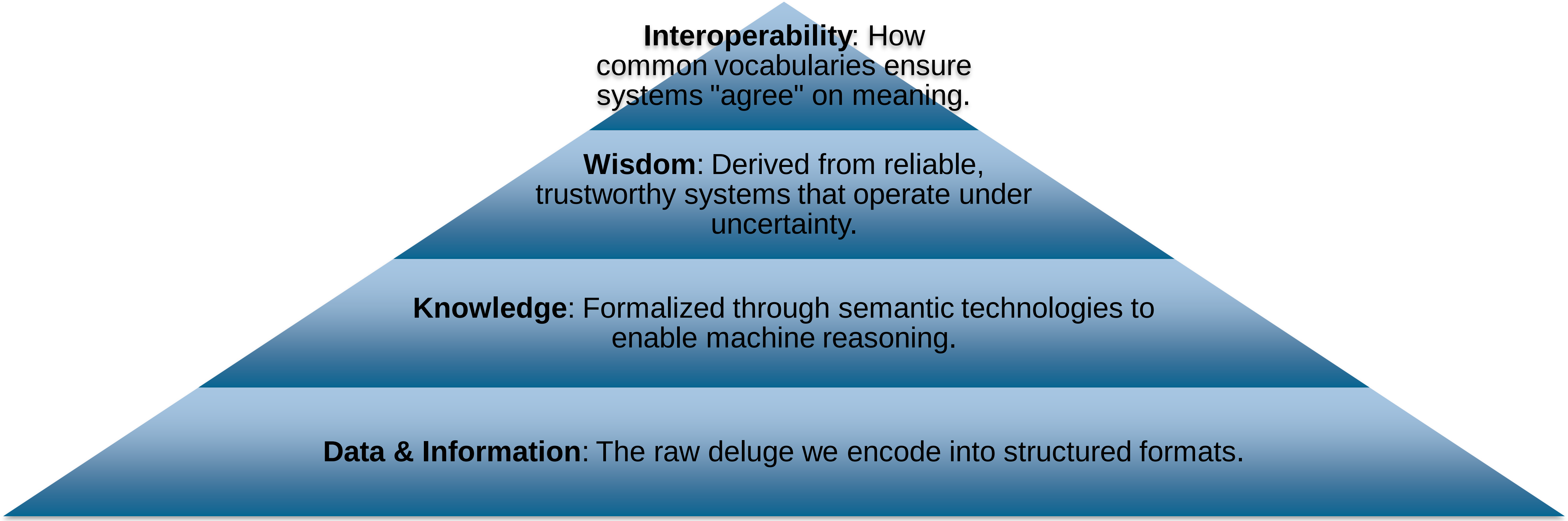
Test Before Invest

Skills Development

Support to Find Investment

Ecosystem Building

THE KNOWLEDGE PYRAMID
IN DATA PIPELINES AND DATA SPACES



Interoperability: How common vocabularies ensure systems "agree" on meaning.

Wisdom: Derived from reliable, trustworthy systems that operate under uncertainty.

Knowledge: Formalized through semantic technologies to enable machine reasoning.

Data & Information: The raw deluge we encode into structured formats.

Why Trustworthiness is Non-Optional

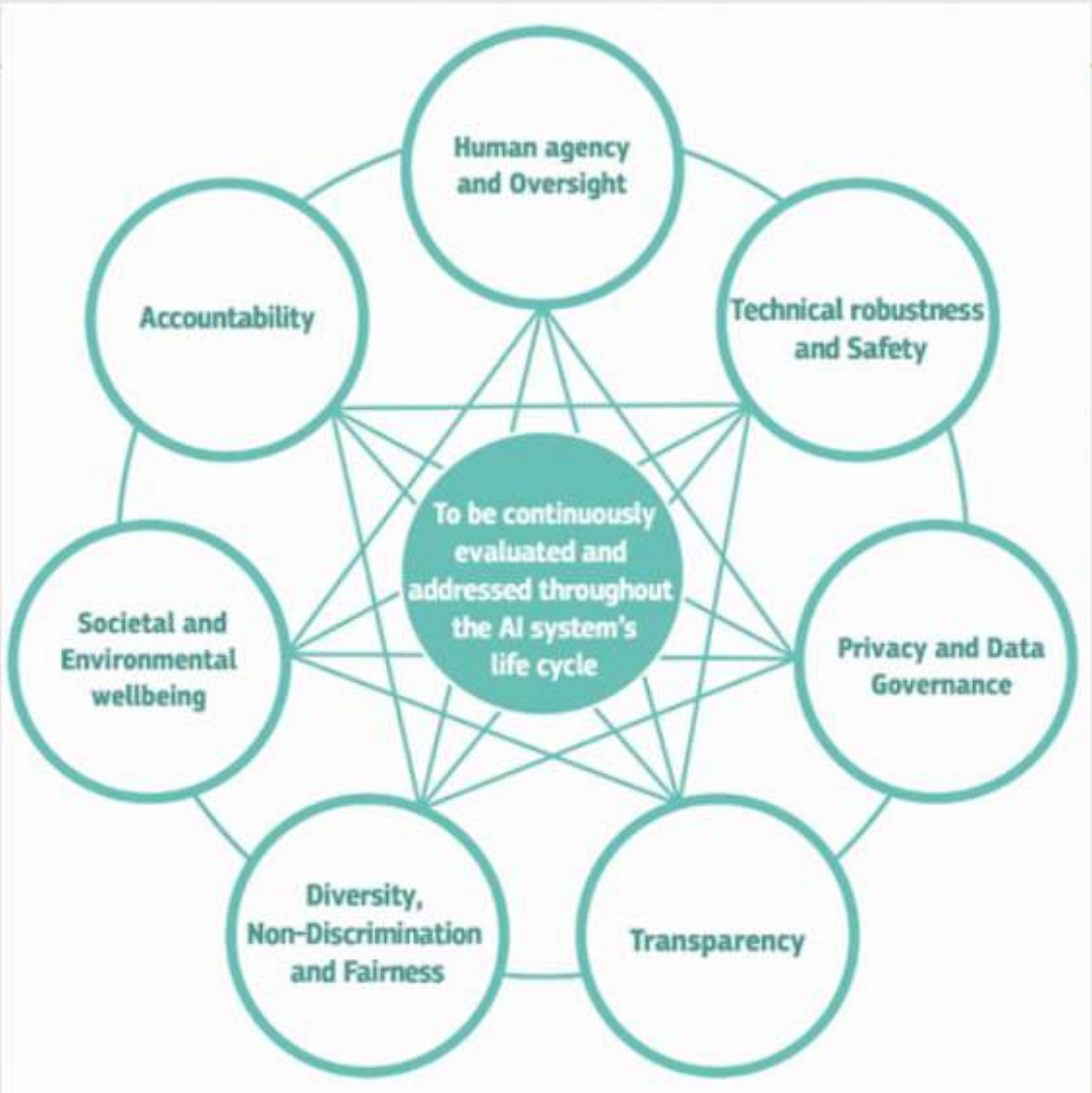
Deep Integration: KBS are no longer silos; they drive public administration and industrial decisions

The Solution: Designing for reliability, reasoning communication, and user empowerment

The Trust Gap: Black-box models erode confidence

The 7 Pillars of Trustworthy AI (HLEG)

Source: Ethics Guidelines for Trustworthy AI. Independent High-Level Expert Group on Artificial Intelligence. European commission, 8 April, 2019.


