# Machine Learning for Clinical EEG

## Achieving State-of-the-Art Classifier Performance and Exploring Trustworthy Modes of Clinical Translation


**David Western**

**Wallscourt Fellow in Health Technology**

University of the West of England, UWE Bristol


7th May, 2025

# The Problem

- Electroencephalography (EEG) is routinely used in the assessment of epilepsy and other neurological conditions.

- Ongoing research suggests its clinical potential is much broader than its current use.
  - AAN recommends continuous EEG monitoring for ICU patients with altered mental state [1].
  - Limited capacity discourages greater use [2].

- Interpreter time is half the cost - $251 of $501 for a 48-hour recording in 2013 [3].

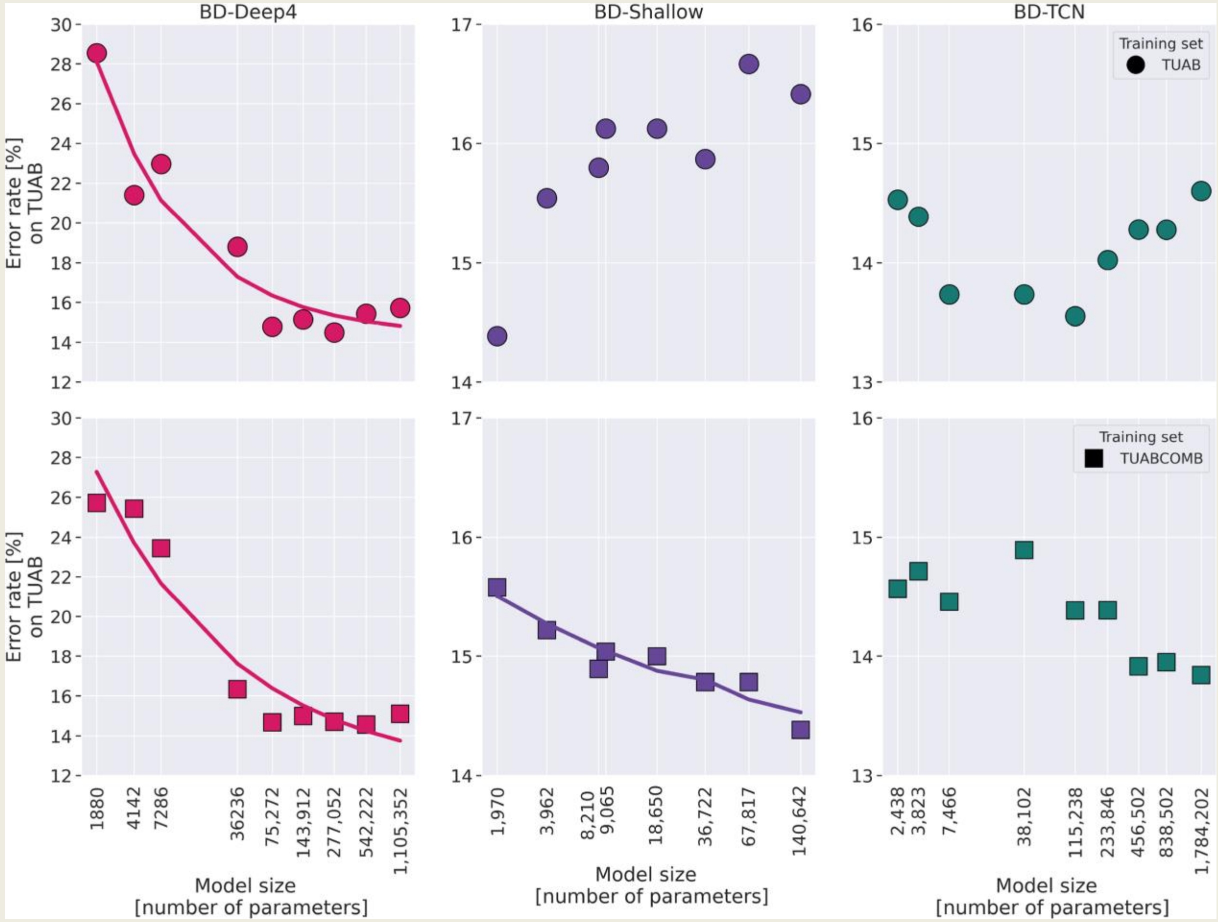

[1] Herman, S.T. *et al.* (2015) Consensus Statement on Continuous EEG in Critically Ill Adults and Children, Part I: Indications. 32 (2).

[2] Park, A., Chapman, M., McCredie, V.A., Debicki, D., Gofton, T., Norton, L. and Boyd, J.G. (2016) EEG utilization in Canadian intensive care units: A multicentre prospectiv e observational study. *Seizure* [online]. 43, pp. 42–47.

[3] Abend, N.S., Topjian, A.A. and Williams, S. (2015) How much does it cost to identify a critically ill child experiencing electrographic seizures? *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society* [online]. 32 (3), pp. 257–264.

