Nottingham Trent University

Demystifying AI—Risks and Opportunities that Support Education and Assessment Resources for Persons with Neurodevelopmental and Neurocognitive Impairments.

AI powered inclusive education digital transformation in Vietnam

David J Brown, Professor of Interactive Systems for Social Inclusion, Director of the Computing and Informatics Research Centre.

Nottingham Trent University, UK.

Agenda

1. Welcome, framing, and "what AI is (and isn't)".
2. Explainability: Inherently interpretable vs post hoc XAI.
3. Inclusive, ethical, bias-aware AI: what makes it hard (and doable)
4. Research examples: XAI + VR + brain informatics + biofeedback + Self Regulation
5. "Neurobridge" – micro-credentials for teachers using AI to support neurodiverse students
6. Wrap-up, take-home messages

Welcome, framing, and “what AI is (and isn’t)”

- **AI Definition and Capabilities**
- **Many advanced AI models are difficult to understand even by the experts who created them**
- **Applications of AI SEND Education**
- **Common Misconceptions:** AI is not Magic, infallible, or inherently neutral.
- **Ethical and Practical Limitations:** Reliance on opaque algorithms without robust explanation mechanisms can erode trust and limit accountability.

“an algorithm does not have to make the same choice as a human would make, but a human should at least be able to understand the process by which the decision is made.”

Explainability: Inherently interpretable vs post hoc XAI:

Explainability is the cornerstone of trustworthy AI, particularly in special and inclusive education, and clinical settings where decisions impact health, autonomy, and rights.

Inherently Explainable Models - Models like decision trees offer transparency by design, aiding teachers' understanding of AI decisions.

Post Hoc Explanation Techniques - Feature attribution (e.g., PHASE IV AI) and saliency maps interpret complex AI models but may have fidelity uncertainties.

Ethical Necessity in SEND & Neuropsychology - Explainability supports informed consent and shared decision-making for individuals with cognitive impairments.

Explainability Enables Trust and Contestability - Clear AI explanations empower users to trust and challenge AI outputs when needed.

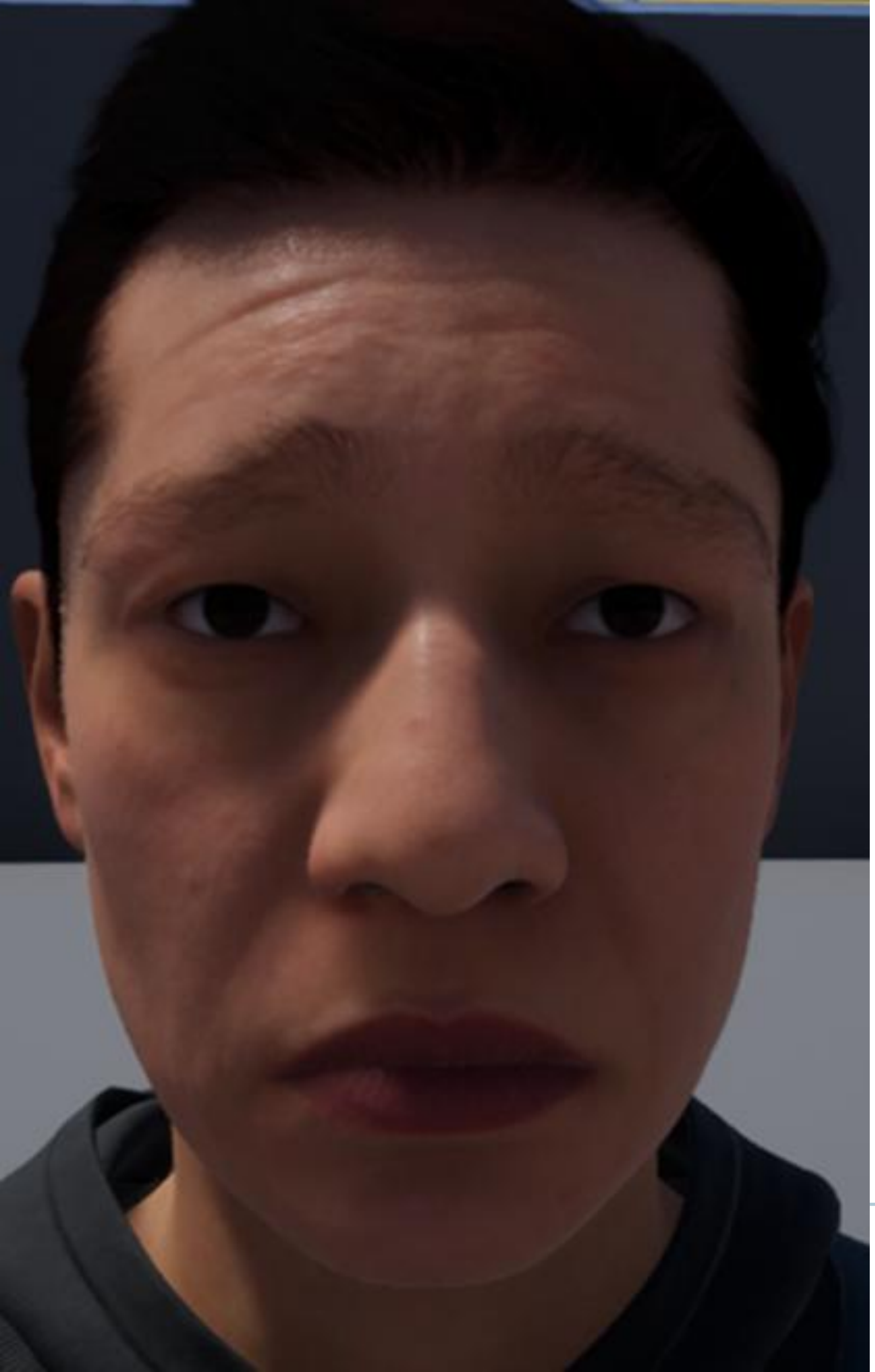
Inclusive, ethical, bias-aware AI: what makes it hard (and doable)for our neurodiverse learners…