Human agency and Oversight
Technical robustness and Safety
Privacy and Data Governance
Transparency
Diversity, Non-Discrimination and Fairness
Societal and Environmental wellbeing
Accountability
To be continuously evaluated and addressed throughout the AI system's life cycle

The 7 Pillars of Trustworthy AI (HLEG) Requirements and sub-requirements

1 **Human agency and oversight:** fundamental rights

2 **Technical robustness and safety:** resilience to attack and security, fall back plan and general safety, accuracy, reliability and reproducibility

3 **Privacy and data governance :**respect for privacy, quality and integrity of data, and access to data

4 **Transparency:** traceability, explainability and communication

5 **Diversity, non-discrimination and fairness:** avoidance of unfair bias, accessibility and universal design, and stakeholder participation

6 **Societal and environmental wellbeing:** sustainability and environmental friendliness, social impact, society and democracy

7 **Accountability** auditability, minimization and reporting of negative impact, trade-offs and redress.

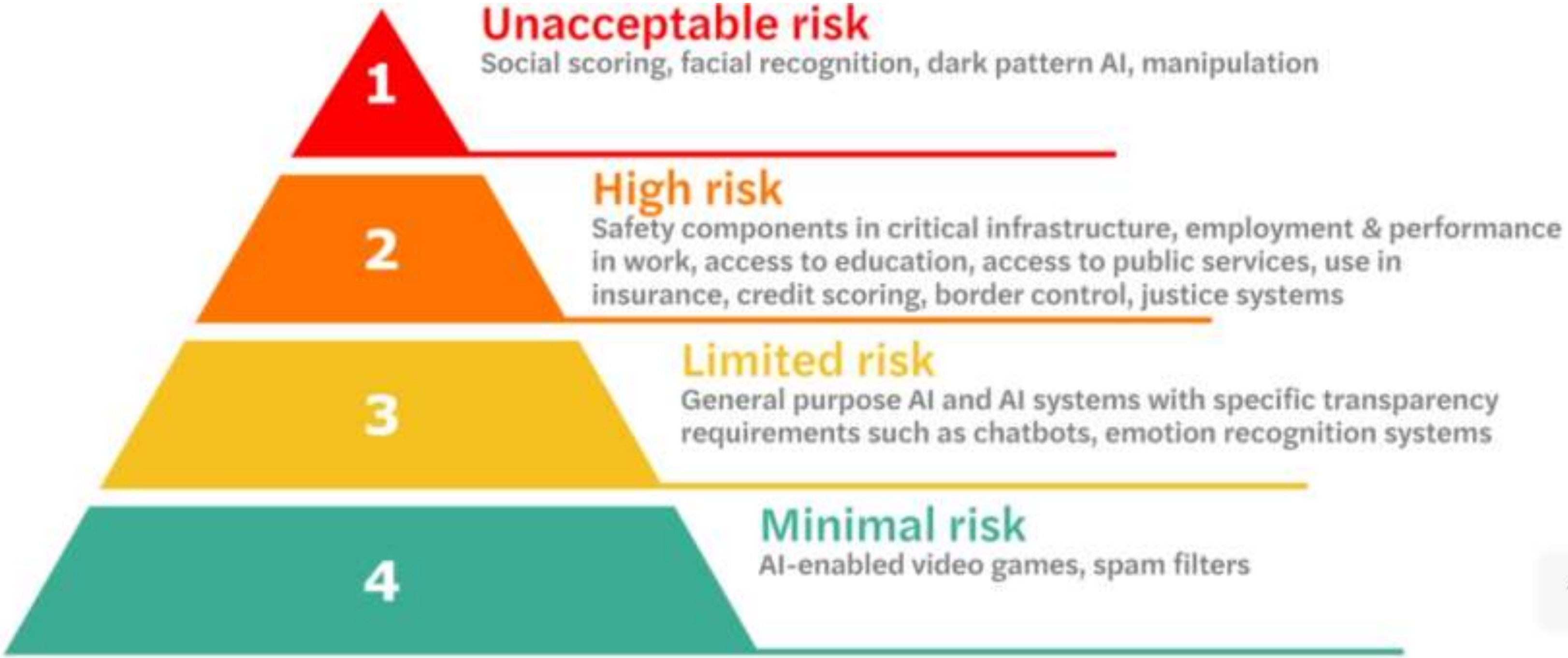
THE LEGAL REALITY: THE EU AI ACT

- **Regulation:** Moving from "Ethics Guidelines" to "Enforceable Law."
- **Risk-Based Approach:** Systems need to meet proportionate and adequate requirements based on the applications (e.g. strict documentation and transparency standards.)
- **KBS Advantage:** Explicit knowledge representation simplifies compliance (Traceability by design) to meet the **"Data Governance"** and **"Technical Documentation"** requirements of the Act.

THE EU AI ACT

The **EU AI Act** (passed in 2024) provides the legal enforcement using a **Risk-Based Approach:**

- **Unacceptable Risk:** Prohibited.
- **High Risk:** Strict compliance, logging, and human oversight required.
- **Limited/Minimal Risk:** Transparency obligations (e.g.,chatbot disclosing it's an AI).



1 **Unacceptable risk**
Social scoring, facial recognition, dark pattern AI, manipulation

2 **High risk**
Safety components in critical infrastructure, employment & performance in work, access to education, access to public services, use in insurance, credit scoring, border control, justice systems

3 **Limited risk**
General purpose AI and AI systems with specific transparency requirements such as chatbots, emotion recognition systems

4 **Minimal risk**
AI-enabled video games, spam filters

Julia Palma
julia.palma@ucd.ie

IMPACT ASSESSMENT FOR HIGH-RISK AI SYSTEMS

EU AI Act: Article 27: Fundamental Rights Impact Assessment for High-Risk AI Systems

- *Deployers prior to deploying AI of high-risk, shall perform an assessment of the impact on fundamental rights that the use of such system may produce.*

Where's the line

LEGAL: A **fundamental rights-based approach** is more closely linked to **existing law** and focuses on aspects that are legally relevant and thus **enforceable**.

ETHICAL: Compared to this, an **ethics-based approach** is much broader and also more open to reflection on **potential** implications that may not be worth considering from a legal perspective.