# AI for EEG Reporting: State-of-the-Art

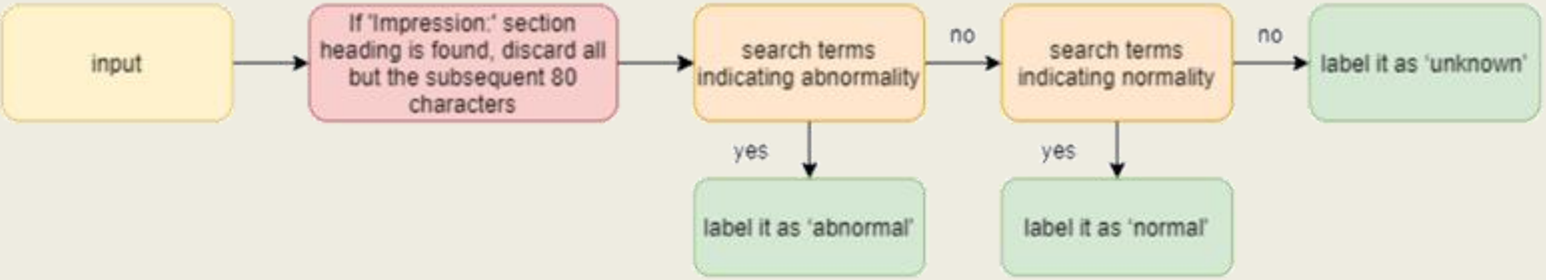
- Most fundamental task:
  - normal/abnormal classification
- One widely used dataset:
  - Temple University Hospital Abnormal EEG Corpus (TUAB)
- Performance ceiling hypothesis [4]
  - Inter-rater agreement limits accuracy to ~90%...?

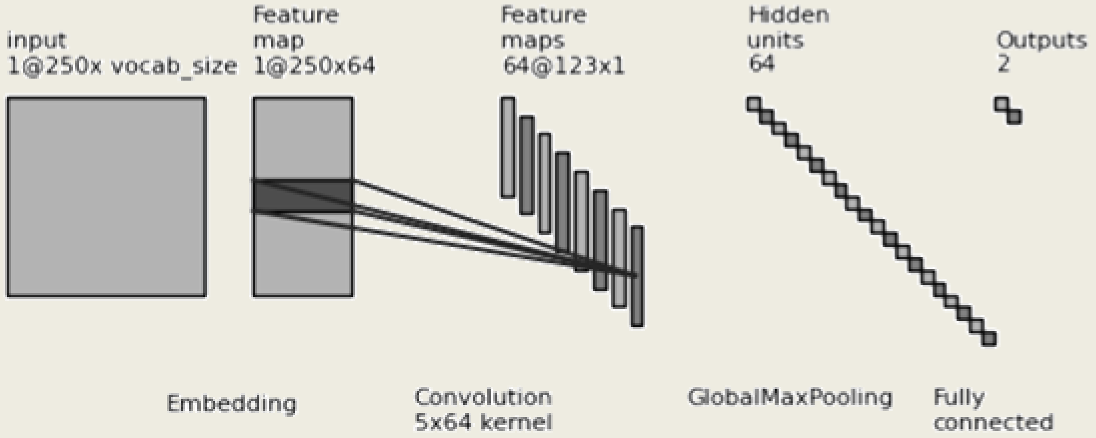



[4] Kiessner *et al*, ‘Reaching the ceiling? Empirical scaling behaviour for deep EEG pathology classification’, *Computers in Biology and Medicine*, vol. 178, p. 108681, Aug. 2024, doi: 10.1016/j.compbiomed.2024.108681.

# Increasing Data Availability

- Automated labelling based on reports increases available training data from 2,717 to 17,402 recordings [5].
- Two simple algorithms to extract labels from report text: rule-based and text CNN.
- No substantial change in EEG classifier performance, but improved ease in curation of new datasets (subject to access!)
- Process increased our understanding of original TUAB labelling.



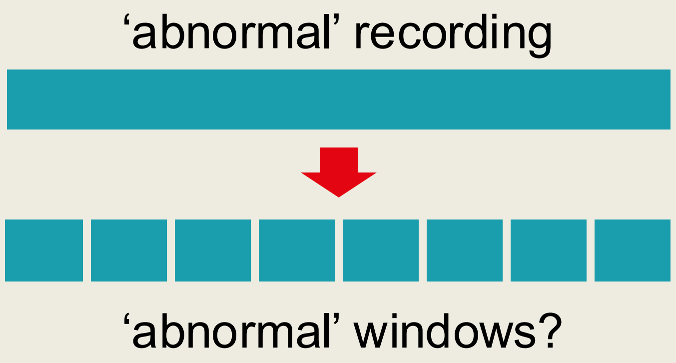




[5] Western et al., 'Automatic Report-Based Labelling of Clinical EEGs for Classifier Training', in IEEE SPMB 2021 (Signal Processing in Medicine and Biology), Philadelphia, PA, USA, 2021.

