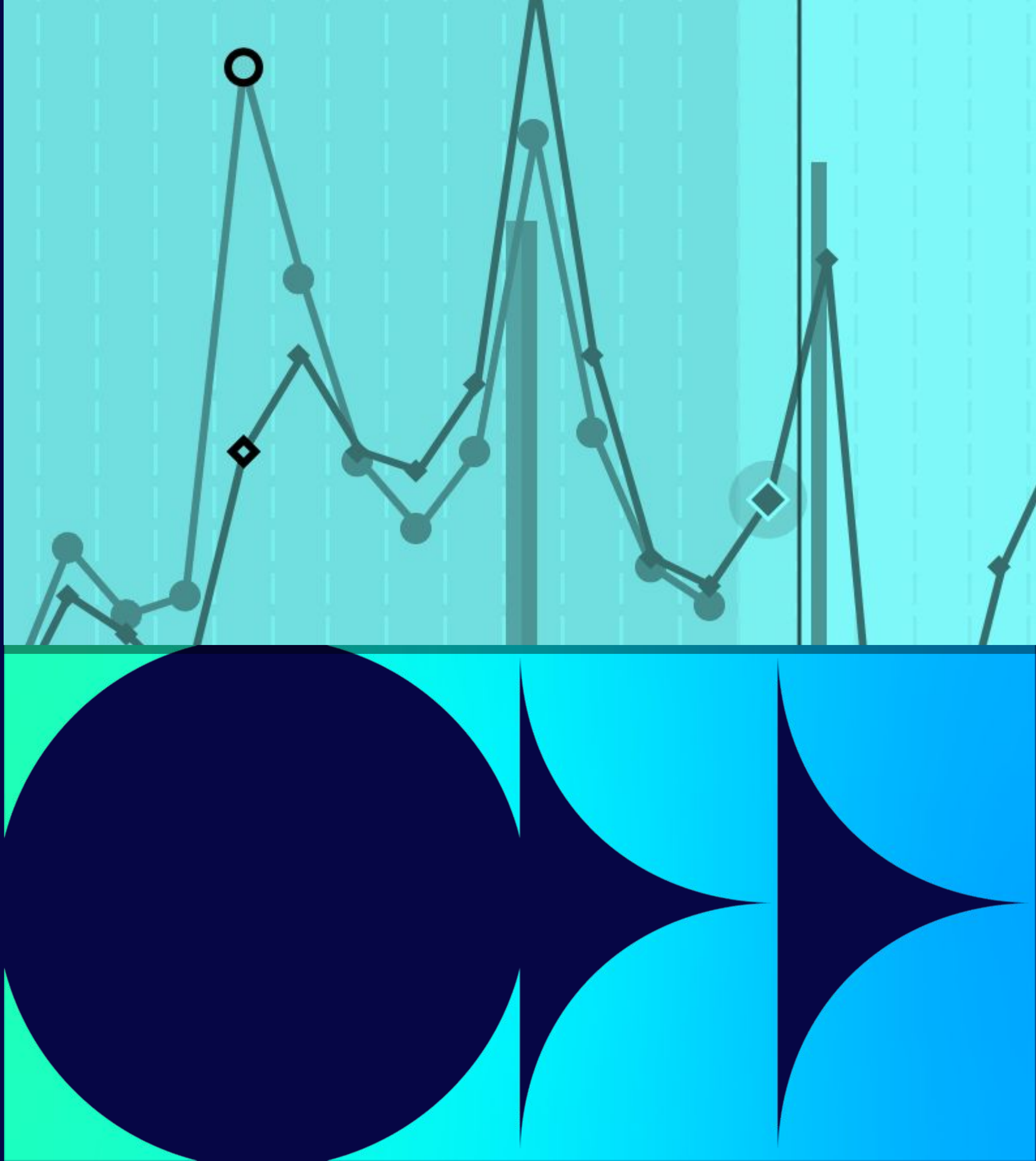
Embedding Responsible AI in Learning Analytics: From Ethical Principles to Institutional Practice

Fairness Innovation Challenge: Promoting Equity in HE



Knowledge Media institute, The Open University

26th November 2024, HESPA Virtual Conference

The Open University

KMi Knowledge Media Institute

UKRI Science and Technology Facilities Council

Engagement with the Community

Exploring Responsible AI principles within Learning Analytics solutions

Context

- **Aim:** gather information about
  - the current status of the adoption and implementation of Learning Analytics systems in UK Higher Education Institutions and,
  - the adoption of Responsible Tech/AI principles