European
Digital Innovation
Hubs Network

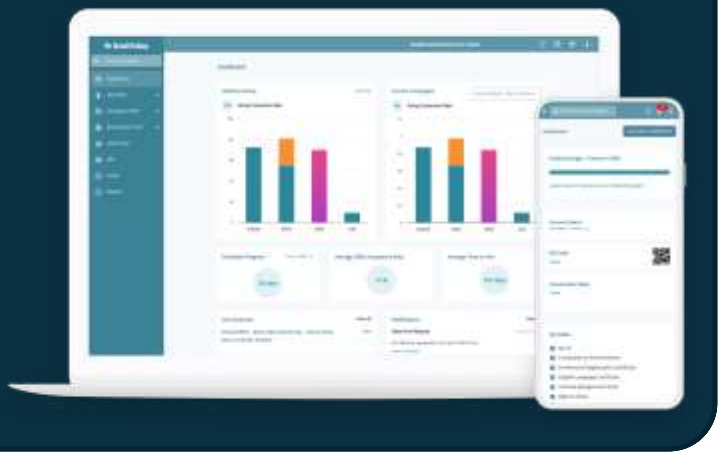
First comes the Regulation; then, Trust Awareness

80% companies reaching out for 'compliance' and obligations of 'transparency'

healthdaq®

LLMs to parse candidates

Human-oversight



An Roinn Fiontar, Turasóireachta agus Fostaíochta
Department of Enterprise, Tourism and Employment

Julia Palma
julia.palma@ucd.ie

CeADAR
Ireland's Centre for AI

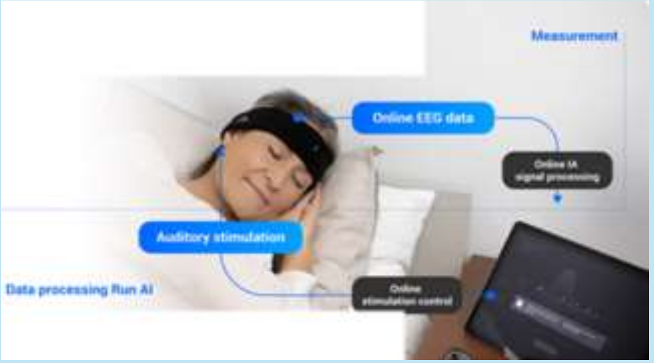
MANOLO is a Horizon Europe project led by CeADAR for **trustworthy and lighter, high-quality AI models** in cloud-edge distributed environments

MANOLO
CLOUD-EDGE EFFICIENT & TRUSTWORTHY AI

PAL

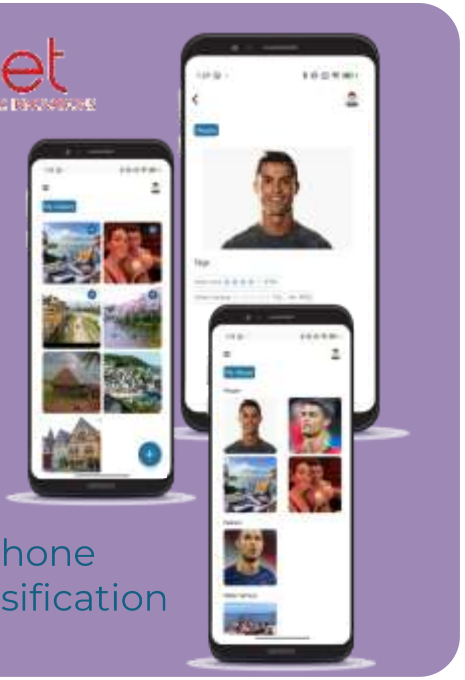
Collaborative Assembly in a Manufacturing Environment

Bitbrain



Auditory Stimulation for Memory Consolidation

arx.net



Edge Phone Image Classification

Part of Adra
The AI Data Robotics Association

Funded by the European Union under GA no. 101135782. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or CNECT. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

Julia Palma
julia.palma@ucd.ie

Embedding Trustworthy AI across the innovation cycle

CeADAR
Ireland's Centre for AI

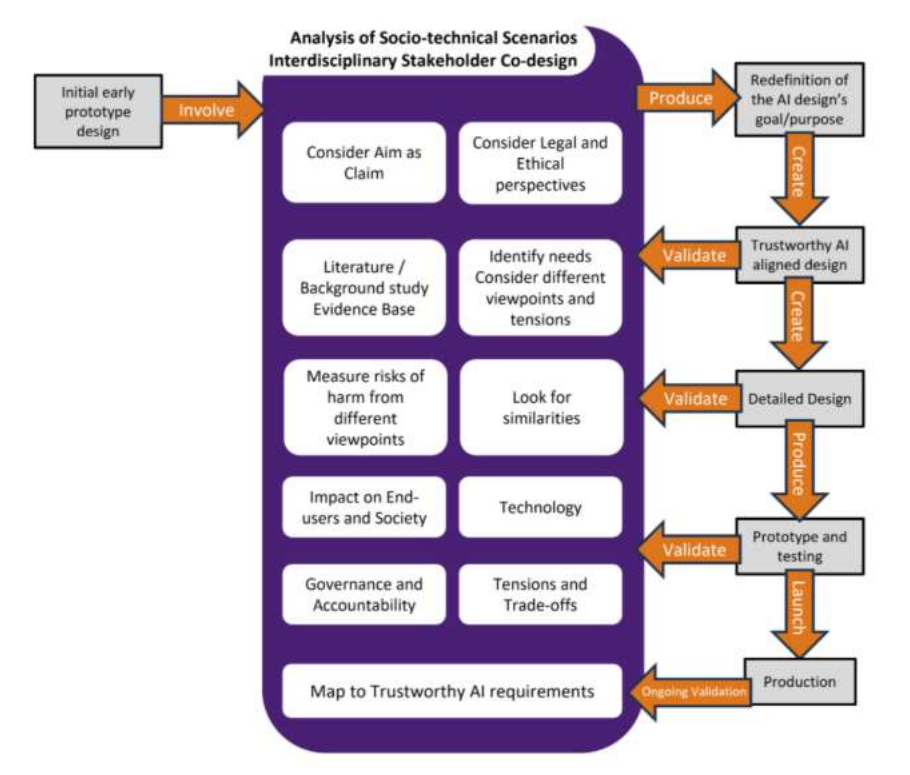

Analysis of Socio-technical Scenarios
Interdisciplinary Stakeholder Co-design
Initial early prototype design
Involve
Consider Aim as Claim
Consider Legal and Ethical perspectives
Literature / Background study Evidence Base
Identify needs Consider different viewpoints and tensions
Measure risks of harm from different viewpoints
Look for similarities
Impact on End-users and Society
Technology
Governance and Accountability
Tensions and Trade-offs
Map to Trustworthy AI requirements
Produce
Redefinition of the AI design's goal/purpose
Create
Validate
Trustworthy AI aligned design
Create
Validate
Detailed Design
Produce
Validate
Prototype and testing
Launch
Production
Ongoing Validation