# Performance Ceiling Revisited

- TUAB labels are from panel consensus, not from original reports.
  - Hence inter-rater agreement does not limit accuracy within the dataset

- Also, the labels apply to recordings, but these are typically divided into smaller windows for training.

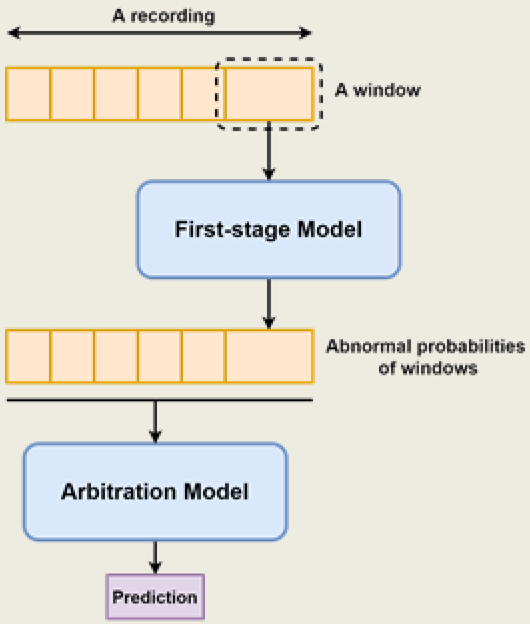



- We went on to ‘break’ the performance ceiling in three ways, increasing SoTA accuracy from 89.8% to 99.0%.
  - Multiple Instance Learning [6, 7]
  - Audio architecture [8]
  - Multimodal learning [9]

[6] Y. Zhu, L. Canham, and D. Western, ‘Scope and Arbitration in Machine Learning Clinical EEG Classification’, in *2023 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, Dec. 2023, pp. 1–7. doi: 10.1109/SPMB59478.2023.10372635.
[7] Y. Zhu, R. Kandasamy, L. J. W. Canham, and D. Western, ‘Window Stacking Meta-Models for Clinical EEG Classification’, Jan. 14, 2024, *arXiv*: arXiv:2401.10283. doi: 10.48550/arXiv.2401.10283.
[8] Y. Zhu and D. Western, ‘Adapting Deep-Learning Audio Models for Abnormal EEG Classification’, in *2024 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, Nov. 2024, pp. 1–8. doi: 10.1109/BHI62660.2024.10913666.
[9] Zhu et al. “Integrating Clinical Context with Signal Analysis for Multimodal EEG Classification” In preparation.

# Multiple Instance Learning [6, 7]

- Common practice is to divide recordings into smaller windows for training.
- Result is 'weak' inherited labels.
- This causes low sensitivity
- We introduce a second machine learning stage to optimise aggregation of per-window outputs.




| Model | Accuracy | Sensitivity | Specificity |
|---|---|---|---|
| 1D-CNN (T5-O1 channel) (Yıldırım et al., 2020) | 79.3% | 71.4% | 86.0% |
| 1D-CNN (F4-C4 channel) (Yıldırım et al., 2020) | 74.4% | 55.6% | 90.7% |
| Deep4 (Schirrmeister et al., 2017) | 85.4% | 75.1% | 94.1% |
| TCN (Gemein et al., 2020) | 86.2% | – | – |
| ChronoNet (Roy et al., 2019) | 86.6% | – | – |
| Alexnet (Amin et al., 2019) | 87.3% | 78.6% | 94.7% |
| VGG-16 (Amin et al., 2019) | 86.6% | 77.8% | 94.0% |
| Fusion Alexnet (Alhussein et al., 2019) | 89.1% | 80.2% | 96.7% |
| Fusion CNN (Muhammad et al., 2020) | 89.8% | 81.3% | 96.9% |
| Scope and Arbitration (Deep4-ANN-Hybrid) Zhu et al. (2023) | 93.3% | 92.0% | 92.9% |
| **Window-Stacking Meta-Model (TCN-XGBoost-Raw)** | **99.0%** | **98.1%** | **100%** |

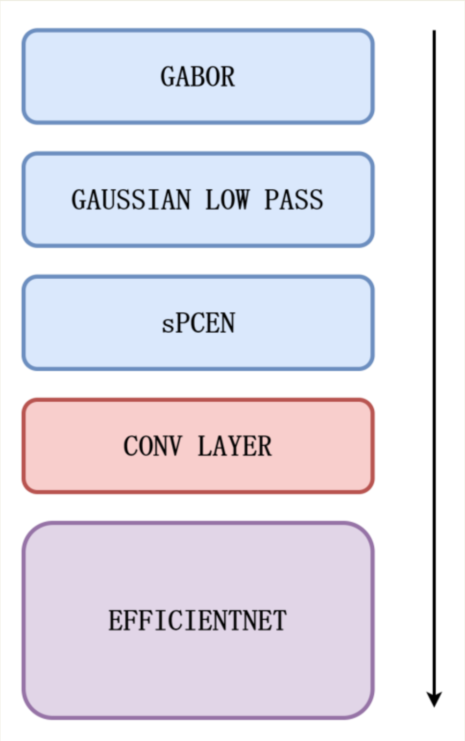

[6] Y. Zhu, L. Canham, and D. Western, 'Scope and Arbitration in Machine Learning Clinical EEG Classification', in *2023 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, Dec. 2023, pp. 1–7. doi: 10.1109/SPMB59478.2023.10372635.
[7] Y. Zhu, R. Kandasamy, L. J. W. Canham, and D. Western, 'Window Stacking Meta-Models for Clinical EEG Classification', Jan. 14, 2024, *arXiv*: arXiv:2401.10283. doi: 10.48550/arXiv.2401.10283.

