Ethics and Accessibility in Artificial Intelligence driven Language Assessment

Professor Barry O'Sullivan OBE

British Council
English Language Research

The English Language Research & Impact team drive innovation through impactful research in product development and education reform.

Team members provide technical expertise, advice and guidance on English teaching, learning and assessment worldwide.

Selected Areas of Expertise

Assessment Design & Development		Ethics in Assessment & Learning
Language Teaching & Learning	Validity	Localisation & Personalisation
Language Assessment Policy	Validation	Equality, Diversity & Inclusion
Building Learning Systems	AI Solutions	Data & Measurement
Monitoring & Evaluation	CEFR Application	Washback & Impact
Application of EdTech	NLP in language assessment	Future of English

Setting the Context

An Ethical Approach

Making it Happen

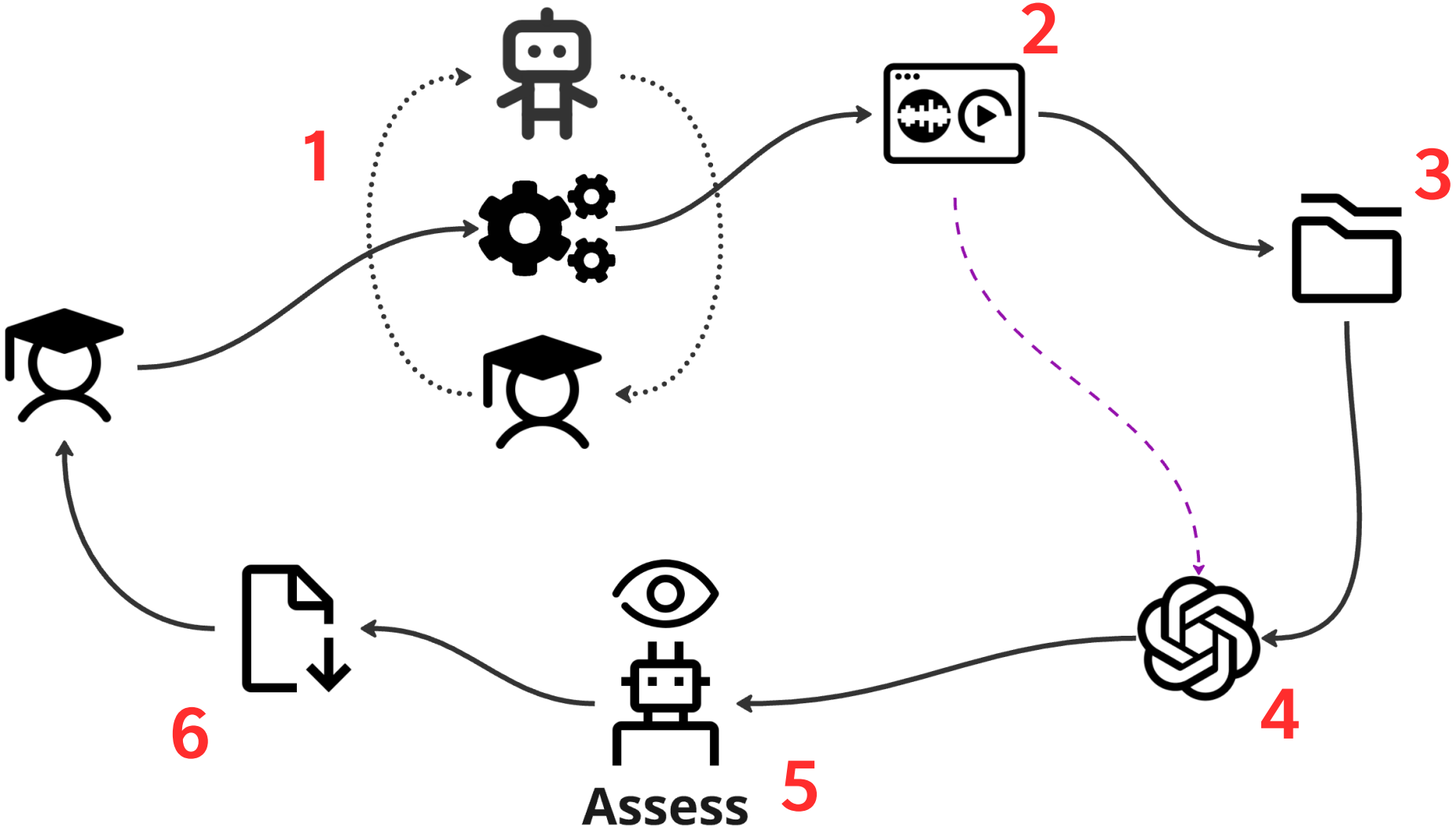
Accessibility

BRITISH COUNCIL

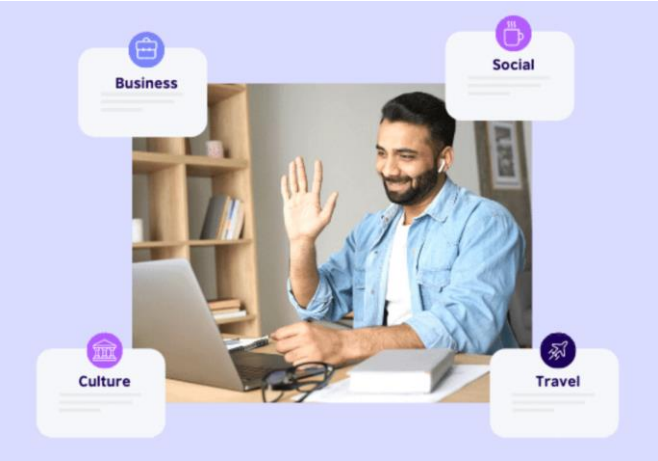
Setting the context

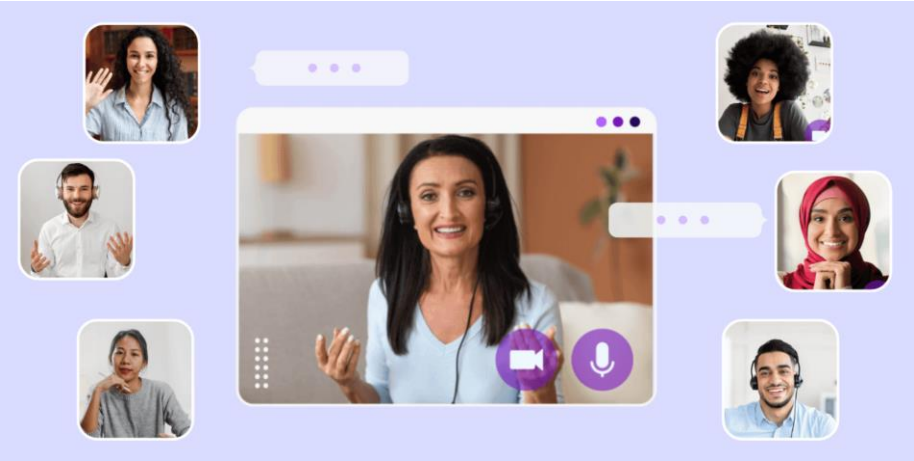
integrating assessment into a digital learning system