Figure 2: MANOLO Z-Inspection® Co-Design Framework

Ethical Issues and Tensions

User privacy, safety and data protection vs Transparency

Accountability

User autonomy vs technical robustness/bias/justice

Regulatory Issues and Tensions

Compliance with GDPR, AI Act, Cybersecurity, domestic/sectoral regulation

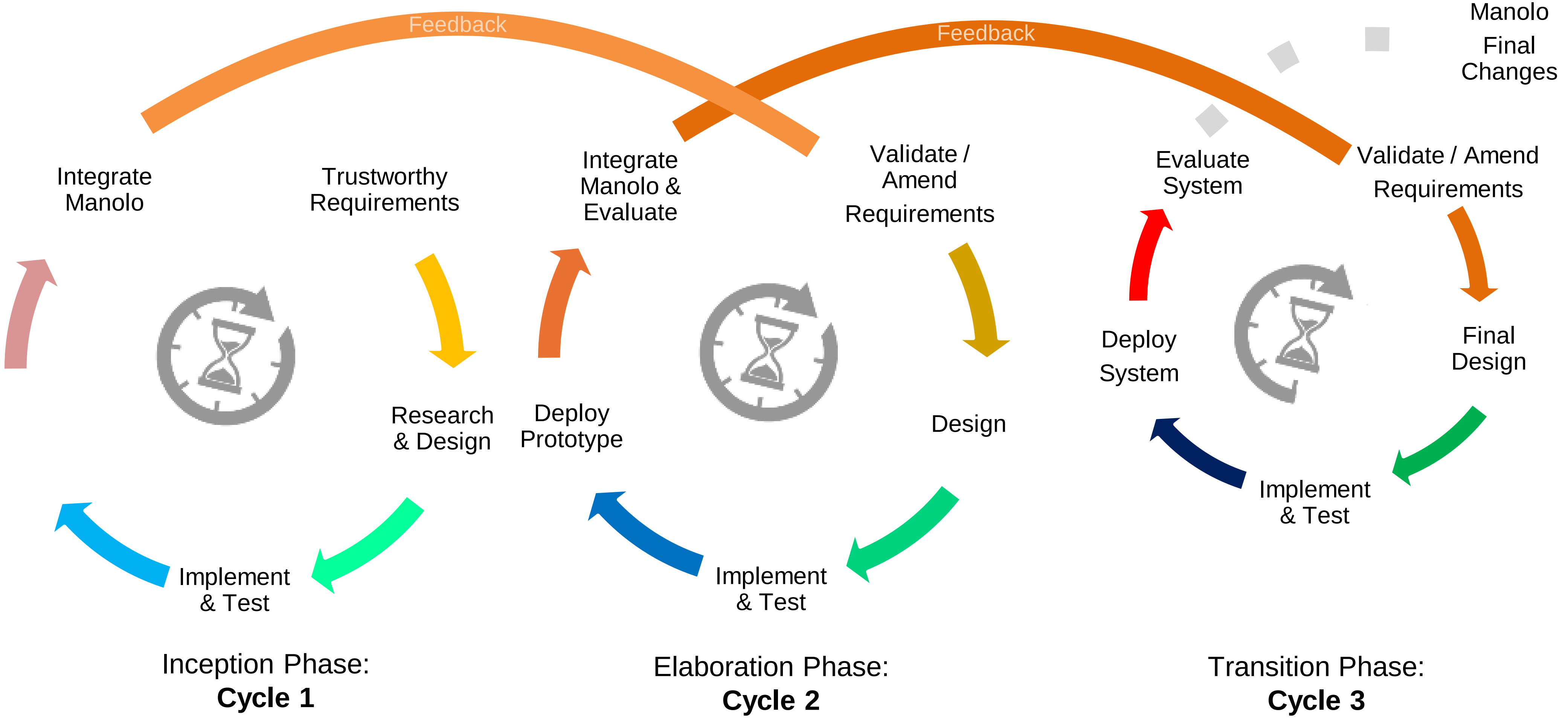
KPI Assessment

Mitigation Measures

Aligned technological innovation with ethical considerations, ensuring transparency, privacy, inclusivity, and accountability based on stakeholders' feedback

Julia Palma
julia.palma@ucd.ie

Integration Approach & Mapping



Feedback

Feedback

Manolo Final Changes

Integrate Manolo

Trustworthy Requirements

Research & Design

Implement & Test

Inception Phase: **Cycle 1**

Integrate Manolo & Evaluate

Validate / Amend Requirements

Design

Implement & Test

Deploy Prototype

Elaboration Phase: **Cycle 2**

Evaluate System

Validate / Amend Requirements

Final Design

Implement & Test

Deploy System

Transition Phase: **Cycle 3**

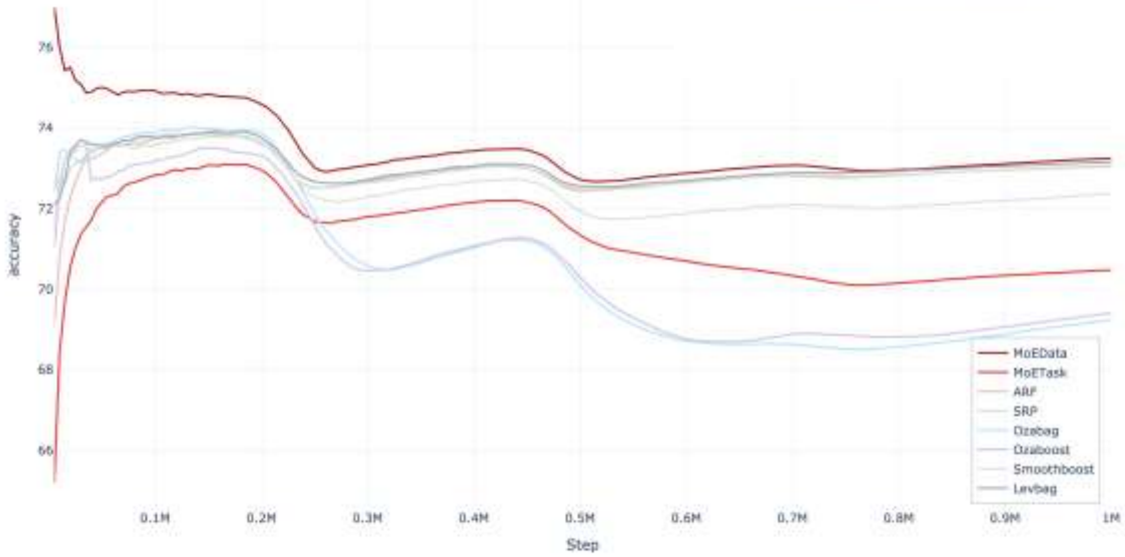
Julia Palma
julia.palma@ucd.ie

MoE-drift (Mixture of Experts Approach to Handle Concept Drift).

The Problem: Concept drift (when real-world data changes and models fail).

Function: A framework using a neural router to detect when specific "expert" nodes in a system are no longer accurate due to data shifts.

Impact: Ensures **Technical Robustness** in non-stationary environments.



Accuracy over time plot for LED_g dataset

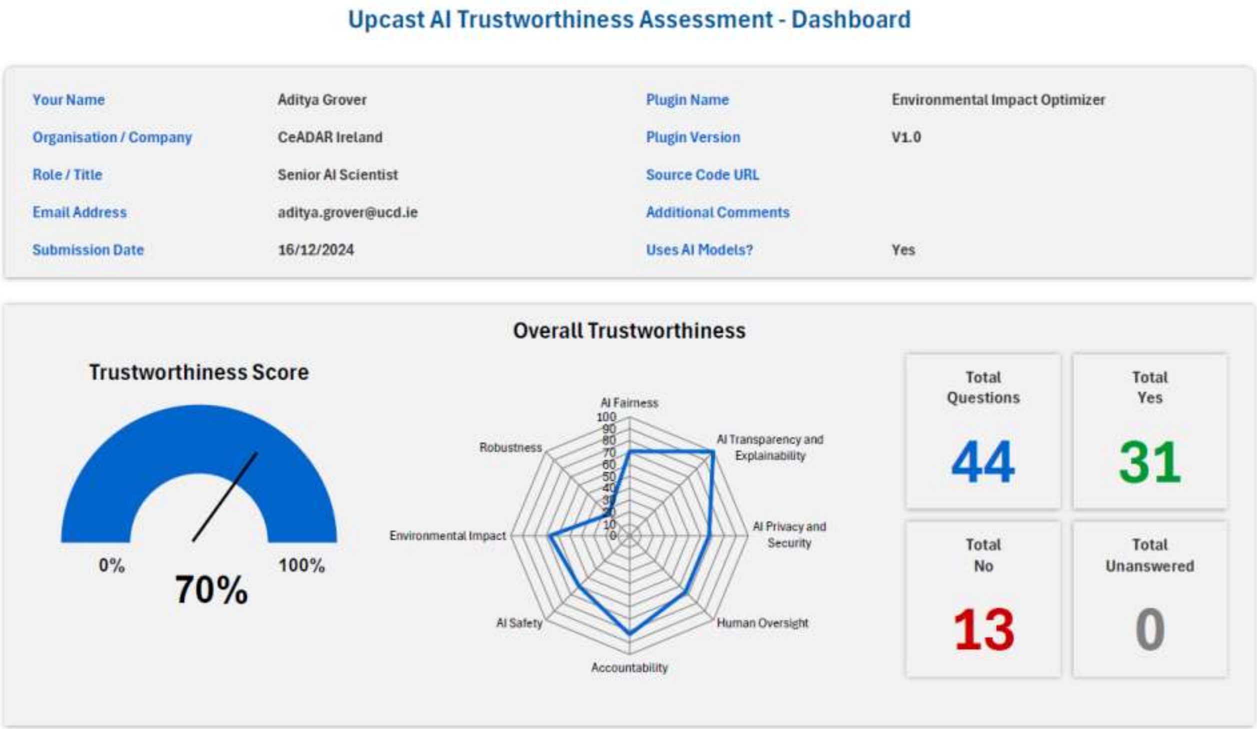
Part of The AI Data Robotics Association

