

Keynote 1: Software reproducibility for data processing and machine learning workflows in the era of AI

Tuesday 25 March 2025 09:10 (30 minutes)

In the era of AI, ensuring software reproducibility for data processing and machine learning workflows is more critical than ever, particularly in computationally intensive fields like bioinformatics. Reproducibility guarantees consistent outputs across different environments, which is essential for validating research findings and disseminating workflows widely. However, the complexity of managing software dependencies, especially as versions rapidly evolve, poses significant challenges.

This talk presents a principled approach to building reproducible analysis pipelines using tools like GNU Guix, demonstrated through the PiGx pipelines for RNA sequencing and other bioinformatics applications. We explore how these pipelines, by encapsulating dependencies and providing standardized outputs, serve as a model for reproducibility in AI-driven workflows. Additionally, we discuss the implications of integrating AI and machine learning in research, emphasizing the need for reproducible practices to maintain the integrity of AI applications across various scientific domains.

Presenter: Dr AKALIN, Altuna (Max-Delbrück-Center (MDC))

Session Classification: Opening