- The data collected will help us understand
  - challenges institutions face
  - principles of interest for HEI when adopting Responsible Tech
  - inform and feedback the creation of a ***Responsible AI Framework for Learning Analytics in Higher Education***

- **HESPA**
  153 HE UK Institutions
- 10 min questionnaire

The Open University

Findings from Engagement with the Community

Questionnaire about the use of Responsible AI principles within LA solutions

Survey Results:

28 participants

3 nations represented:
England, Wales, Scotland

Roles represented*:

Strategy and planning	6
Technology design and implementation	5
Data analytics	3
Teaching and student success	3
Other	1

The Open University

** Answers were optional*

Poll No 1

Findings from Engagement with the Community

Definitions

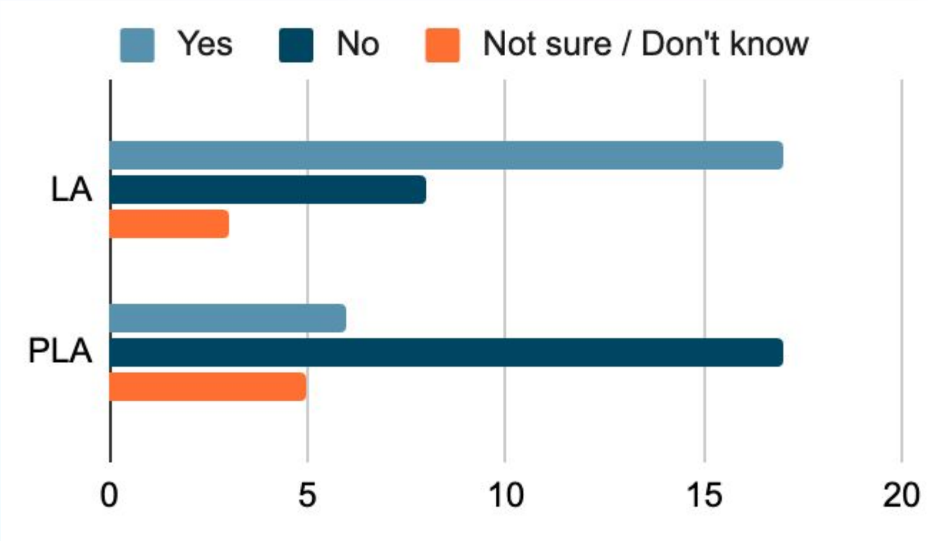
Learning Analytics (LA) [1]:

The measurement, collection, analysis, and reporting of data about learners and their contexts, for the purposes of understanding and optimising learning and the environments in which it occurs.

Predictive Learning Analytics (PLA):

A branch of Learning Analytics that aims to forecast what might happen in a learning context.

Does your organisation use of any type of LA or PLA solutions?



- *61% use LA or PLA solutions*
- *39% do not use any LA or PLA solution*

The Open University

[1] https://www.solaresearch.org/about/what-is-learning-analytics/

Poll No 2

Findings from Engagement with the Community

What type of LA/PLA solutions?

Third party tools, for example:

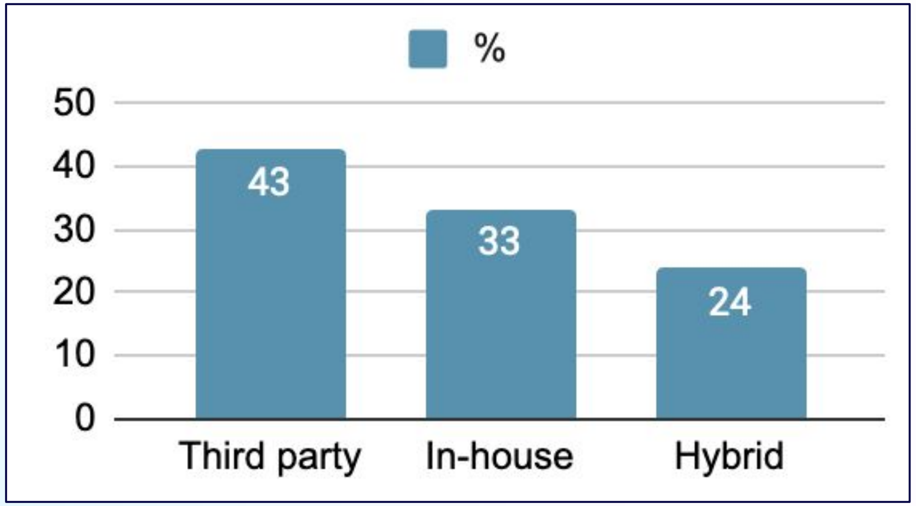
- Blackboard Analytics
- Civitas Learning
- IntelliBoard
- JISC Learning Analytics
- SEAtS One
- StREAM