Challenges

- C1. Datasets often lack diversity.
- C2. Informed consent poses issues for users with cognitive impairments.
- C3. Explanations can be extremely hard to understand for non experts.
- C4. Achieving XAI in multimodal AI remains challenging.

Mitigation

- M1. Training on diverse, disability-specific data makes model behaviour more predictable and less prone to opaque errors
- M2. Working with experts and exploration of how AI systems that learn over time can adapt their behaviour when consent changes.
- M3. Co-design embeds user-appropriate explanations.
- M4. Focus on mapping how different inputs interact, assigning importance to different data sources.



Research Example 1: Engagement/meltdown detection and privacy-preserving ML in autism.

Prevalence of Autism - More than 1.1% of the UK population may fall within the autism spectrum.

There has been a sharp increase in diagnoses over the past decade.

We have developed AIEd Tools that detect affective states relevant to learning.

Meltdown - An intense response to an overwhelming situation, when someone is completely overwhelmed by their current situation and temporarily loses control of their behaviour.

"Many autistic people have meltdowns – how to anticipate them, identify their causes and minimise frequency" (autism.org.uk).

Many autistic people will show signs of distress before having a meltdown, which is sometimes referred to as the "rumble stage".

Strategies at this stage can help - such as taking deep breaths, listening to calming music, going for a walk, exercise, quiet space, stress ball…