# Adapting Audio Models for EEG [8]

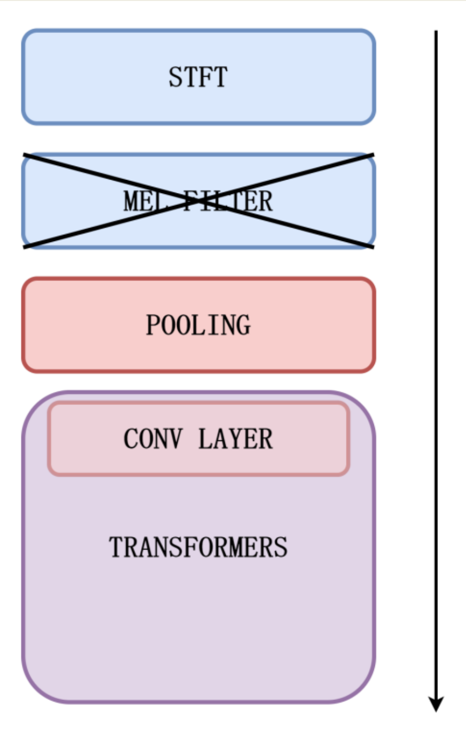
- Pre-training EEG classifiers on audio data doesn't help.
- But architectures developed for audio can be effectively transferred to EEG.

| **Model** | **Learning Rate** | | |
|---|---|---|---|
| | **0.001** | **0.0001** | **0.00001** |
| Deep4 | $\boxed{0.854}$ | 0.860 | 0.800 |
| **EEG-PaSST** | 0.837 | 0.949 | $\boxed{0.957}$ |
| **EEG-LEAF** | 0.880 | $\boxed{0.940}$ | 0.878 |