Hybrid, for example:

- Power BI customisation
- AWS tools customisation

In-house

- Python scripts
- Dashboards



The Open University

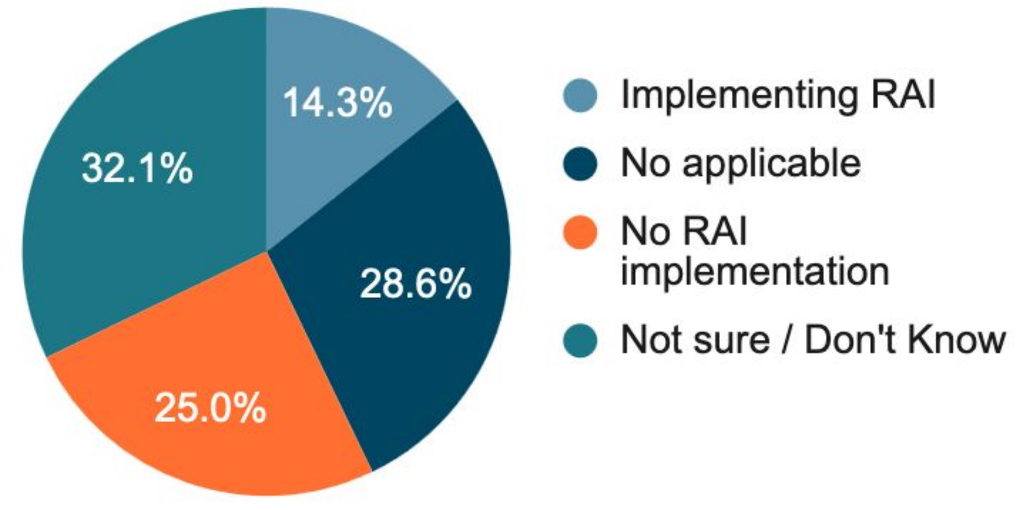
Poll No 3

Findings from Engagement with the Community

R-AI implementation

Does your organisation implement or consider Responsible AI principles within your LA and PLA solutions?

Responsible principles:
fairness & bias, transparency, security, privacy, safety, accountability, explainability



No R-AI implementation feedback:

"Early implementation of LA, not considered at this stage

"In consideration, but not formalised yet"

Third party tools seen as "Black-boxes" and difficult to understand

"Lack of policy"

"Need of right infrastructure/tools"