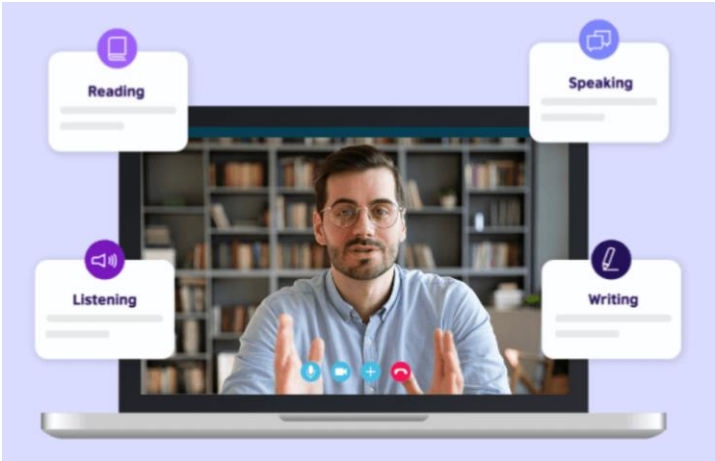
Developmental Assessment



English Online







Available 24/7

Themes adapted to your needs

British Council expert teachers

AiBC

Formative

FRED

Summative

BRITISH COUNCIL

An Ethical Approach

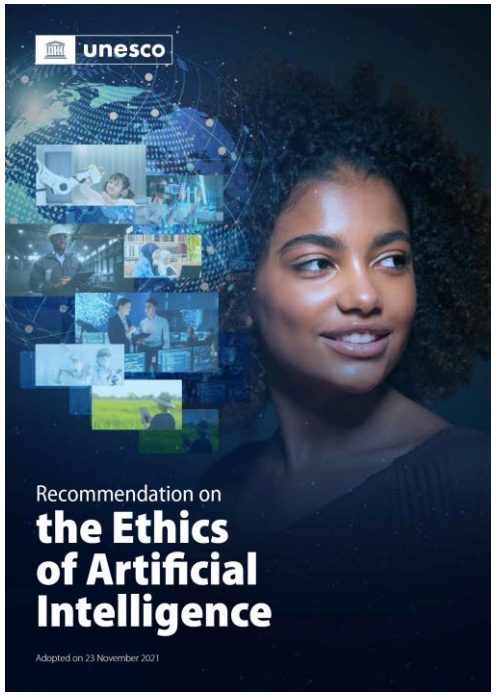
overview of the British Council's position

Guidelines, frameworks and legislation

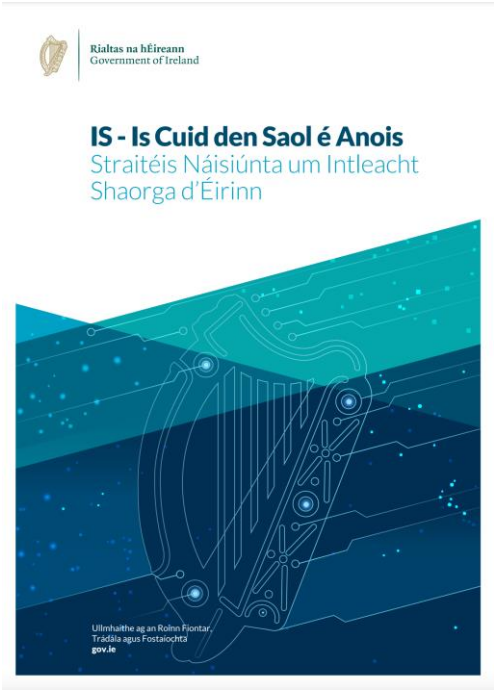
GDPR.EU

EU Artificial Intelligence Act

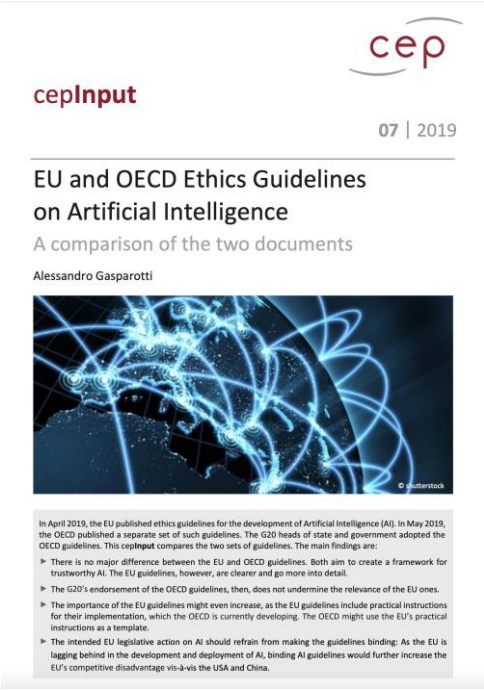
EU



UNESCO



National AI policies and regulations



Recommendations by organisations

Ethics by Design

- Six general ethical principles for AI systems:

Respect for Human Agency

Privacy and Data Governance

Fairness

Individual, Social and Environmental Well-being

Transparency

Accountability and Oversight

1. Respect for Human Agency – Three principles

Autonomy

Allow people to think for themselves, decide for themselves what is right and wrong, and choose how they want to live their lives.

Dignity

Human beings possess an intrinsic worth and should not be instrumentalized, objectified or dehumanized, but treated with respect at all times.

Freedom

People should not be constrained and be guaranteed freedom of movement, freedom of speech, freedom of access to information, and freedom of assembly.