EEG-LEAF



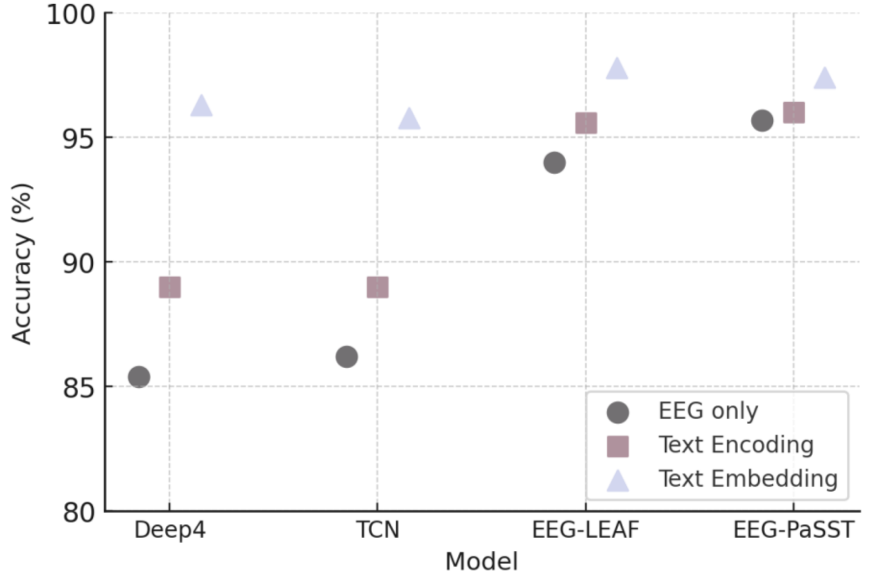

EEG-PaSST

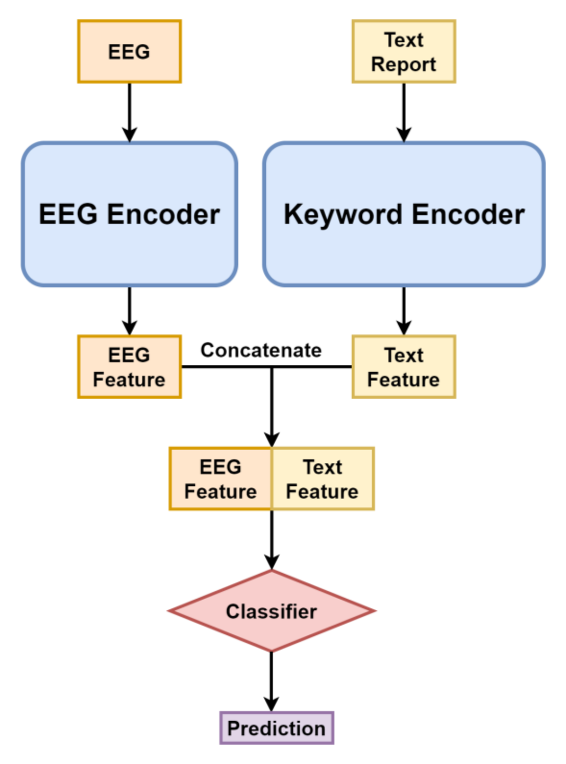


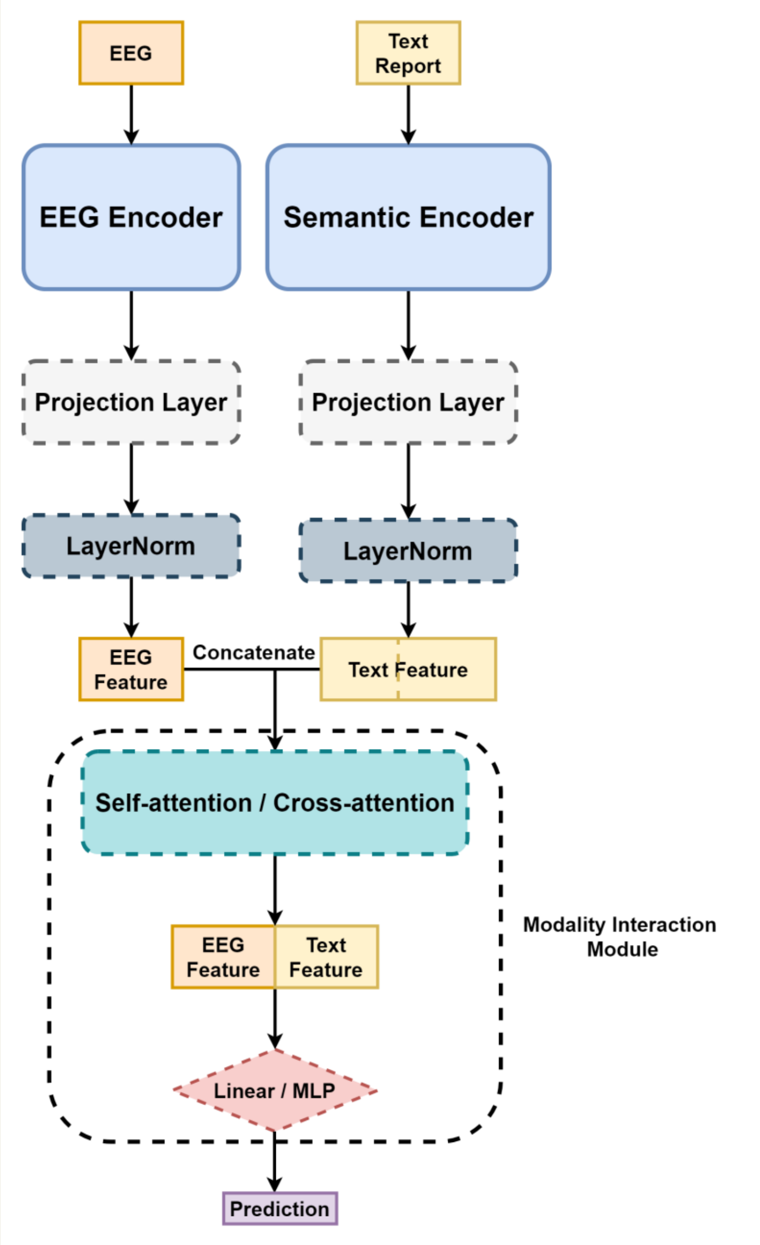

[8] Y. Zhu and D. Western, 'Adapting Deep-Learning Audio Models for Abnormal EEG Classification', in *2024 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, Nov. 2024, pp. 1–8. doi: 10.1109/BHI62660.2024.10913666.

# Multimodal EEG+Text Classifier [9]

- Conventional EEG report text includes some *a priori* contextual information
- We explored two approaches to integrate this info into an EEG classifier: keyword encoding (gender, age, consciousness, epilepsy history) and semantic embedding (LLM).




[9] Zhu et al. “Integrating Clinical Context with Signal Analysis for Multimodal EEG Classification” In preparation.

# Future Work

- Shift in focus to clinical translation

- Need more data for validation and further development

  Existing widely accessible databases
  - Temple University Hospital EEG Corpus
  - Harvard EEG database
  - NMT Scalp EEG Dataset (South Asian)

  Limitations:
  - Widely used for many years – overfitting
  - Not from NHS cohorts
  - Limited ground-truth quality*
  - Limited metadata / linked data for multimodal AI

- Data linkage will enable further performance gains
  - Multimodal data for richer input
  - Training on clinical outcomes to surpass human performance.