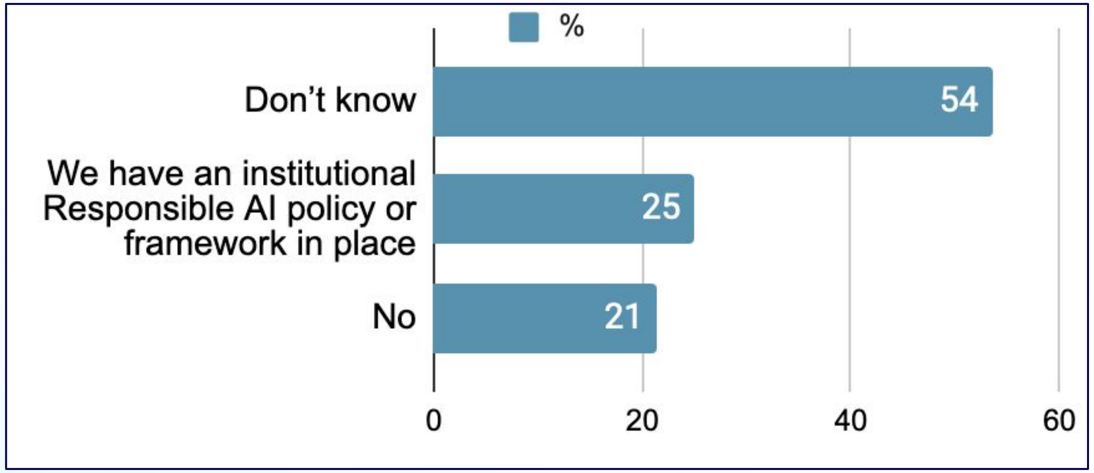
The Open University

Poll No 4

Findings from Engagement with the Community

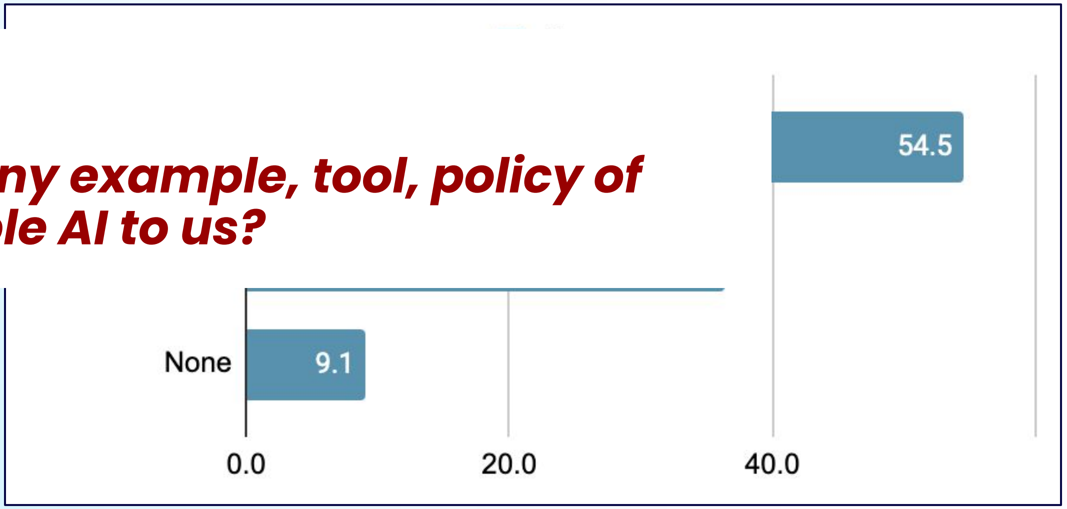
R-AI policy or framework

Does your organisation have an institutional Responsible AI policy or framework, external or internal?



Aspects considered in the Responsible AI policy/framework

- ***Social aspects:*** *in*
  aligning with stak
  the broader socie
  technology
- ***Technical aspect***
  and mitigating algorithmic biases, ensuring robust security measures, and safeguarding data integrity



Open Question:

Would you recommend any example, tool, policy of Responsible AI to us?