2. Privacy and Data Governance

- AI systems must process personal data in a **lawful, fair and transparent** manner.
- Ensure data **minimisation** and data **protection**.
- Set **security measures** and safeguard the rights and freedoms of data subjects.
- Data should be acquired, stored and used in a manner which can be **audited by humans**.

3. Fairness

- All people are entitled to the same fundamental rights and opportunities.
- No discrimination.
- Identify and avoid algorithmic bias.
- Ensure universal accessibility (to the extent possible).
- Identify and mitigate negative social impacts.

4. Individual, Social and Environmental Well-being

- AI systems must not unduly or unfairly reduce the psychological and emotional well-being of its end-users and stakeholders.
- Be mindful of principles of environmental sustainability.
- Mitigate any negative impact on communication, social interaction, information, democratic processes or social relations.
- AI and robotics systems must not reduce safety in the workplace

5. Transparency

- The purpose, inputs, and operations of AI programs must be **knowable and understandable** to its stakeholders.
- It must be made clear to end-users that they are **interacting with an AI system**.
- **Traceability** during the system's lifecycle.
- **Explainability**, i.e. how decisions were made by the system.
- **IP rights, confidentiality or trade secrets** claims cannot prevent transparency but can be confidentially shared.

6. Accountability and Oversight

- The actors involved in the development or operation of systems must **take responsibility** for the way in which they function and for the resulting consequences.
- They must be able to **understand, supervise and control** the design and operation of the AI system.
- Document how **socially undesirable effects** of the system will be detected, stopped, and prevented from reoccurring.
- Specify how stakeholders will be able to **report complaints, ethical concerns, or adverse events** and how these will be dealt with

Making it happen

the british council – ucl project

The British Council – UCL collaboration

British Council AiBC Project

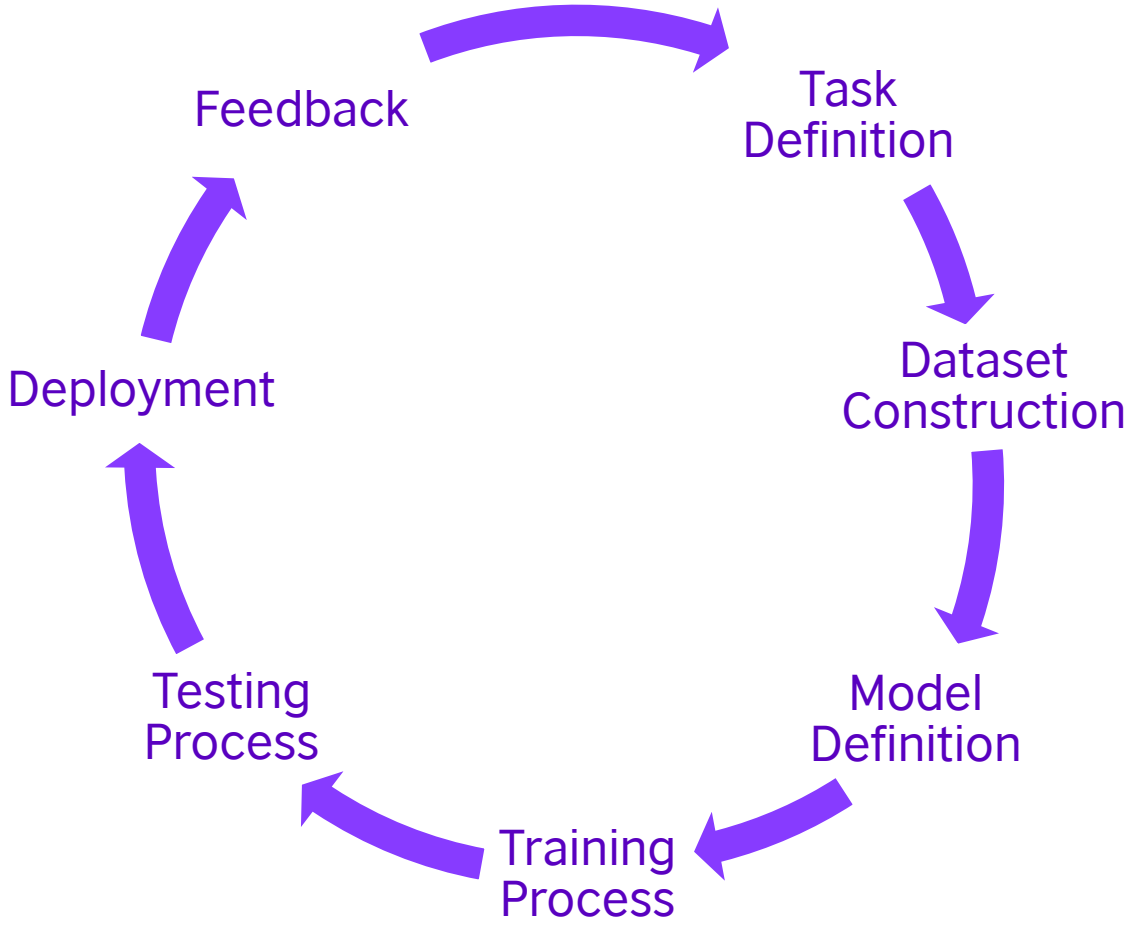
- An ambitious and innovative AI-powered product.
- Combines different kinds of AI technology, such as ASR, GenAI, speech synthesis, NLU, etc.
- Early stages.
- Keen to collaborate with other organisations.