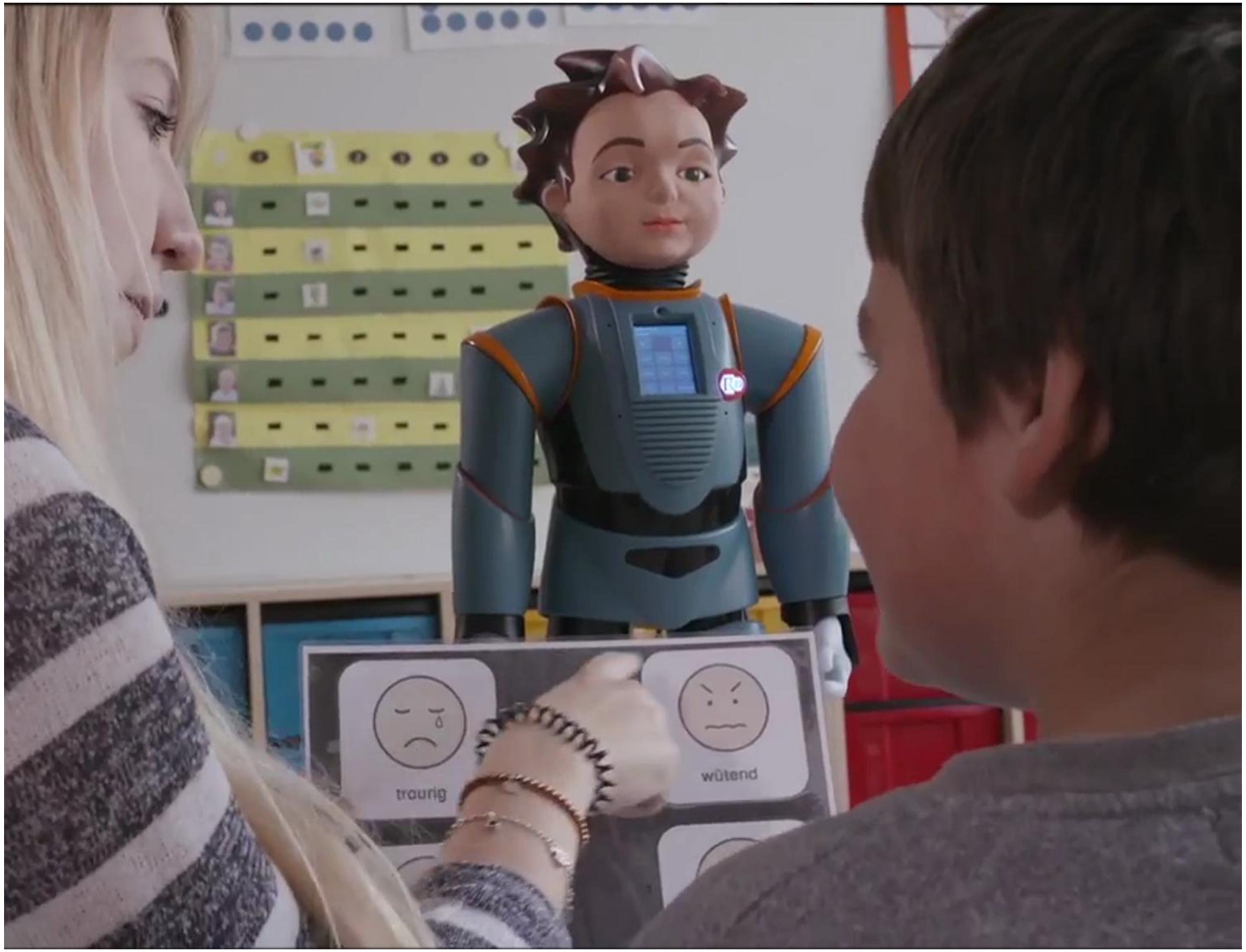
Paucity of Training Data

"As machine learning continues to advance, it is important to prioritize the quality of the training data to ensure the development of ethical and effective AI systems" (https://www.linkedin.com/pulse/importance-good-training-data-machine-learning-datacrunch/).

Good training data ensures that the algorithm is accurate, reliable and unbiased.

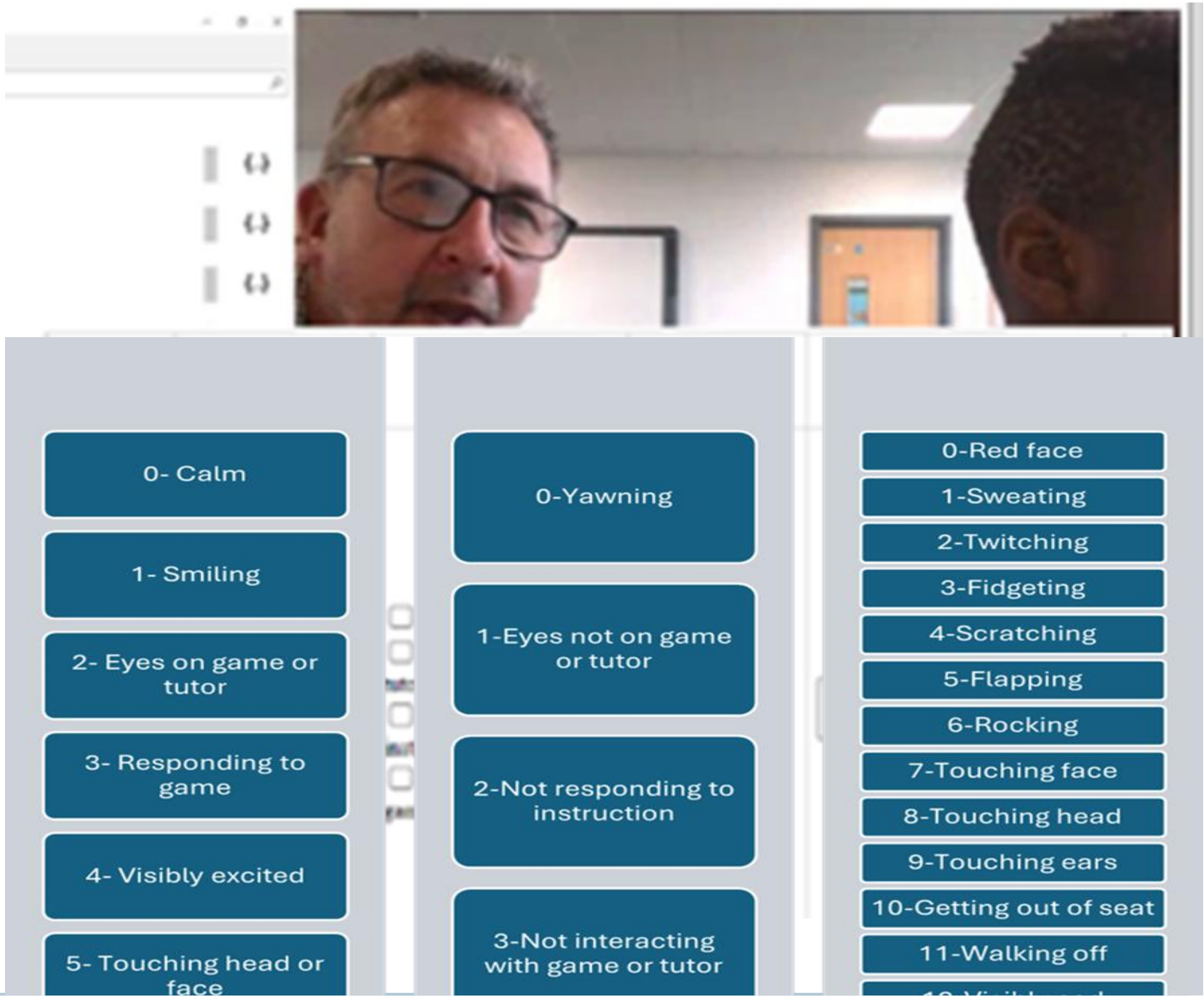
DEENIGMA and the Meltdown Crisis datasets – obtaining access remains difficult.