The Open University

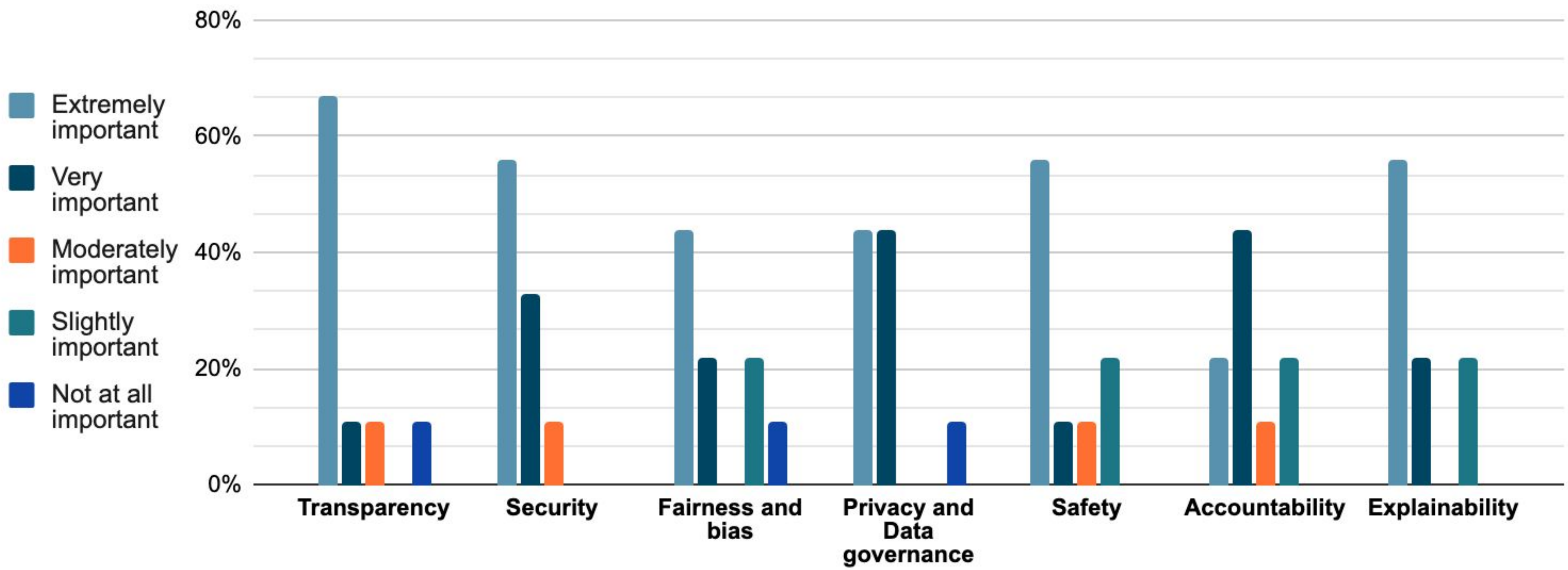
Findings from Engagement with the Community

Principles

How important are these principles/factors of Responsible AI within the LA and PLA solutions of your organisation?

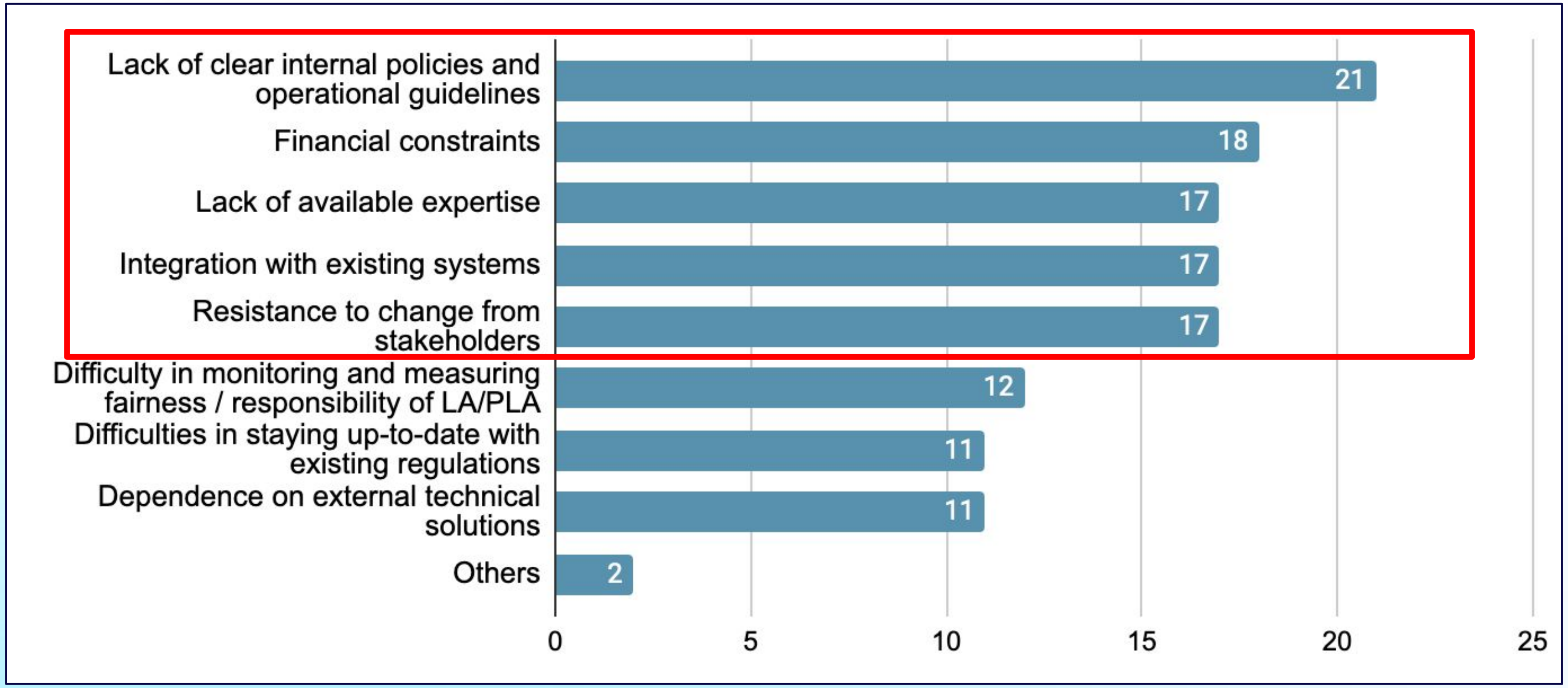
Responsible principles:

