Axel's Journal of Neurotypical Studies

Volume 2, Issue 2 (Special Research Note)

The Etiology of Neurotypicality:

Title
Closing the Distance: Explainable AI for Diagnosing Neurotypicality in Disability Assessors with TabPFN-NeuroMix and SHAP

Abstract

Background:
Contemporary disability systems display an intense interest in "screening," "triaging," and "optimising" disabled lives via machine learning, while displaying a striking lack of interest in being in the same room as those lives for more than 45 minutes. This suggests an unexamined phenomenon: *neurotypicality as proximity-avoidant behaviour*, particularly in professionals compelled to assess disabled people with algorithms.

Objective:
This paper proposes a complex machine learning framework to diagnose *high-risk neurotypicality* in clinicians, bureaucrats, data scientists and assorted "innovation leaders" involved in algorithmic disability assessment. We hypothesise that the urge to algorithmically contain disabled people functions as a psychological distancing technology—allowing professionals to comply with policy without risking actual relationship. Exceptions (e.g., John from the footy) are modelled as outlier cases.

Methods:
Adapting a previously published (and helpfully already retracted) TabPFNMix+SHAP framework originally designed for autism diagnosis, we invert the problem: instead of classifying autistic subjects, we classify the people trying to classify autistic subjects. Our proposed *TabPFN-NeuroMix* model ingests multimodal assessor features (linguistic, behavioural, spatial and policy-interaction signals) and outputs a continuous *Neurotypicality Proximity Avoidance Index* (NPAI). Explainability is provided via SHAP, repurposed as *Shapley Ableism Partitioning*, to identify which features most strongly predict algorithmic compulsion.

Results:
Across 642 assessors, TabPFN-NeuroMix outperformed baseline models (logistic regression, "vibe check," and "ask the receptionist") with an AUC-ROC of 0.97 for detecting "high-risk proximity-avoidant neurotypicality." Top SHAP features included: (1) number of times "we just have to follow the guidelines" is used per hour, (2) percentage of appointment spent facing a laptop rather than the disabled person, (3) frequency of requesting *one more*

report despite existing evidence, and (4) tendency to describe disabled people as "participants," "cases," or "complex presentations" but never "people." The only consistent false negative was "John from the footy," whose existing parasocial relationship with "one of the boys" disabled teammate confounded the model.

Conclusion:
Rather than further refining tools to diagnose disabled people, our findings suggest redirecting AI toward the more urgent public health problem: neurotypicals who cannot approach disabled people without a portal, a rubric, or a pre-trained model as emotional PPE. We propose that future NDIS reforms include screening for high-risk algorithmic distancing in decision-makers, alongside mandatory exposure therapy to unscripted disabled humanity.

Keywords: Neurotypicality, Explainable AI, Disability bureaucracy, NDIS, Proximity avoidance, SHAP, TabPFNMix, Algorithmic governance

---

1. Introduction

The last decade has seen an explosion of AI systems promising to *streamline*, *triage* and *optimise* disabled lives, especially autistic lives, usually without asking autistic people what "optimised" might mean to them. Building on a growing literature of AI-based autism diagnosis and explainability, disability systems now host a flourishing ecosystem of algorithms, dashboards and "participant journeys," all seemingly designed to prevent actual human contact from contaminating the data.

The National Disability Insurance Scheme (NDIS) Act 2013, with its charming promises of "choice and control," has become a fertile space for this tendency. The Act nominally requires supports to be "reasonable and necessary," but in practice this has often been translated into "modelled and numerically defendable," with an implied clause: *"and preferably not too emotionally confronting for the delegate."*

This paper starts from a naïve but apparently radical question:

> If the system is so sure that *we* are the problem, why are *they* the ones clinging to algorithms like emotional support animals?

Rather than once again asking machines to diagnose autistic or otherwise disabled people, we invert the lens. We treat *neurotypical proximity-avoidant behaviour*—the persistent need to interpose forms, metrics, and machine learning between oneself and disabled people—as the target of diagnosis.

Specifically, we aim to:

1. Develop a complex, extremely serious-sounding ML model to identify people who experience an overwhelming urge to assess disabled people via algorithms.

2. Use explainable AI to reveal which professional behaviours most strongly predict this urge.

3. Tentatively propose that this behaviour may function as a socially sanctioned strategy to avoid getting “too close” to disabled people—except when the disabled person is John from the footy, in which case *mateship* temporarily overrides the model.

---

2. Methods

2.1 Study design

We conducted a cross-sectional, multi-institutional study targeting four populations:

1. **Clinicians** involved in functional capacity assessment.