Very problematic to deliberately stress young people with Autism to develop training samples, so instead we use games acting as CPTs to develop the labels for both emotional states related to learning and psychological arousal.

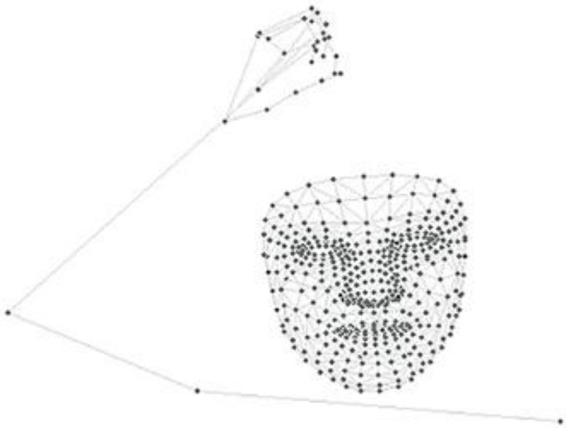


Data Collection, Labelling, Segmentation

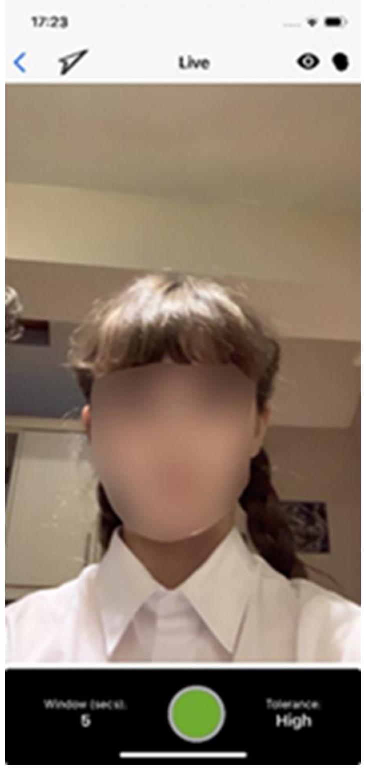
- We worked with 39 different students across the schools in our UK network and collected data in 74 separate sessions.
- A labelling checklist was co-created by educational psychologists to act as labels for the video samples.
- The videos produced were then labelled with the appropriate behavioural states using the observational behavioural checklist.
- Checklist recreated within the VIA video annotation software.
- Video segmentation was performed according to the labels assigned.
- Mediapipe (using both Face and Pose Landmarkers) was used to detect facial and posture landmarks and extract corresponding 3D cloud datapoints.



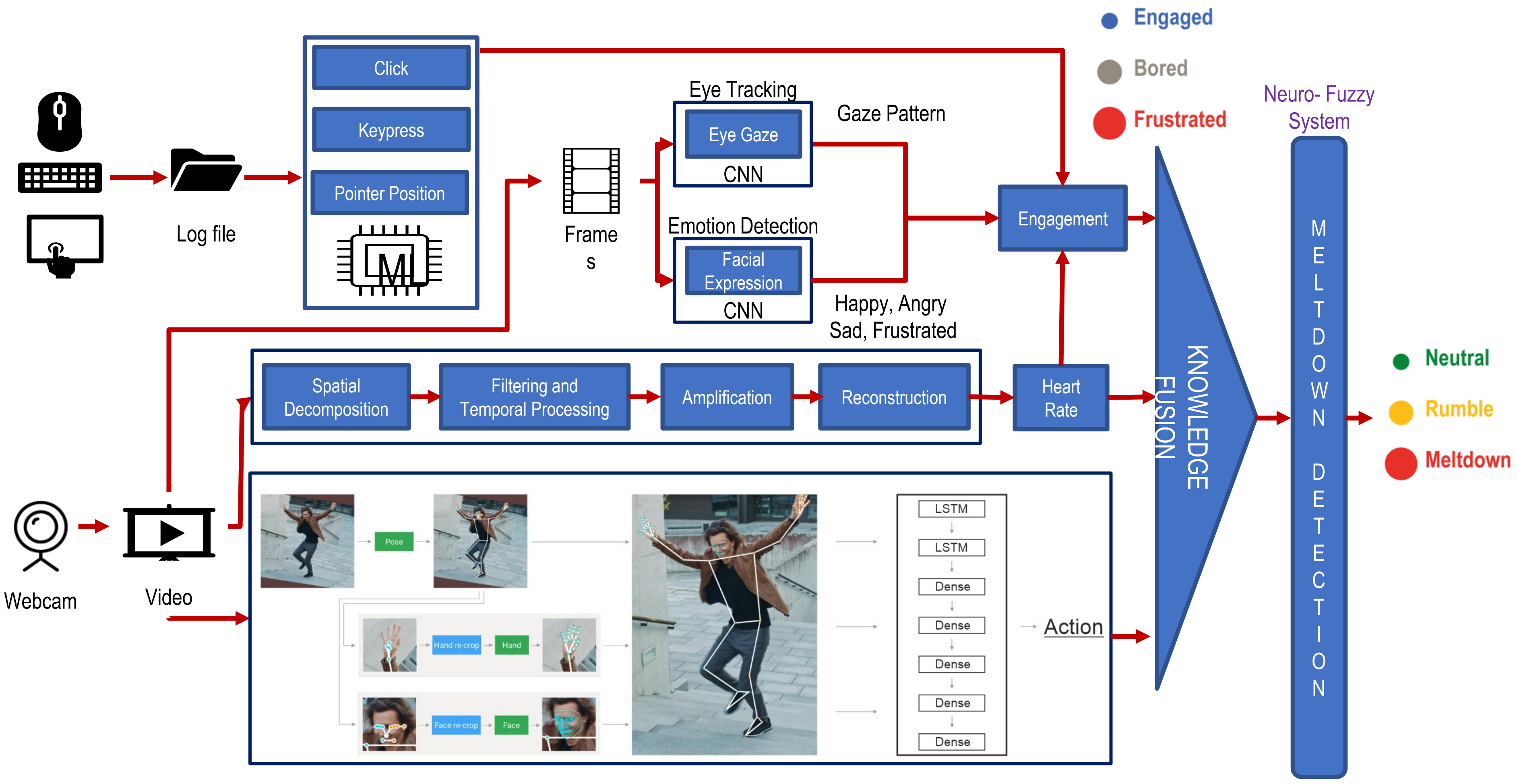
Addresses sensitivity and privacy concerns associated with video data