- *Fairness & bias*
- *Transparency*
- *Security*
- *Safety*
- *Accountability*
- *Explainability*
- *Privacy & data governance*



Findings from Engagement with the Community

What are the key challenges your organisation faces when implementing Responsible AI principles for LA/PLA solutions?



Others:

"...reliability, accuracy and usefulness have not been proven.."

"Concerns over the metrics used to generate insights..."

"...lack of clarity over when and how much is weighted to predictors."

The Open University

Open Question:

What do you think are the main challenges Higher Education Institutions in the UK face when implementing Responsible Tech principles?

Promoting Equity in HE

A specific focus on addressing bias and discrimination in AI-driven learning analytics systems.

Our project seeks to establish a framework of best practices, risks, and compliance with UK regulations.

The primary aim of this framework is to provide higher education institutions with *actionable guidance* on how to incorporate responsible AI principles effectively into their LA initiatives.

Latest publication

https://fairai4edtech.kmi.open.ac.uk/outcomes/

Towards an Operational Responsible AI Framework for Learning Analytics in Higher Education

ALBA MORALES TIRADO, PAUL MULHOLLAND, and MIRIAM FERNANDEZ, Knowledge Media Institute, The Open University, United Kingdom

Universities are increasingly adopting data-driven strategies to enhance student success, with AI applications like Learning Analytics (LA) and Predictive Learning Analytics (PLA) playing a key role in identifying at-risk students, personalising learning, supporting teachers, and guiding educational decision-making. However, concerns are rising about potential harms these systems may pose, such as algorithmic biases leading to unequal support for minority students. While many have explored the need for Responsible AI in LA, existing works often lack practical guidance for how institutions can operationalise these principles. In this paper, we propose a novel Responsible AI framework tailored specifically to LA in Higher Education (HE). We started by mapping 11 established Responsible AI frameworks, including those by leading tech companies, to the context of LA in HE. This led to the identification of seven key principles such as transparency, fairness, and accountability. We then conducted a systematic review of the literature to understand how these principles have been applied in practice. Drawing from these findings, we present a novel framework that offers practical guidance to HE institutions and is designed to evolve with community input, ensuring its relevance as LA systems continue to develop.

CCS Concepts: • **Applied computing** → **Learning management systems**.

Additional Key Words and Phrases: Responsible Artificial Intelligence, Learning Analytics, Higher Education

The Open University

Responsible AI Framework

For learning analytics in Higher Education Institutions

Framework development

- Built based on an exhaustive analysis of established Responsible AI frameworks (including leading technology companies) and mapped to the context of LA in HE
- Identified seven common principles including: *fairness and bias, transparency, privacy, accountability, explainability, safety and security*
- Conducted research to review how studies addressed these principles in practice within the LA domain
- Collected the solutions, challenges, and lessons learned from these studies
- Finally, we propose a framework aimed at guiding HE institutions in the responsible implementation of LA systems.