UCL Knowledge Lab

- Research laboratory focused on the interactions between society and technology.
- Expertise in AI and education.
- Experience in supporting companies during system development.
- Developed an ethics framework.

Audit-D

- Audit tool (questionnaire) by Porayska-Pomsta, Holmes and Nemorin (2023).
- Iterative audit during seven stages of system development.



Porayska-Pomsta, K, Holmes W, Nemorin S. (2023). The ethics of AI in education. In DuBoulay, B., Mitrovic, A., Yacef, K. (Eds.), *Handbook of Artificial Intelligence in Education*, (pp.571-604), edward elgar publishing

WHO?	WHAT?	WHY?	HOW?	BIASES?

TASK DEFINITION (14)

DATASET CONSTRUCTION (13)

MODEL DEFINITION (14)

TRAINING PROCESS (11)

TESTING PROCESS (11)

DEPLOYMENT (11)

FEEDBACK (11)

	TASK DEFINITION	DATASET CONSTRUCTION	MODEL DEFINITION
WHAT?	What is the product?	What type of data is necessary to enable the interaction with task and system?	What is being modelled?
	Does the product require any pre-requisite knowledge?	What type of data is sufficient to enable the interaction with task and system?	To what extent does your choice of objective function correspond to the task's/system's goals and desired outcomes?
	How is it to be used? (what are the decisions to be made about the user?)		Are there alternatives to the objective function chosen?
WHO?	Who is the product for?	Who will provide the data?	Who will provide the data?
	Will all learners/users have access to knowledge needed?	In what contexts will they provide the data?	In what contexts will they provide the data?
	Will all learners/users feel comfortable performing the task?	Who will the data be representative of?	Who will the data be representative of?
WHY?	Why this task (and not another)?	Why are the specific types of data needed?	What do you want to infer and how/what will you use it for?
	What is the purpose of the task in terms of outcomes?	What are your assumptions about the data and context(s) in which they were collected?	What is the evidence for your assumptions and choices regarding your objective function/predictor variables/algorithmic choices
HOW?	What is the implementation scenario?	If you have the data available: how old is it?	What modelling techniques will be used?
	How will you develop it and test it?	If you have the data available: where does it come from	What are the assumptions needed to use these techniques?
	Does the learner/user have a choice over what task they do?	If you have the data available: has it been audited for any prejudices and stereotypes?	What are the known issues with these techniques?
BIASES?	Is the product culture, gender, age appropriate?	Is the data representative of all potential users' backgrounds, needs, learning trajectories? If so/not - in what way?	Is the objective function congruent with the task objectives?
	Is the intervention safe for everyone?	Is the data collected congruent with the intended pedagogies/learning support?	Are the objective function and the corresponding outcome measures fair in terms of the process or outcomes for all intended users?
	Are there any potential adverse long-term consequences for any users?	Does the data collected allow you to encode pedagogical supports that are appropriate for all/some of your target users and their contexts?	Is the model baseline data representative of all intended users?

	TRAINING PROCESS	TESTING PROCESS	DEPLOYMENT	FEEDBACK
WHAT?	What baseline/training data will you use?	What are you testing for (e.g., system performance, learners' outcomes, user experience, …)?	In what context or contexts is your technology deployed?	What will be captured as user feedback, e.g., (dis-)likes, complaints; detailed reports of use (including what worked and what didn't)?;
	How voluminous is the dataset; is it sufficient for model learning?	What are your assumptions about what constitutes a good outcome of the test?	What are your assumptions about the deployment context or contexts for your system? Are these assumptions valid under all possible known circumstances?	What data will be collected?
	What pedagogical and learning approaches will your model support?	What are your assumptions about the context in which the system will be used? Can you guarantee that these assumptions will hold for all possible contexts of use?	What is the level of your system's transparency about what it can and cannot do, for whose use it is aimed, and in what contexts?	Will all users' feedback have the same weight in impacting further development, changes, etc.?
WHO?	Who will provide the data?	Does your test data match the benchmark data?	Has your data (learning, test, or benchmark) been collected in the deployment context(s)? Is the data representative of the deployment context	Whose feedback will be captured (learners, educators, employers, …)?
	Is the training dataset representative enough of the target population to generalise across all learners and contexts?	Is your test data representative of the deployment contexts (in terms of target users, intended vs actual environment, etc.)?	Does the design context match the deployment context?	Who will have access to feedback?
WHY?	In what way do you consider the baseline data necessary and sufficient to achieve your system's/task's goals?	In what way is the test data an adequate basis for the evaluation of your system?	Do your system's models encode the contextual information needed for their performance efficacy and relevance to those contexts?	Why will the feedback be collected (what will you want to learn from it)?
	In what way are the pedagogical support capabilities of your system appropriate for the task and target population?	In what way does the system stand up to scrutiny with respect to what the system promises it can do?	Have you built in sufficient flexibility into your system to allow for changes in the way that the system adapts to the user/can be personalised to their needs?	Why will the specific forms of feedback be collected
HOW?	How will the training data be generated?	What performance measures are you using to evaluate your system?	How have you audited for shifts in user populations and deployment contexts?	How user's interaction with the system will be monitored?
	Have you checked for proportionality in your training data vis a vis target population?	What challenges/known weaknesses do your chosen performance measures have (what can they/cannot capture?)?	How have you assessed whether any issues, such as mismatches between the intended and the deployment context are a one-off issue or indicative of a more persistent problem?	How will feedback from different users be differentiated, if at all?
BIASES?	Does your data represent all or only some of your target users?	Is your test data representative of the users and contexts for which the system is designed?	Have you considered external audits of your system's efficacy, fairness, ethical quality?	Will the feedback be representative of all intended users?
	Are all inferences and predictions based on correct, relevant and complete information?	Is your model(s)'s performance an outcome of overfitting?	Are your assumptions about the deployment context generalisable?	Are all the data collected for feedback necessary?

Recommendations and actions

- UCL provided **recommendations** for each stage of the development pipeline, based on our answers to questions.
- UCL raised **“red flags”** in some cases, which required further attention by the GEEP Project team (self-reflection and actions).
- The outcome of the audit had a **direct impact on design decisions**, e.g. task design, choice of services, data handling, etc.

Outcomes and lessons learnt

- The ethics audit was a **thorough and time-consuming** process.
- It exposed **differences** in terminology and interpretation.
- It helped **identify and reflect on issues** that would have otherwise been missed.
- It gave the British Council a tool to perform **continuous ethics assessment**.
- It provided UCL with a **case study** for their methodology.

BRITISH COUNCIL

Accessibility

one implication

Aspects	Progression		
	Acquire	Deepen	Create
Human-centred Mindset	Human agency	Human accountability	AI social responsibility
Ethics of AI	Ethical principles	Safe and responsible use	Co-creating AI ethical rules
AI Foundations & Applications	Basic AI techniques and applications	Application skills	Creating with AI
AI pedagogy	AI-assisted teaching	AI-pedagogy integration	AI-enhanced pedagogical transformation
AI for professional development	AI enabling lifelong professional learning	AI to enhance organizational learning	AI to support professional transformation

AI competency framework for teachers (AI CFT)

Guiding principles AI CFT

1. Human-AI social contract towards inclusive digital futures
2. A human-centred approach
3. Protecting teachers right and dynamically defining teachers' roles
4. Promoting trustworthy and climate-friendly AI for education
5. Ensuring applicability for all teachers and reflecting digital evolution
6. Teacher development as lifelong learning

AI competency framework for school students (AI CFS)