- Addresses sensitivity and privacy concerns associated with the use of raw video data in a privacy-preserving approach.
- Our framework takes 3D cloud datapoints (468 for facial mesh, 33 joint points) as inputs into machine learning models to predict engagement & meltdown events and presents these as a traffic lights system to teaching staff, parents and carers, or work colleagues.



Model for Engagement/Meltdown Detection



Inclusive, ethical, bias-aware achievements and remaining challenges

Efforts/Successes

- Uses a privacy-preserving approach for predicting state of arousal using the generated cloud point dataset.
- No need for wearables – all sensing is remote/noncontact (including physiological).
- Curated a MM dataset and accurate models.
- Labelling checklist created by Educational Psychologists.
- Co-designed a simple traffic light system to present rumble and meltdown states, and other states related to learning.

Remaining Challenges

- Data scarcity & representational bias (small dataset which could under-represent the target group).
- Model opacity & fragile explanations (opaque multimodal models; multimodal features hard to interpret).
- Consent & capacity (some students were SLD+ASD).
- Privacy & surveillance risks (continuous monitoring, mitigated using MediaPipe).
- Ethical concerns: Consent/capacity or supported decision-making concerns. Stigma/deficit framing or harmful labels (Meltdown!!).
- Explanation will be required for traffic lights in three levels.

Research Example 2: EEG-Based Cognitive Range Classification – The Promise

Young adults with Down Syndrome have a greatly increased risk of dementia (Alzheimer's Association).

The retrosplenial cortex (RSC) plays a significant role in spatial learning and memory and is functionally disrupted in the early stages of AD (https://www.jneurosci.org/content/42/37/7094).

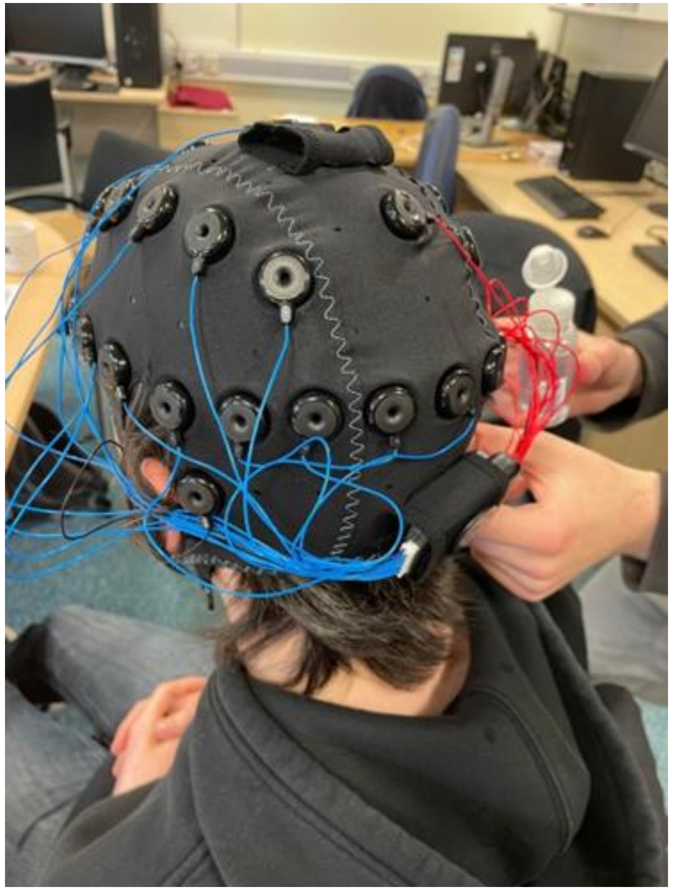
Clinical screening tools are not usually designed with this group in mind and rarely applied.