- More data linkage means more privacy risk - Trust is essential

# Thanks to Collaborators!

- Yixuan Zhu
- Luke Canham
- Rohan Kandasamy
- Felix May
- Sammie Taylor

- And many more!

# Part 2: Exploring Trustworthy Modes of Clinical Translation

# Common Scenario

"We developed this tool, so now we need to start a company to get it into clinical use. Not to make lots of money, but because we need an entity to demonstrate compliance, accountability, etc."

- Could a social enterprise be more effective than a profit-led company in achieving impact?

clinician/academic

# Hypothesis

- When trust is highly valued, a social enterprise can outcompete profit-led companies in a free market…
- … and machine-learning-for-health is such a market.

# Defining ‘social enterprise’ etc.

- Charity vs Limited Company

- ‘Social’ flavours of limited company:
  - **B Corp**
    - Certification framework
    - Emphasis on ‘how’ rather than ‘why’
  - **Social Enterprise**
    - Not a formal legal term. SEUK certification.
    - >50% income through trading, >50% profit reinvested
  - **Community Interest Company (CIC)**
    - Social enterprise with community purpose registered with CIC regulator at Companies House
    - Legal assurances against repurposing include:
      - Asset lock
      - Dividend cap

# Commercial Value of Trust

- Hypothesis: "**When trust is highly valued**, a social enterprise can outcompete profit-led companies in a free market, and machine-learning-for-health is such a market."

- Supply of vs. demand for trust; is adoption of machine-learning-for-health constrained by a trust deficit?

- If so, does a social-enterprise approach address that deficit?

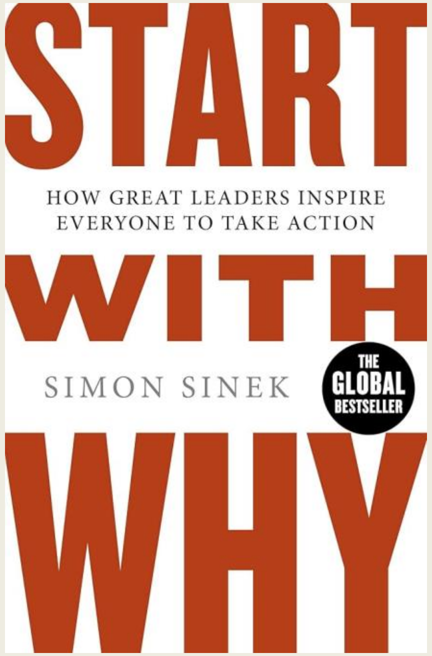
# Where Does Trust Come From?

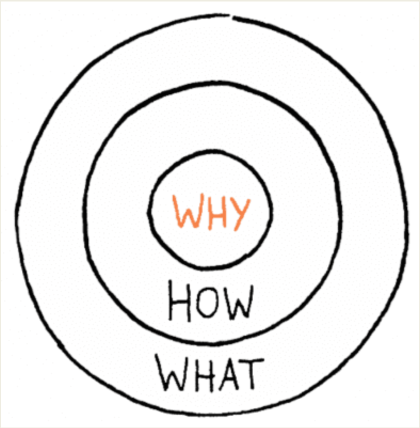
- General factors: Ability, benevolence, and integrity (+ propensity to trust) [1]

- Application-specific considerations:
    - Trustors: Patients, Clinicians, Institutions
    - Factors: Explainability, reliability+bias, liability, impact on profession, ease-of-use, privacy, autonomy, human care relationship, integration, organisational culture (employee 'buy in')

- Many guidelines on how to achieve trust in AI – e.g. FUTURE-AI [2] – focussed on **how** to operate, not **why**.

- Where does the deficit lie and where can a social-enterprise approach help?

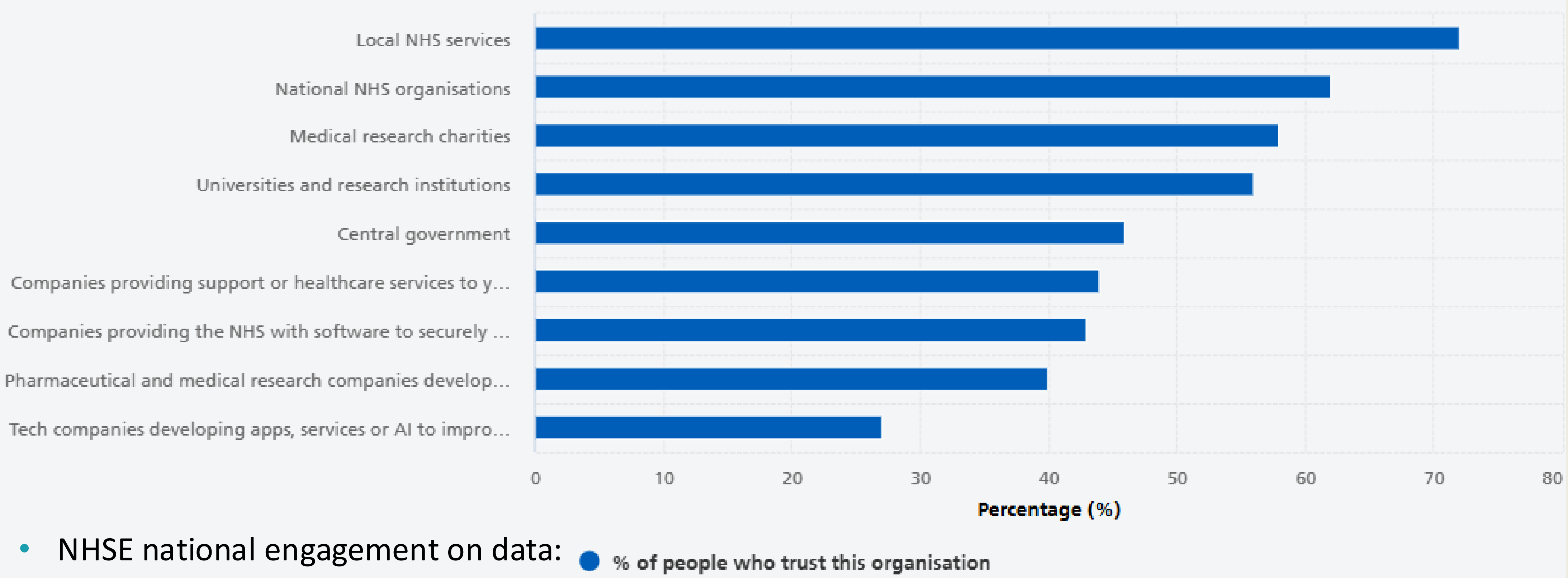



[1] R. C. Mayer, J. H. Davis, and F. D. Schoorman, 'An Integrative Model of Organizational Trust', *The Academy of Management Review*, vol. 20, no. 3, pp. 709–734, 1995, doi: 10.2307/258792.
[2] K. Lekadir *et al.*, 'FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare', *BMJ*, vol. 388, p. e081554, Feb. 2025, doi: 10.1136/bmj-2024-081554.

# Trust in Organisations

Figure 2: Trust in organisations with patient data, from most trusted to least trusted




- NHSE national engagement on data:


[3] 'Public attitudes to data in the NHS and social care', NHS England Digital. https://digital.nhs.uk/data-and-information/keeping-data-safe-and-benefitting-the-public/public-attitudes-to-data-in-the-nhs-and-social-care

[4] NHS England, 'National engagement on data: Cohort 1 report', NHS Transformation Directorate. Accessed: May 07, 2025. [Online]. Available: https://transform.england.nhs.uk/key-tools-and-info/data-saves-lives/national-public-engagement-on-the-use-of-health-data/national-engagement-on-data-cohort-1-report/

# Trust in Organisations

- NHSE national engagement on data [4] - Case study 7:
  - Public asked to consider “how an AI tool developed by a **medical research charity** could be used to improve diagnosis and treatment for breast cancer”…
  - … then consider same, swapping charity for “**start-up pharmaceutical company**”

*“I worry [with it being a pharmaceutical company] because it's about lining someone's pockets.”*

*North London, female, Workshop 1*

*“I'm okay with them making a profit that’s what they exist to do. So long as it is for the greater good and the correct processes in place. I am okay with it, but there has to be a lot of safeguards in place.”*

*North London, male, Workshop 1*

*“It makes me question it a bit, how much of it is for the greater good. If they find something more accurate and more precise but is it accessible to everyone or is it for their greed.”*

*North London, female, Workshop 1*

***Participant preference:*** Overall, participants felt significantly less comfortable with their data being shared with a private pharmaceutical company, than with an academic research charity. This feeling stemmed from a belief that a pharmaceutical company would be more financially motivated, with profits ultimately going to shareholders, whereas a charity would be more focused on improving health outcomes for patients.

[4] NHS England, ‘National engagement on data: Cohort 1 report’, NHS Transformation Directorate. Accessed: May 07, 2025. [Online]. Available: https://transform.england.nhs.uk/key-tools-and-info/data-saves-lives/national-public-engagement-on-the-use-of-health-data/national-engagement-on-data-cohort-1-report/