Guiding principles AI CFS

1. Fostering critical thinking on the proportionality of AI for the real-world challenges
2. Prioritizing competencies to make human-centred collaboration with AI
3. Steering the design and use of more climate friendly AI
4. Facilitating transferable AI foundation for lifelong learning
5. Promoting inclusivity in AI competency development

Competency Aspects	Progression Levels		
	Understand	Apply	Create
Human-centred mindset	Human Agency	Human Accountability	AI Society Citizenship
Ethics of AI	Embodied Ethics	Safe and Responsible Use	Ethics by Design
AI techniques and applications	AI Foundations	Application Skills	Creating AI Tools
AI system design	Problem Scoping	Architecture Design	Iteration and Feedback Loops

unesco

ENNHRI European Network of National Human Rights Institutions

AI can provide tailored educational materials and support for students with learning disabilities. Adaptive learning platforms powered by AI can adjust the pace and content to meet individual needs and foster a more inclusive learning environment.

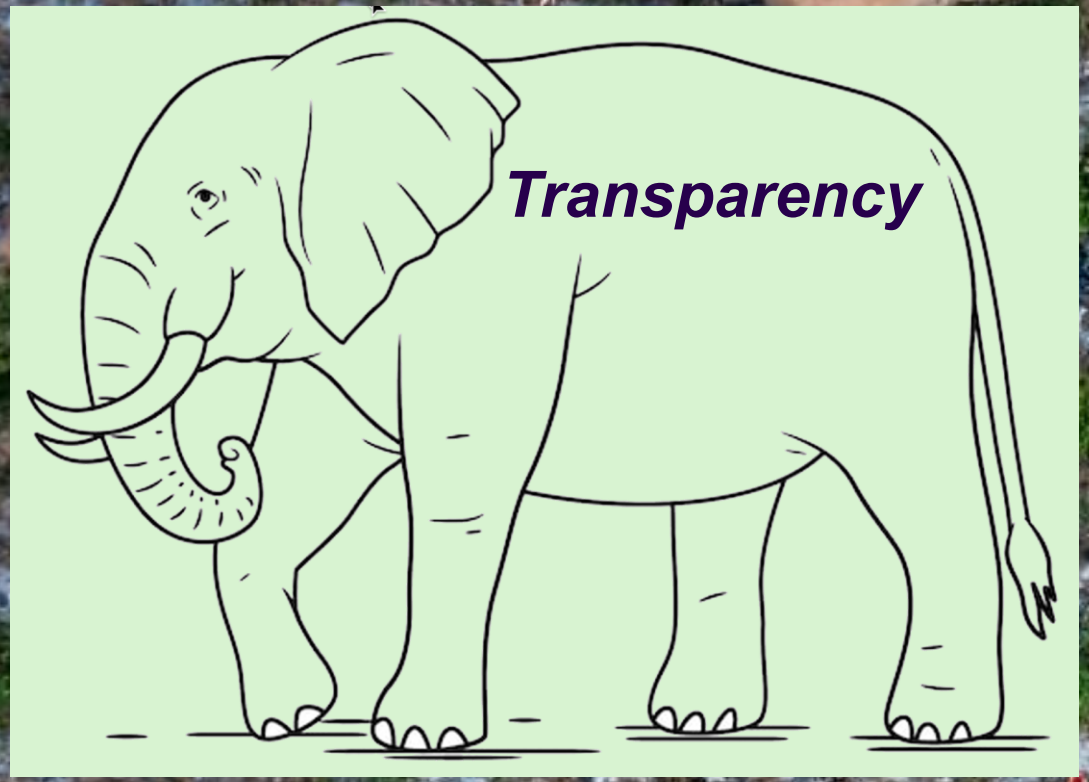
(Turin, 2023)

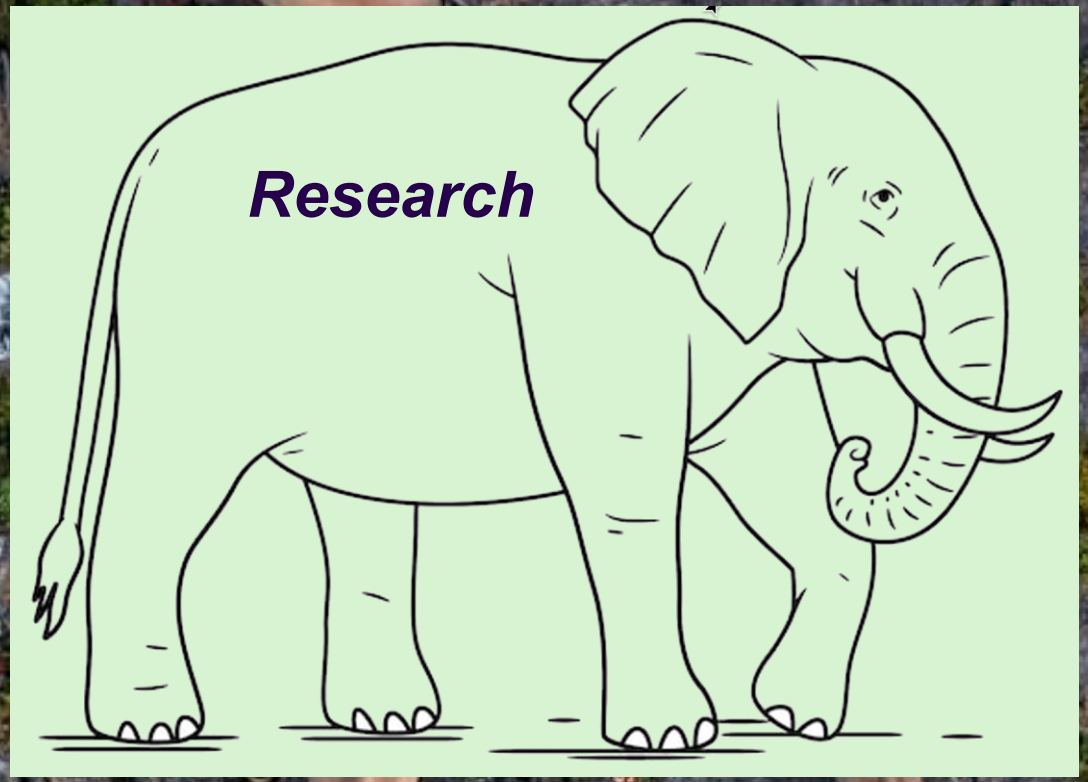
Turin, J. (2023). *Artificial intelligence and its impact on the human rights of persons with disabilities*. Guest blog on the ENNHRI website. https://ennhri.org/news-and-blog/artificial-intelligence-and-its-impact-on-the-human-rights-of-persons-with disabilities/#:~:text=AI%20holds%20significant%20potential%20to,avenues%20for%20communication%20and%20interaction.

Accessible Exam Formats	• Text-to-speech (TTS) • Speech-to-text (STT)
Adaptive Learning Platforms	• Personalised Learning • Real-time Feedback
Exam Proctoring and Monitoring	• Remote Proctoring • Behaviour Analysis
Customised Exam Timing	• Flexible timing
Assistive Technologies	• Braille Translation • Visual and Hearing Aids
Emotional and Mental Support	• AI Chatbots • Cognitive Behavioural Therapy (CBT) Tools
Language and Communication Support	• Language Translation (multilingual assessment) • Augmentative and Alternative Communication (AAC)
Data Analysis for Accommodations	• Needs Assessment • Performance Tracking

adaptive learning	combines computer algorithms and AI to curate and the customise resources and assessment tasks
facial expression analysis	Allows for accurate interpretation of candidates' emotions and engagement levels – potentially triggering support
chatbots	Provide personalised guidance through the test while offering emotional and subject-specific support where appropriate
communication aides	Support those with difficulties speaking or understanding what is being said (autistic, or who have cerebral palsy or learning disabilities)
mastery learning systems	Provide candidates with as much time and support as they need to demonstrate their understanding/ability
intelligent tutoring systems	A system that imitates human interlocutors and provides immediate and customised feedback or guidance to learners
interactive robots	Range from small student programmable robots to larger robots that are designed to interact with students in a more social way

Bhatti, I., Mohi-U-din, S. F., Hayat, Y., & Tariq, M. (2024). Artificial Intelligence Applications for Students with Learning Disabilities: A Systematic Review. *European Journal of Science, Innovation and Technology*, *4*(2), 40-56. Retrieved from https://ejsit-journal.com/index.php/ejsit/article/view/397


Transparency


Research


Gracias
Go raibh míle maith agaibh
Thank You