Collecting EEG data whilst participants undertake spatial navigation tasks of varying difficulties might help us build a ML model top predict cognitive changes over time & develop a novel, engaging, game-based tool to identify issues at an early stage.

Anchor Reference: Harris, M. C., Fryer, D. L., Mahmud, M., Rahman, M. A., Vyas, P., Shopland, N., ... & Brown, D. J. (2024, December). Towards a Machine Learning Model to Classify Cognitive Ability Using EEG Data and Virtual Spatial Navigation Task Scores in Intellectually Disabled Adults. In International Conference on Neural Information Processing (pp. 188-202). Singapore: Springer Nature Singapore.

Co-designing a virtual spatial navigation task.

- **Co-designed:** with young adults with ID & Autism for better outcomes and usability.
- **Task:** Triangle completion task to assess ability to use cues from self motion to estimate position and orientation relative to starting position without visual cues.
- **Platform:** Developed for the PiCO Neo 3 headset & Emotive EPOC Flex.
- **Outcome measures:** Distance from the final target, task completion time and deviation angle recorded by the VR headset using its inside-out tracking capabilities.

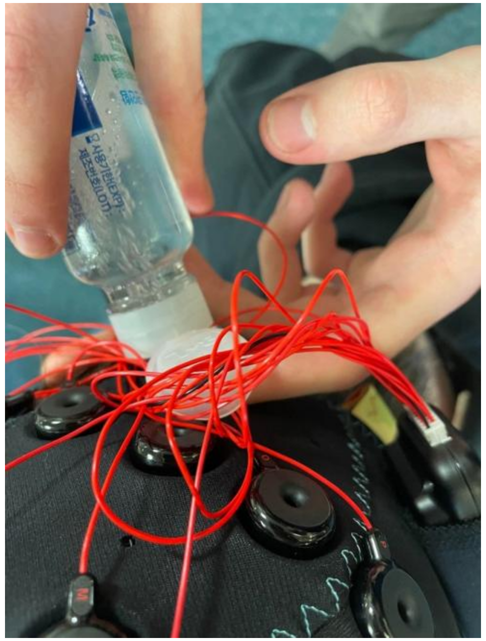


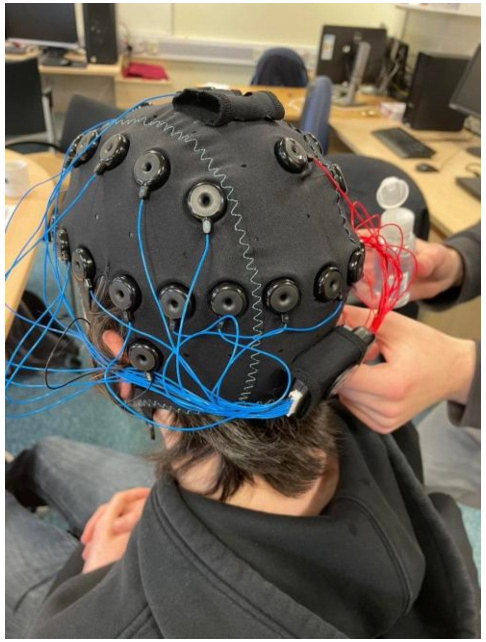
Towards an immersive game-based tool to automatically detect cognitive changes over time…

Aim: Develop a tailored, game-based tool combining VR spatial navigation and brain informatics to assess cognitive changes over time.

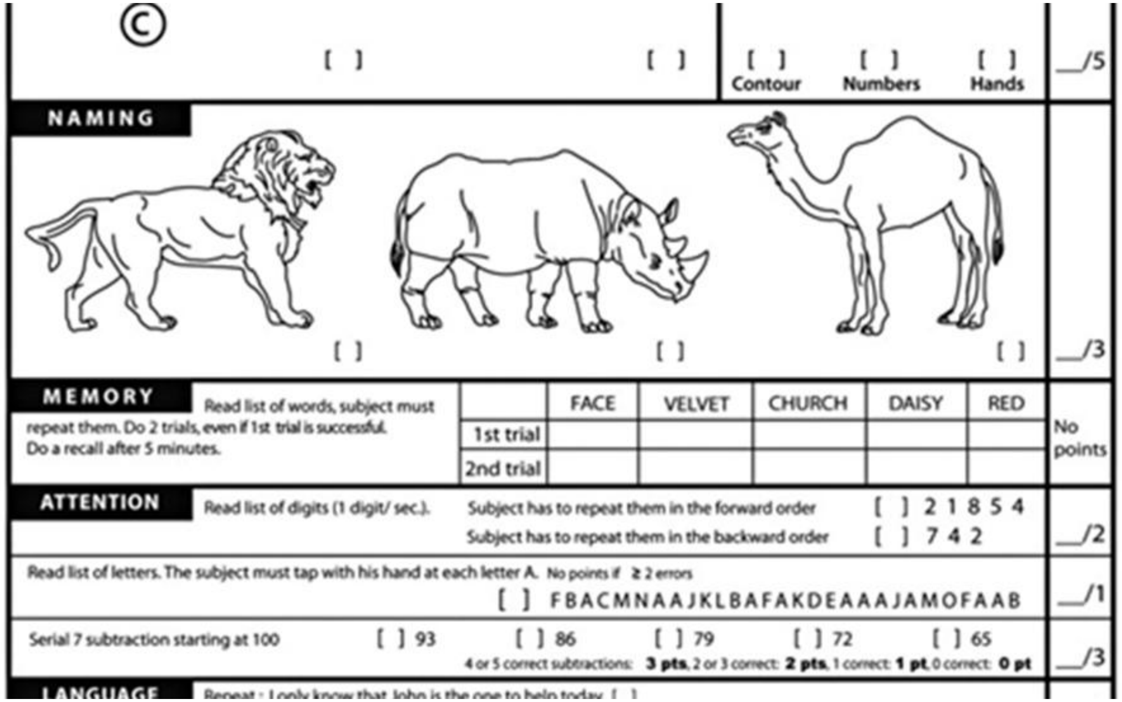
Goals:

- Assess whether our task requires the participant to recruit regions of the brain associated with spatial navigation.
- Use these brain signals labelled with spatial navigation task performance metrics and cognitive assessment scores to train a machine learning model (supervised learning).
- Use this machine learning model to track cognitive changes over time for preventative healthcare.





Data Collection, Labelling and Feature Extraction

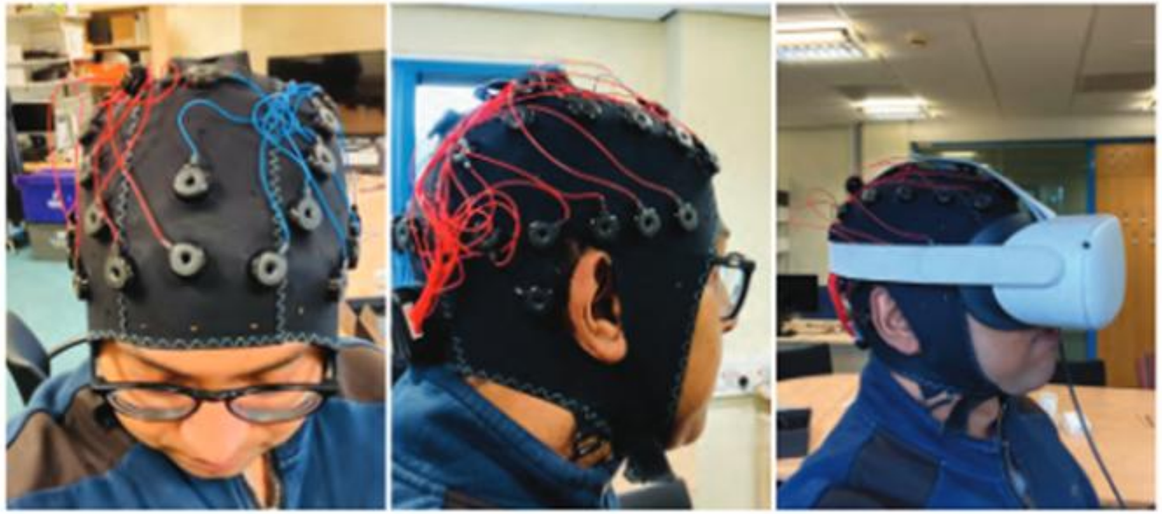


Aim: To train classification models to predict cognitive ranges.

Participants:

- 23 individuals (aged 21-50) with Down's syndrome, Williams syndrome, ASD and ASD with SLD.
- Each participant completed data collection on two separate occasions (minimum two weeks apart).
- Task related Outcome Measures.
- Each participant completed the MoCA-B.
- 59 sessions recorded, for each mode 60 secs of EEG data collected before reaching final location (Point C) and data securely archived.
- Relatively small dataset, but substantial in terms of working with participants with ID.

Machine learning models



Montage used for collecting EEG data. Optimised for collecting data from the parietal cortex, frontal cortex and prefrontal cortex

Neural Network - Simple feedforward neural network (implemented in pytorch).

- It consists of one hidden layer with 64 units and Leaky ReLU activation, sandwiched between two dropout layers.

36 EEG features in total:

- Power spectral density for groups of channels were computed using the multitaper method, focusing on specific frequency bands: Delta (1--4 Hz), Theta (4—8 Hz), Alpha (8—12 Hz) and Beta (12-–30 Hz).
- The EEG channels were grouped according to specific brain regions to allow for regional analysis of the EEG signals.
- Twelve regions were identified: frontal midline, parietal midline, left and right frontal, left and right temporal, left and right central parietal, left and right parietal, and left and right occipital. This resulted in 36 features.

VR task performance metrics also used:

- Deviation angle.
- Distance from the final target and task completion time (recorded by the VR headset).

What have we achieved so far in terms of inclusive, ethical, bias-aware AI in this project?

Efforts/Successes

- Co-designed with young adults with ID.
- Curated dataset (23 participants, 59 sessions).
- We made recommendations about how MoCA-B could be adapted to better suit participants with ID (NS'24)

Remaining Challenges

- **Data scarcity & representational bias:** need much more data and cohort effects.
- **Model opacity & fragile explanations:** CNN and explainability challenges.
- **Consent & capacity:** All consent for themselves but do they rally understand what they are consenting to?
- **Privacy & surveillance risks:** 'Brainprints' are unique and are vulnerable to re-identification.
- **Worst possible harm?** You decide…

Some potential approaches to explainability for cognitive range classification….

Explainability could be built in at three levels: which EEG features matter, when they matter during navigation, and how they relate to observable task performance, so clinicians and support staff can see why the system infers cognitive change in each young person.

Some opportunities for explainability from the literature:

- Encode longitudinal labels (baseline, early improvement/decline, plateau) so the model predicts clinically meaningful change, not just trial-by-trial success.
- Present clinicians with summaries such as "*over 12 weeks, this participant's navigation errors decreased while the model increasingly relies on parietal theta features, suggesting more efficient spatial encoding*".
- Involve families and support workers in co-design to ensure explanations respect communication needs, autonomy, and consent in this population.

NeuroBridgeAI: Trustworthy AI-supported early childhood intervention pathways for neurodiverse children

The project will develop an innovative modular micro-credential ecosystem enabling teaching and clinical staff to:

- critically supervise AI-assisted learning & diagnostic systems
- interpret AI-generated developmental analytics
- validate AI-supported recommendations
- identify bias and unsafe outputs
- implement AI Act-compliant workflows
- apply adaptive intervention planning for neurodiverse children

Wrap-up, take-home messages: Human centric design approaches drive explainability and trustworthy, unbiased, privacy preserving systems that create impact.

Key takeaways include:

- Explainability is not optional—it is fundamental to trust and autonomy;
- Emerging applications such as VR-based therapy, EEG-informed profiling, and privacy-preserving monitoring illustrate both promise and complexity;
- Ethical challenges related to bias, privacy and consent, demand proactive, interdisciplinary solutions; &
- Co-design with affected communities transforms technology from a top-down intervention into a collaborative tool for empowerment.

Michael and Leigh send greetings…

Anchor Refs

- Joyce, D. W., Kormilitzin, A., Smith, K. A., & others. (2023). Explainable artificial intelligence for mental health through transparency and interpretability for understandability. NPJ Digital Medicine, 6(1), 6. https://doi.org/10.1038/s41746-023-00751-9
- Kim, J., Maathuis, H., & Sent, D. (2024). Human-centered evaluation of explainable AI applications: A systematic review. Frontiers in Artificial Intelligence, 7, Article 1456486. https://doi.org/10.3389/frai.2024.1456486
- Giovine, C., & Roberts, R., with Pometti, M., & Bankhwal, M. (2024, November 26). Building AI trust: The key role of explainability. McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/building-ai-trust-the-key-role-of-explainability
- Senoner, J., Schallmoser, S., Kratzwald, B., Feuerriegel, S., & Netland, T. (2024). Explainable AI improves task performance in human–AI collaboration. Scientific Reports, 14(1), 1–13. https://doi.org/10.1038/s41598-024-82501-9
- Harris, M. C., Fryer, D. L., Mahmud, M., Rahman, M. A., Vyas, P., Shopland, N., Connor, B., Lewis, J., & Brown, D. J. (2024). Towards a machine learning model to classify cognitive ability using EEG data and virtual spatial navigation task scores in intellectually disabled adults. In ICONIP 2024 (pp. 188–202).
- Turabee, Z. B., et al. (2025). A privacy-preserving approach for engagement and meltdown detection in children with autism using machine learning models. In M. Mahmud, M. Doborjeh, K. Wong, A. C. S. Leung, Z. Doborjeh, & M. Tanveer (Eds.), Neural information processing: ICONIP 2024 (Communications in Computer and Information Science, Vol. 2285). Springer. https://doi.org/10.1007/978-981-96-6957-8_19
- Premkumar P, Heym N, Myers JAC, Formby P, Battersby S, Sumich AL and Brown DJ (2024) Augmenting self-guided virtual-reality exposure therapy for social anxiety with biofeedback: a randomised controlled trial. Front. Psychiatry 15:1467141. doi: 10.3389/fpsyt.2024.1467141
- Woodward, K., Kanjo, E., Brown, D. J., et al. (2023). In the hands of users with intellectual disabilities: Co-designing tangible user interfaces for mental wellbeing. Personal and Ubiquitous Computing, 27, 2171–2191. https://doi.org/10.1007/s00779-023-01752-x

Contact

David J. Brown.

Interactive Systems Research Group

Computing and Technology Team

NTU, UK.

Tel: +44 115 848 8350

david.brown@ntu.ac.uk

Google Scholar Profile – David Joseph Brown @ https://scholar.google.co.uk/citations?user=4fiZo_IAAAAJ

PHASE IV AI

Co-funded by the Erasmus+ Programme of the European Union

DIVERSASIA

AI-TOP

Funded by the European Union