Responsible AI principles

- **Fairness and bias:** Ensuring that LA systems are free from biases that could disadvantage certain student groups.
- **Accountability:** Establishing clear lines of responsibility for the design, deployment, and outcomes of LA systems.
- **Transparency:** Providing clear and accessible information to all stakeholders.
- **Explainability:** Ensuring that predictions and decisions made can be understood by non-expert users
- **Security:** implementing robust technical safeguards to protect the integrity, confidentiality, and availability of data. Secure predictive models from manipulation or misuse.
- **Privacy:** Protecting the personal data of students and staff, ensuring that LA systems comply with legal and ethical standards around data privacy.
- **Safety:** Ensuring that LA systems are designed to minimise harm to staff and students

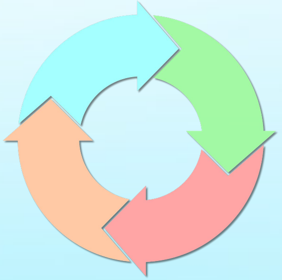
The Open University

Responsible AI Framework

For learning analytics in Higher Education Institutions

Structure

Recognising that institutions are at various stages of their LA adoption, we have structured the framework to follow the stages of the software development lifecycle



Framework proposal V0.1

RESPONSIBLE AI FRAMEWORK

Principles/ Stages	Requirements & Data Collection	Design	Development	Testing	Release & Monitoring
Accountability					
Explainability					
Fairness & bias					
Safety					
Security					
Transparency					
Privacy					

Actionable guidance

The Open University

Responsible AI Framework

For learning analytics in Higher Education Institutions

Fairness & Bias

Tools & examples

AI FAIRNESS-> https://github.com/Trusted-AI/AIF360

- AI Fairness 360 - Python: AIF360 is an extensible open-source library containing techniques developed by the research community to help detect and mitigate bias in machine learning models throughout the AI application lifecycle. This document will provide an overview of its features and conventions for users of the toolkit.

Fairlearn (python) -> https://fairlearn.org/

- **Fairlearn** is an open-source, community-driven project to help data scientists improve fairness of AI systems.

WhatIfTool -> https://pair-code.github.io/what-if-tool/

- Using WIT, you can test performance in hypothetical situations, analyze the importance of different data features, and visualize model behavior across multiple models and subsets of input data, and for different ML fairness metrics.

Using AI Fairness360 & Fairlearn

- Deng, W. H., Nagireddy, M., Lee, M. S. A., Singh, J., Wu, Z. S., Holstein, K. and Zhu, H. (2022) 'Exploring How Machine Learning Practitioners (Try To) Use Fairness Toolkits', *2022 ACM Conference on Fairness, Accountability, and Transparency*. DOI: 10.1145/3531146.3533113 .

Common accuracy metrics

- false positive rate (FPR), false negative rate (FNR), true positive rate (TPR), true negative rate (TNR), Positive predictive value (PPV)
  - Examples and definitions: Agarwal, S. and Mishra, S. (2021) Responsible AI: Implementing Ethical and Unbiased Algorithms. DOI: 10.1007/978-3-030-76860-7

Fairness metrics

- Equal Opportunity, Predictive Equality, Equalized Odds, Predictive Parity, Demographic Parity, Average Odds Difference
  - Examples and definitions: Agarwal, S. and Mishra, S. (2021) Responsible AI: Implementing Ethical and Unbiased Algorithms. DOI: 10.1007/978-3-030-76860-7

Bias audit toolkit

- **AEQUITAS.** University of Chicago's open source bias audit toolkit for machine learning developers. https://www.datasciencepublicpolicy.org/our-work/tools-guides/aequitas/

Software for detecting algorithmic discrimination

- Repositories: https://github.com/megantosh/fairness_measures_code/

Ethics

- An ethics checklist for data scientists. https://deon.drivendata.org/

Responsible AI Framework

Open conversation

While many works have explored the need for Responsible AI principles in LA, existing works often lack practical guidance for how institutions can operationalise these principles

1. What are the **barriers** to moving from guidance to practical implementation?
2. What **best practices** exist for implementing these principles?
3. Which **new types of support or guidance** do organisations need for implementation?
4. What are the **implications in terms of policy** and **regulations**?

Responsible principles:
fairness & bias, transparency, security, privacy, safety, accountability, explainability

The Open University

Thank you!

Meet the speakers:

Professor Miriam Fernandez

Dr Paul Mulholland

Dr Alba Morales Tirado

For more information visit:

https://fairai4edtech.kmi.open.ac.uk

The Open University

KMi
Knowledge Media Institute

UKRI
Science and Technology Facilities Council

Thank You!

To find out more about HESPA, scan the QR code below: