# OpenEBench4AI: Integrating Execution, Metrics, and Resource Flexibility for Benchmarking AI models

**Vizcarra, Karla[1,2], Redondo, Anna[1,2], Fernández, José María[1,2], The OpenEBench Team, Gelpí, J.L.[1,2,3,] Capella-Gutierrez S.[1,2]**

[1] Spanish National Bioinformatics Institute (INB/ELIXIR-ES).ES.
[2] Barcelona Supercomputing Center (BSC).ES.
[3] Dept Biochemistry and Molecular Biomedicine, University of Barcelona (UB).ES.

**OpenEBench4AI** extends the **OpenEBench** platform with an **automated**, **scalable** framework for executing and benchmarking AI models. It supports diverse computing environments (CPU/GPU) and both public and private datasets, enabling **flexible, reproducible** evaluations.

## Key New features

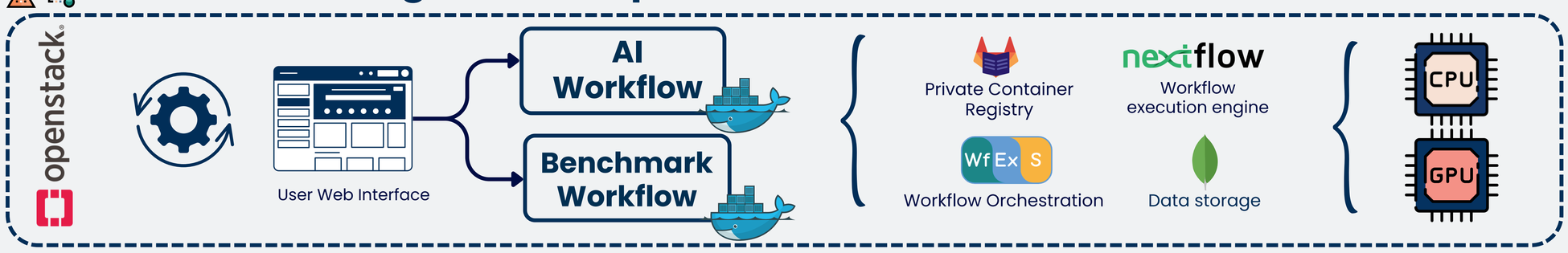
1. New modules for OpenEBench website
2. AI Benchmark Execution Platform
3. AI Submission Workflow Design
4. Integrated Metrics Implementation

Key features include a **user-friendly web** interface, **integrated metrics**, and a **streamlined experience** for participants and organizers. By combining execution, metrics, and resource management, OpenEBench4AI fosters a **transparent and efficient benchmarking ecosystem for AI research**.

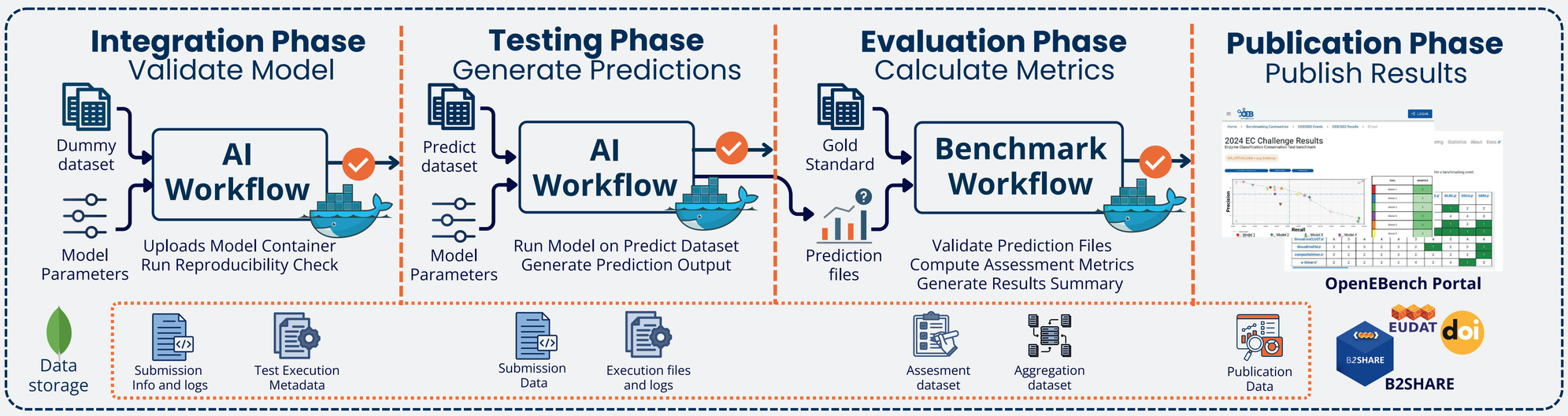
## 1 Web modules for AI

- Improved participant dashboard
- Participant enrollment for a challenge
- Participant submission
- Submission tracking status and configuration
- Challenge metrics configuration
- Workflow orchestration administration
- User administration
- AI plattform administration and configuration

## 2 AI Benchmarking Execution platform




## 3 AI submission Workflow design




## 4 Integrated metrics implementation

Metrics created and predefined following **FAIR principles**

- **Accesible** for selection during challenge configuration
- Automatically **configured in the** benchmark **workflow**
- **Ready** to be calculated without code manipulation
- **Integrated** with web widgets at publication stage

## Conclusions

- **OpenEBench4AI** enables **reproducible**, cross-platform **benchmarking of AI models** via a fully automated, container-based workflow.
- Simplifies challenge participation and **transparent performance** evaluation across diverse infrastructures.
- **Standardized metrics** and **robust execution** pipelines for FAIRer comparisons and more **trustworthy AI**.

## Next steps

- Expand support for **multimodal** datasets and challenges that require **additional features**.
- Explore new AI cases: **Generative** model-based submissions, **federated** and **encripted** data models.
- Enable automated integration of **custom metrics** for specific domains.

## References

OpenEBench new web design
OpenEBench Widgets
WfExS
AI Workflow

## Contact

**Karla Vizcarra Loayza**
karla.vizcarra@bsc.es
**INB/ELIXIR-ES Coordination Hub**
Barcelona Supercomputing Center (BSC)
**OpenEBench Team**
https://openebench.bsc.es/team

Spanish National Bioinformatics Institute
Barcelona Supercomputing Center
Centro Nacional de Supercomputación
AI4S BSC AI 4 Science Fellowships
This initiative is part of the talent attraction and retention programs within investment 4 of component 19 of the Recovery, Transformation, and Resilience Plan, C005/24-ED CV1
GENERACIÓN D
Financiado por la Unión Europea NextGenerationEU
red.es
Plan de Recuperación, Transformación y Resiliencia
ELIXIR ALL HANDS 2025