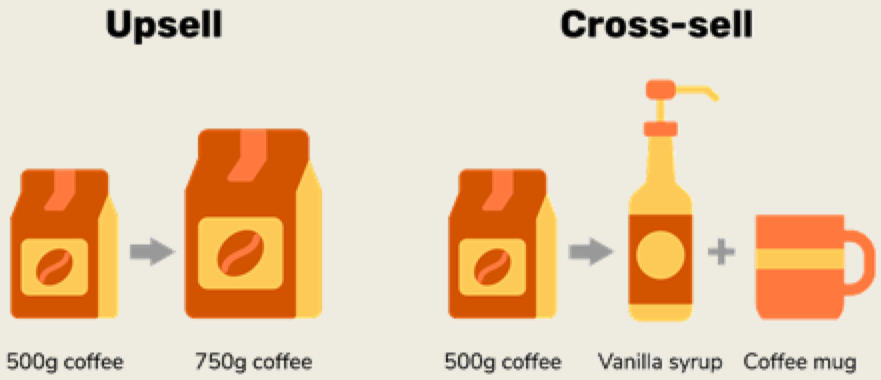
# AI Adoption and Profit Motives

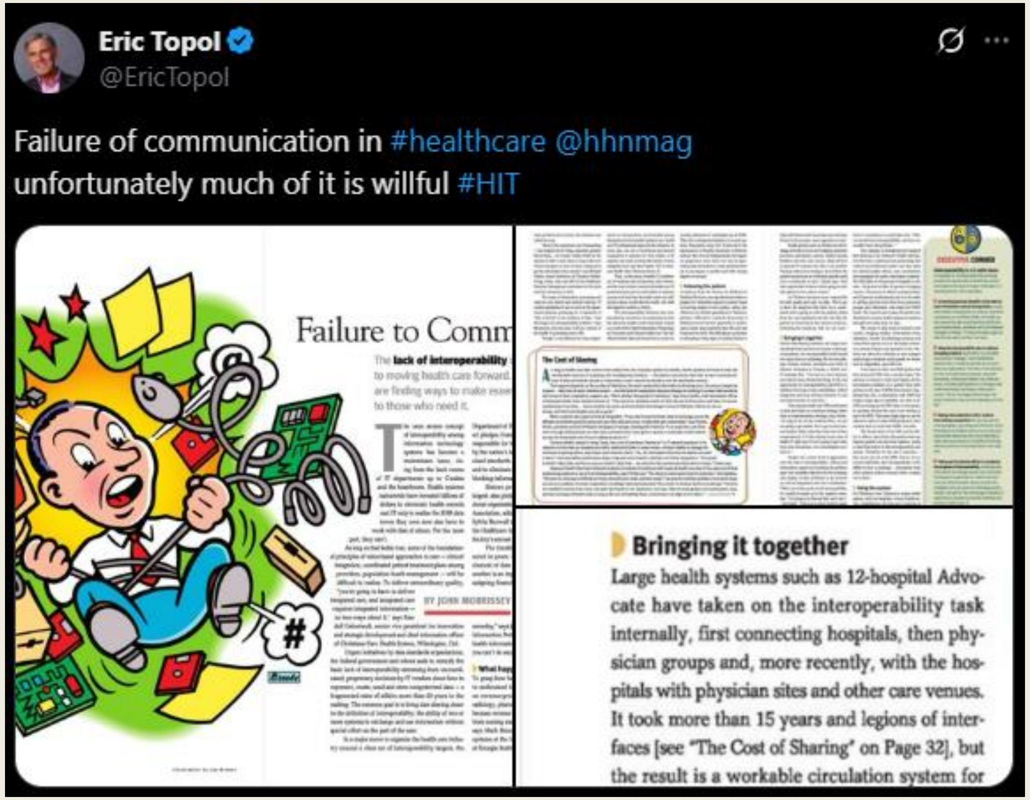
Areas with potential for misaligned incentives:

- Transparency vs proprietary interests
  - Explainability
  - Evaluation


- Upselling/cross-selling

- Limited interoperability



# Precedent for Social Enterprise in Health

- Non-profit care homes outperform for-profit counterparts in quality and access [5]
- Sirona CIC delivering NHS services across BNSSG region
- OxVent Ltd – ‘social venture’ – low-cost ventilators
- Nothing in health data technologies?

[5] A. A. Amirkhanyan, H. J. Kim, and K. T. Lambright, ‘Does the public sector outperform the nonprofit and for-profit sectors? Evidence from a national panel study on nursing home quality and access’, *Journal of Policy Analysis and Management*, vol. 27, no. 2, pp. 326–353, 2008, doi: 10.1002/pam.20327.

# When Does Trust Matter

- R&D
  - Engagement for scoping and design input
  - Data access
- Commissioning
  - Evidence of efficacy
  - Economic case
  - Public perception
- In use
  - Patient experience
  - Clinician experience

# Hypothesis Revisited

- When trust is highly valued, a social enterprise can outcompete profit-led companies in a free market…
- … and machine-learning-for-health is such a market.

- But will it work in practice?
    - Purpose alone is not enough to ensure trustworthiness.
    - Likely positive effect on informal interactions.
    - Formal difference to concrete processes? Perhaps not. Pending ongoing developments e.g. 'Data Pact'.

# Choose My Adventure

- What approach would likely maximise the 'impact' of the AI-for-EEG work presented?
  - Start a charity
  - Start a 'normal' company
  - Social enterprise - re-invest 51% of profits in health tech dev
  - CIC - re-invest 100% of profits
  - Sell/license the tech to a more established for-profit company
  - Other?