Funded by the European Union under GA no. 101135782. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or CNECT. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

Julia Palma
julia.palma@ucd.ie

UPCAST is a Horizon Europe project where CeADAR developed the **AI Trustworthiness Assessment Tools and Dashboard**



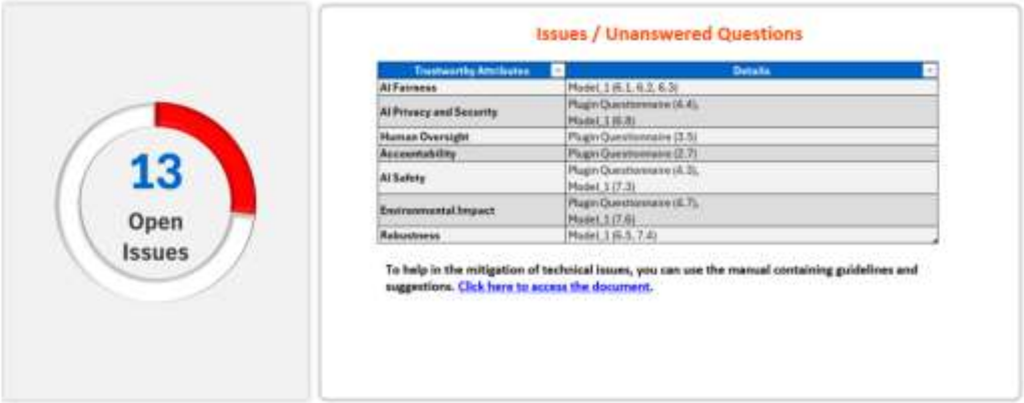
Upcast AI Trustworthiness Assessment - Dashboard

Your Name	Aditya Grover	Plugin Name	Environmental Impact Optimizer
Organisation / Company	CeADAR Ireland	Plugin Version	V1.0
Role / Title	Senior AI Scientist	Source Code URL	
Email Address	aditya.grover@ucd.ie	Additional Comments	
Submission Date	16/12/2024	Uses AI Models?	Yes

Overall Trustworthiness

Total Questions	Total Yes	Total No	Total Unanswered
44	31	13	0

UNIVERSAL PLATFORM COMPONENTS FOR SAFE FAIR INTEROPERABLE DATA EXCHANGE, MONETISATION AND TRADING



Issues / Unanswered Questions

Trustworthy Attributes	Details
AI Fairness	Model_1 (6.1, 6.2, 6.3)
AI Privacy and Security	Plugin Questionnaire (4.4), Model_1 (6.8)
Human Oversight	Plugin Questionnaire (3.5)
Accountability	Plugin Questionnaire (2.7)
AI Safety	Plugin Questionnaire (4.3), Model_1 (7.3)
Environmental Impact	Plugin Questionnaire (4.7), Model_1 (7.6)
Robustness	Model_1 (6.5, 7.4)

To help in the mitigation of technical issues, you can use the manual containing guidelines and suggestions. Click here to access the document.

Figure 6: Open Issues

Julia Palma
julia.palma@ucd.ie

CeADAR
Ireland's Centre for AI

UPCAST Valuation Plug-in

•**The Principle:** Fairness and Accountability. Transparency.

• **Function:** for data providers - setting fair prices for their data products; for data consumers - estimate costs for data acquisition. Using explainable AI techniques (XAI), also provides transparency by identifying the key factors influencing the price.

• **Impact:** Prevents power asymmetries in the Common European Data Spaces.

UPCAST PROJECT

UNIVERSAL PLATFORM COMPONENTS FOR SAFE FAIR INTEROPERABLE DATA EXCHANGE, MONETISATION AND TRADING

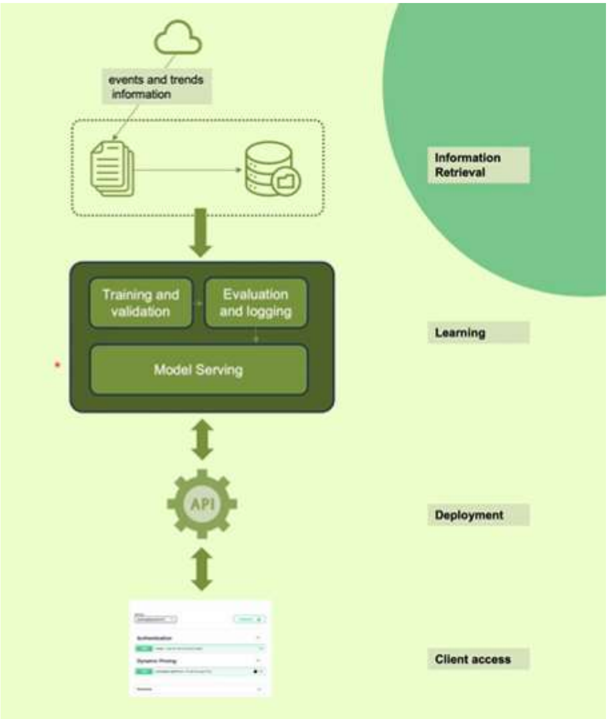
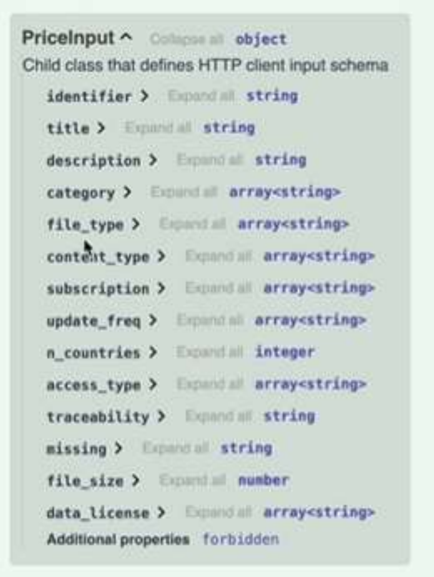
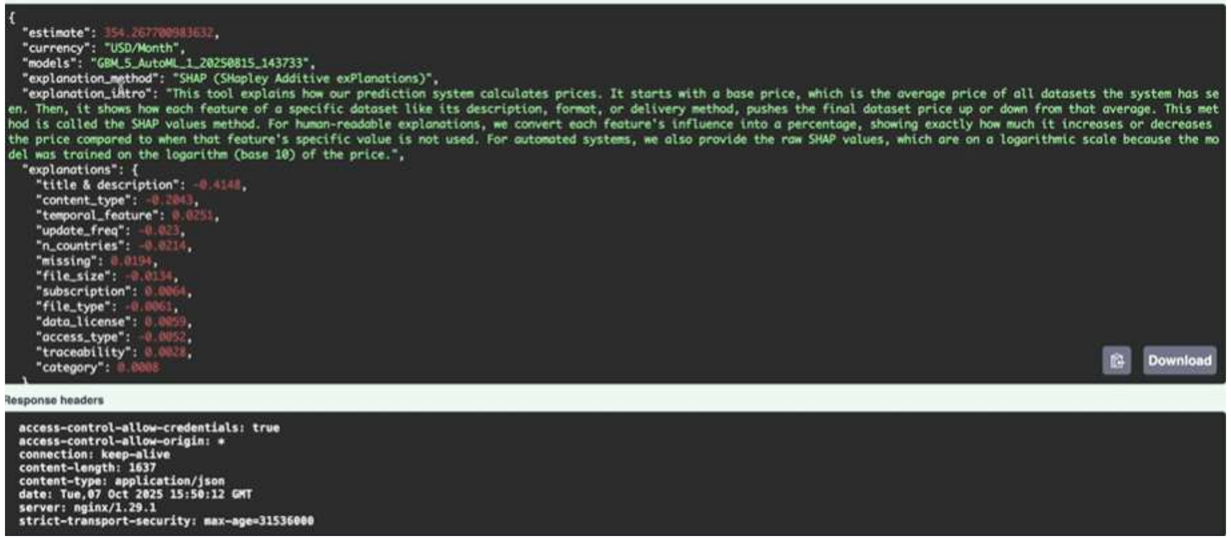


Figure: Overall plugin architecture



PriceInput ^ Collapse all object
Child class that defines HTTP client input schema

```
identifier >       Expand all  string
title >            Expand all  string
description >      Expand all  string
category >         Expand all  array<string>
file_type >        Expand all  array<string>
content_type >     Expand all  array<string>
subscription >     Expand all  array<string>
update_freq >      Expand all  array<string>
n_countries >      Expand all  integer
access_type >      Expand all  array<string>
traceability >     Expand all  string
missing >          Expand all  string
file_size >        Expand all  number
data_license >     Expand all  array<string>
Additional properties  forbidden
```



```
{
  "estimate": 354.267700983632,
  "currency": "USD/Month",
  "models": "GBM_5_AutoML_1_20250815_143733",
  "explanation_method": "SHAP (SHapley Additive exPlanations)",
  "explanation_intro": "This tool explains how our prediction system calculates prices. It starts with a base price, which is the average price of all datasets the system has se
en. Then, it shows how each feature of a specific dataset like its description, format, or delivery method, pushes the final dataset price up or down from that average. This met
hod is called the SHAP values method. For human-readable explanations, we convert each feature's influence into a percentage, showing exactly how much it increases or decreases
the price compared to when that feature's specific value is not used. For automated systems, we also provide the raw SHAP values, which are on a logarithmic scale because the mo
del was trained on the logarithm (base 10) of the price.",
  "explanations": {
    "title & description": -0.4148,
    "content_type": -0.2043,
    "temporal_feature": 0.0251,
    "update_freq": -0.023,
    "n_countries": -0.0214,
    "missing": 0.0194,
    "file_size": -0.0134,
    "subscription": 0.0064,
    "file_type": -0.0061,
    "data_license": 0.0059,
    "access_type": -0.0052,
    "traceability": 0.0028,
    "category": 0.0008
  }
```

Response headers

```
access-control-allow-credentials: true
access-control-allow-origin: *
connection: keep-alive
content-length: 1637
content-type: application/json
date: Tue,07 Oct 2025 15:50:12 GMT
server: nginx/1.29.1
strict-transport-security: max-age=31536000
```

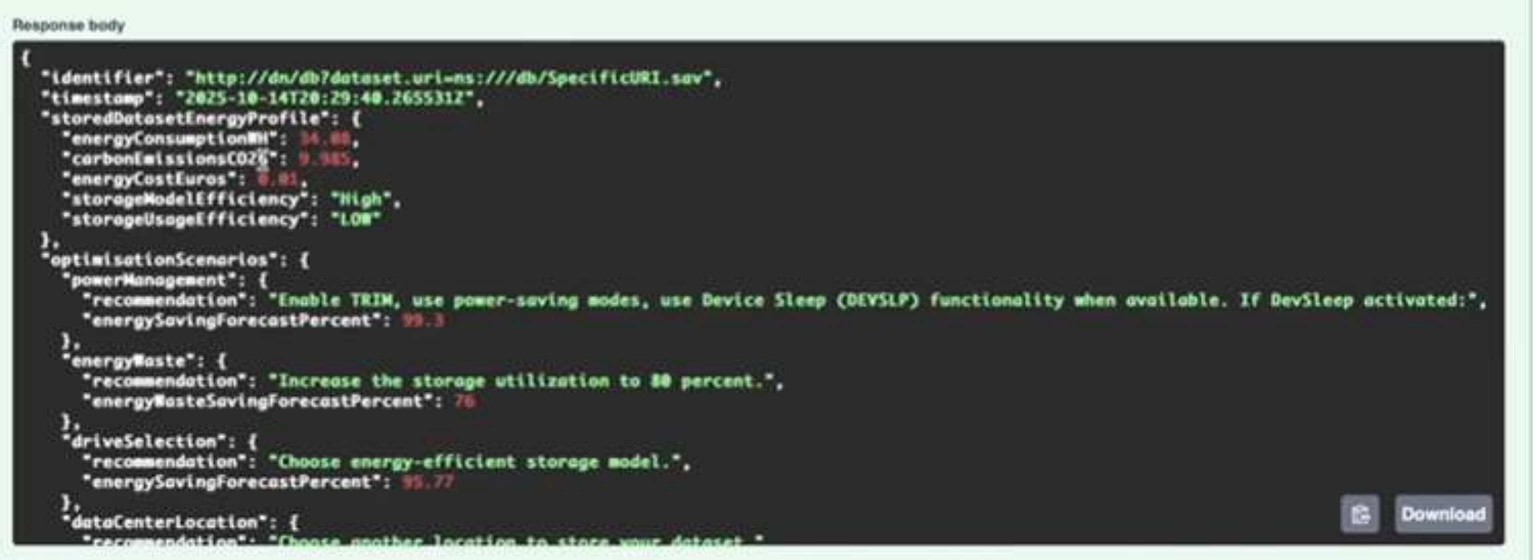
Julia Palma
julia.palma@ucd.ie

UPCAST Environmental Plug-in

UPCAST PROJECT

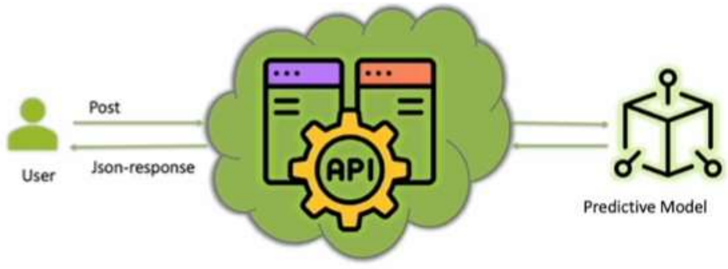
UNIVERSAL PLATFORM COMPONENTS FOR SAFE FAIR INTEROPERABLE DATA EXCHANGE, MONETISATION AND TRADING

•**The Principle:** Societal and Environmental Well-being. Transparency.
•**Function:** estimates energy costs for data storage and processing, identifies and explains the factors affecting energy consumption using explainable AI techniques (XAI). It also provides real-time monitoring of energy usage during data processing.
•**Impact:** Enables "Green Data Spaces" where users choose low-impact reasoning paths.



Response body

```
{
  "identifier": "http://dn/db?dataset.uri=ns:///db/SpecificURI.sav",
  "timestamp": "2025-10-14T20:29:40.265531Z",
  "storedDatasetEnergyProfile": {
    "energyConsumptionWH": 34.08,
    "carbonEmissionsCO2G": 9.985,
    "energyCostEuros": 0.01,
    "storageModelEfficiency": "High",
    "storageUsageEfficiency": "LOW"
  },
  "optimisationScenarios": {
    "powerManagement": {
      "recommendation": "Enable TRIM, use power-saving modes, use Device Sleep (DEVSLP) functionality when available. If DevSleep activated:",
      "energySavingForecastPercent": 99.3
    },
    "energyWaste": {
      "recommendation": "Increase the storage utilization to 80 percent.",
      "energyWasteSavingForecastPercent": 76
    },
    "driveSelection": {
      "recommendation": "Choose energy-efficient storage model.",
      "energySavingForecastPercent": 95.77
    },
    "dataCenterLocation": {
      "recommendation": "Choose another location to store your dataset."
```



Enables estimation of key environmental metrics such as energy consumption, carbon emissions, and costs

Offers two primary endpoints to estimate energy profiles:

- One for dataset storage energy profile at idle
- One for data processing workflow (DPW) energy profile

Julia Palma
julia.palma@ucd.ie

CeADAR
Ireland's Centre for AI

CyclOps

We are the partner providing….

- AI-based orchestration (AIMP)
- AI-Trustworthiness tools (XAI)
- Requirement collection coordination (T2.3 lead)

…in WPs

- WP3: Runtime Layer
- WP5: IKB Exploitation

We are in the project since…

CyclOps facilitates **collaboration** with diverse partners in **data space ecosystems**. Providing technological innovation with industrial relevance through the development of **automated AI-based trustworthy** components.

As an outcome we expect…

- Facilitating the connection to various data sources and repositories through the **CyclOps AI tools & models marketplace**.
- Early access to **knowledge and CyclOps tools** for the management of the end-to-end data lifecycle for data space interoperability.
- Building **ecosystem connections** for potential collaborations, knowledge exchange, and opportunities in future projects.

Our key people are…

Ricardo Simón-Carbajo
Director of Innovation and Development

Julia Palma
Innovation Programmes Manager

Hanene Jemoui
Senior Software Engineer

Vicky Kumar
Research Software Engineer

Saul Burgess
Research Software Engineer

CyclOps

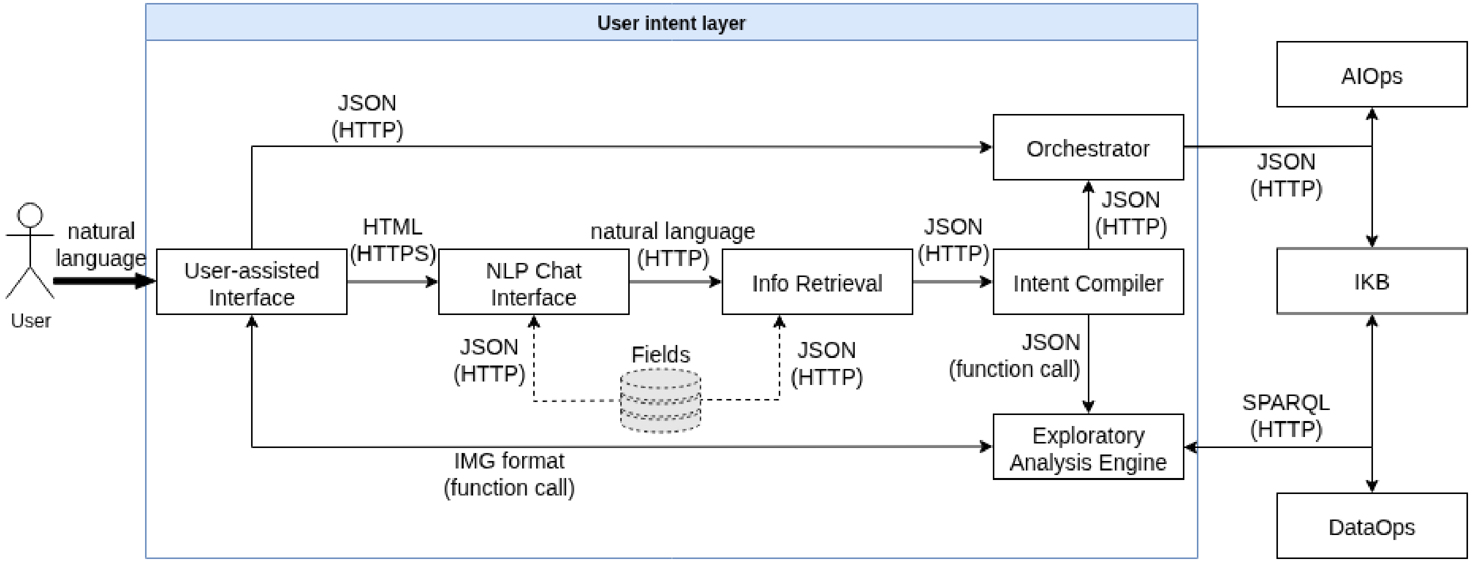
AI Marketplace (AIMP)

CyclOps AI Marketplace Module

•**The Principle:** Robustness. Transparency.

• **Function:** AI-based dashboard supporting automated algorithm matchmaking for users to efficiently select algorithms from a large catalogue of algorithms in use case scenarios such as Tourism, Green Deal, Public Procurement, and Manufacturing

•**Impact:** Efficiency for data processing and AI-based solutions development



•Algorithm Metadata Schema

```
{
  “id”: “Identifier for the algorithm”,
  ”repo_id”: “Identifier for the repository hosting the algorithm”,
  “dataset_id”: “Identifier for the selected dataset”,
  “name”: “Name of the algorithm”,
  “description”: “Textual description of the algorithm”,
  “category”: [“Machine Learning”, “Deep Learning”, “Other”],
  “type”: [“Classification”, “Regression”, “Clustering”, “Anomaly Detection”, “Forecasting”, “Recommendation”],
  “applications”: [“Remote sensing”, “Land cover classification”, “Change detection”, “Tourism data segmentation”],
  “data_format”: [“Numerical”, “Categorical”, “Textual”, “Imagery”, “Time series”, “Sequential”, “Acoustic”, “Geospatial”],
  “optimizer”: [“sgd”, “mini-batch gradient descent”, “adam”, “adamw”],
  “url”: “Web address of the algorithm’s source code or library”
}
```

•User Intent Metadata Schema

```
{
 “user_intent”: {
     “id”: “Identifier for the user intent”,
     ”session_id”: “Identifier for the user session”,
     “dataset_id”: “Identifier for the selected dataset”,
     “application_requirements”: {
        “application_type”: “Selected application type such as Classification, Regression, Clustering, etc”,
        “application_domain”: “Selected application domain such as Tourism, Green Deal, Public Procurement, etc”,
        “description”: “Textual description of the user query”
     },
     “functional_requirements”: {
        “accuracy”: [“High”, ”Medium”, “Low”],
        “training_time”: [“High”, ”Medium”, “Low”],
        “inference_time”: [“High”, “Medium”, “Low”],
        “scalability”: [“High”, “Medium”, “Low”]
     },
     “infrastructure_constraints”: {
        “cpu_cores”: “Number of CPU cores available”,
        “ram_gb”: “Amount of RAM available in GB”,
        “gpu_gb”: “Amount of GPU memory available in GB”,
        “platform”: [“on-premise”, “cloud”, “edge”]
     }
}
```

CyclOps

AI Marketplace (AIMP)

Julia Palma
julia.palma@ucd.ie

CyclOps

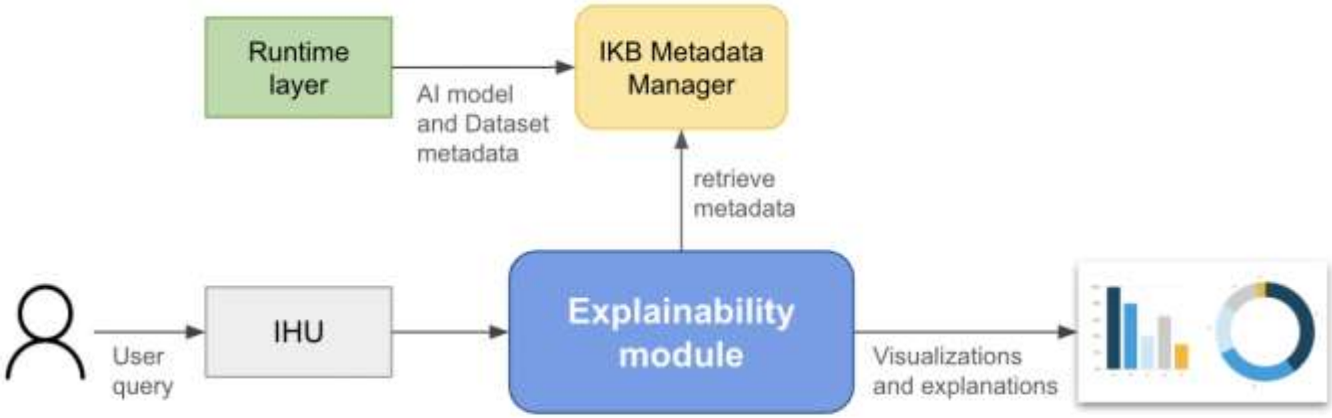
Multimodal Explainability Module (MXAI)

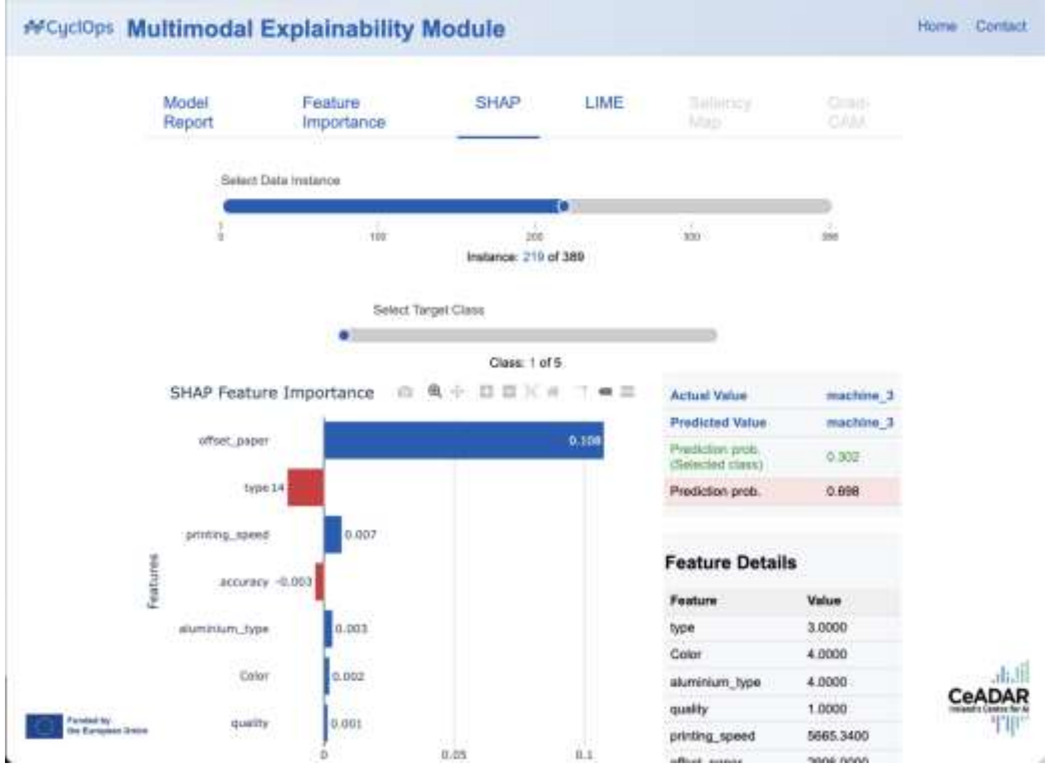
CyclOps Multimodal Explainability Module

•**The Principle:** Transparency. Human oversight.

• **Function:** AI-based dashboard supporting automated algorithm matchmaking for users to efficiently select algorithms from a large catalogue of algorithms in use case scenarios such as Tourism, Green Deal, Public Procurement, and Manufacturing

•**Impact:** Explainability, enhancing trust and interpretability.




CyclOps Multimodal Explainability Module

Instance: 219 of 389

Class: 1 of 5

Actual Value	machine_3
Predicted Value	machine_3
Prediction prob. (Selected class)	0.302
Prediction prob.	0.698

Feature Details

Feature	Value
type	3.0000
Color	4.0000
aluminium_type	4.0000
quality	1.0000
printing_speed	5665.3400

- Integrate sustainability and fairness metrics into the very first lines of your architecture.

Takeaways

Shift focus from optimizing performance to **optimizing for accountability**

Integrate **sustainability** and **fairness** into the very first lines of your architecture

AI and Knowledge-Based Systems are the **means**; human-centric wisdom and societal well-being are the **ends**

Work with us!
We are hiring

Julia PALMA
CeADAR
Programme Manager

julia.palma@ucd.ie

CeADAR
Ireland's Centre for AI

EDIH
European
Digital Innovation
Hubs Network

AIF IRL EuroHPC
EuroHPC AI Factory Antennas

CyclOps

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission as the granting authority. Neither the European Union nor the granting authority can be held responsible for them

Funded by
the European Union