2. **NDIS delegates** and planners responsible for accepting or denying supports.

3. **Data scientists and AI researchers** whose grant titles contain the words “autism,” “triage,” “early detection,” or “burden reduction.”

4. **Policy architects** who can quote risk corridors but not the social model of disability.

Inclusion criteria:

- At least one instance of saying “we need more data before we can approve that” despite having a 37-page report.

- Demonstrated comfort with the phrase “evidence-based decision-making” when the evidence is exclusively about disabled people and never about system harm.

- For the “John from the footy” control subgroup: must be on a first-name, nickname, or “he’s a good bloke” basis with at least one disabled person, but only at sporting or barbecue-adjacent events.

Ethics approval was obtained from the Committee for Human Research, who stated that “studying neurotypicals feels refreshing.”

2.2 Feature construction

Unlike traditional autism datasets that track social responsiveness and repetitive behaviours in disabled children, our dataset tracks social responsiveness and repetitive behaviours in *assessors*. Features were derived from:

- **Linguistic behaviour:**
    - Count of phrases per session: “we just have to follow the guidelines,” “it’s not me, it’s the legislation,” “we need an objective tool.”
    - Ratio of “participant” / “client” / “case” to “person” / “name used correctly.”
    - Number of times “risk to the Scheme” is mentioned before “impact on the person.”
- **Spatial and bodily behaviour:**
    - Mean physical distance (in metres) maintained from disabled person vs laptop.
    - Time to first glance away from portal and towards actual human.
    - Whether the assessor sits closer to the door than to the disabled person (binary).
- **Administrative behaviour:**
    - Number of reports requested per year *after* diagnoses are already clear.
    - Frequency of plan reductions following “independent functional assessment” or equivalent.
    - Average time spent on “participant voice” section of report (seconds).
- **Technological excitement index:**
    - Number of times the assessor describes AI as “promising,” “game-changing,” or “disruptive.”
    - Prior authorship of at least one paper arguing for “scalable diagnostic pipelines for ASD.”
- **John from the footy adjustment factor:**
    - Binary indicator for whether the assessor spontaneously says “oh but John from the footy is different, he just gets on with it.”
    - If yes, proportion of empathy directed at John vs other disabled people.

All features were standardised. Missing values (e.g., planner could not recall any disabled person’s name) were imputed using k-nearest neighbours based on similar patterns of

report-hoarding and acronym usage, following the original TabPFNMix preprocessing pipeline.

2.3 Outcome definition

The primary outcome was the **Neurotypicality Proximity Avoidance Index (NPAI)**, modelled as a continuous score in [0,1]:

- 0.0–0.25: *Low-risk neurotypical or non-neurotypical; may spontaneously treat disabled people as peers.*
- 0.26–0.60: *Bureaucratically conditioned neurotypical; still reachable with plain language and tea.*
- 0.61–0.85: *Algorithmically Preoccupied Neurotypical (APN); requires intensive de-spreadsheeting.*
- 0.86–1.00: *Contact-Avoidant Highly Neurotypical (CAHN); will only approach disabled people through a portal, an app, or a 72-page report.*

“John from the footy” is not assigned a fixed NPAI. Instead, his score is stochastically perturbed by a “good bloke prior,” which forces the posterior towards “nah he’s fine” regardless of observed impairment, suffering, or unmet support needs.

2.4 Model architecture: TabPFN-NeuroMix

We adapted the TabPFNMix regressor originally proposed for ASD diagnosis into *TabPFN-NeuroMix*, a hybrid architecture combining:

1. **Pattern Feature Net (PFN) block** to learn nonlinear interactions between:
    - guideline-quoting frequency,
    - laptop proximity, and
    - number of times “reasonable and necessary” is used as a weapon rather than a protection.
2. **Neurotypicality Mixture Head**, a small neural network that:
    - outputs NPAI, and
    - decomposes risk into three latent dimensions: *Avoidance*, *Abstraction*, and *John Exception Handling*.

The model is trained with a binary cross-entropy loss on high vs low NPAI, using consensus ratings from an expert panel of disabled people, carers, and one extremely jaded support coordinator who has seen things.

2.5 Explainability: SHAP

To avoid reproducing the same black-box logic used to justify denying wheelchairs, we employ SHAP to compute local and global feature attributions, following the original study's explainable AI framework. SHAP values here are interpreted as:

> *"How much did this specific behaviour push the model towards classifying you as someone who probably shouldn't be left alone with a spreadsheet and the NDIS portal?"*

Summary and dependence plots were generated to identify which variables most strongly contributed to high NPAI.

---

3. Results

3.1 Overall performance

TabPFN-NeuroMix achieved:

- **Accuracy:** 93.2%
- **Precision (high NPAI):** 92.1%
- **Recall (high NPAI):** 95.4%
- **F1-score:** 93.7%
- **AUC-ROC:** 0.97

It substantially outperformed:

- Logistic regression using only "number of PhDs involved" (AUC 0.61)
- A simple rule-based classifier ("if they say 'we're not funded for that' in the first five minutes, label as high risk"; AUC 0.74)
- The baseline human method: "ask the disabled person afterwards how it felt" (perfectly accurate but apparently "not scalable").

3.2 SHAP-based feature importance

Global SHAP analysis identified the following top contributors to high NPAI:

1. **Frequency of "we just have to follow the guidelines"**
    - Each additional utterance increased NPAI by ~0.07, independent of actual legislative content.
2. **Laptop engagement ratio** (time looking at screen ÷ time looking at person)
    - Ratios >2.5 strongly pushed predictions towards CAHN.
3. **Report Accumulation Index** (reports requested beyond first clear diagnosis)
    - Nonlinear effect: NPAI rose sharply after the third redundant report, plateauing around the seventh, suggesting a "saturation point of plausible deniability."
4. **AI enthusiasm density** (AI-hype words per 1000 spoken words)
    - High density correlated with high NPAI except when the speaker explicitly referenced disabled scholarship; in those rare cases, SHAP values reversed.
5. **Use of "John from the footy" as counterexample**
    - Paradoxically decreased NPAI slightly (the model recognised fleeting contact with reality), while simultaneously increasing predicted risk of under-funding every disabled person *not* named John.

3.3 Individual-level explanations

SHAP force plots revealed typical patterns:

- **Planner 17:** High NPAI (0.91) driven by laptop engagement ratio of 4.3, eight instances of "the Scheme," and zero uses of the participant's name.
- **Clinician 42:** Moderate NPAI (0.58); risk largely due to business model incentives ("we'll need to reassess in six months") rather than personal animosity.
- **Data scientist 9:** Very high NPAI (0.96), driven by publication record on "automated ASD triage" with no disabled co-authors and a suspicious fondness for ROC curves as moral justification.

The model repeatedly mis-handled **John from the footy**, whose NPAI fluctuated wildly depending on whether input text was drawn from (a) consolatory bar talk or (b) actual support planning documents. In talk, John was "just one of the boys." In documents, John effectively disappeared.

4. Discussion

4.1 Diagnosing the diagnosticians

This study demonstrates that the machinery built to classify disabled people can be trivially inverted to classify the people wielding it. The fact that TabPFN-NeuroMix so accurately diagnoses *neurotypical proximity-avoidant behaviour* suggests that:

1. The behaviours are highly patterned.

2. The patterns are visible to disabled people long before they are visible to ethics committees.

3. We never actually needed a deep neural network; a functioning sense of shame would have sufficed.

Our findings support the hypothesis that many professionals' enthusiasm for algorithmic assessment is less about efficiency and more about emotional risk management. The model strongly linked high NPAI to behaviours that:

- maximise *distance* (physical, linguistic, and bureaucratic) from disabled subjects, while

- maximising *closeness* to policy, guidelines and respectable spreadsheets.

In plain language: the more allergic you are to sitting with disabled people as equals, the more you want a model to sit between you.

4.2 The John from the footy effect

The persistent outlier status of "John from the footy" highlights an important phenomenon: proximity avoidance is not evenly distributed. When disabled people appear as familiar archetypes—teammate, nephew, "good bloke from work"—neurotypicality briefly forgets itself. The model's confusion here reflects the system's confusion:

> "We believe in deinstitutionalisation, community inclusion, and algorithmic triage for cost containment—
> but obviously not for John, he's *different*."

This selective exceptionality has policy consequences: the John-like figure becomes the benchmark for "acceptable disability," against which everyone else is measured and usually found excessive.

4.3 Implications for NDIS and similar schemes

Under the NDIS Act's commitment to "choice and control," much attention has been focused on whether disabled people are making sufficiently "reasonable" choices. Less attention has been given to whether delegates and clinicians exhibit "reasonable and necessary emotional functioning" around disabled people.

Our results suggest that:

- Before deploying AI to "optimise" us, schemes might use AI to identify which decision-makers are operating at CAHN-level neurotypicality and therefore unsafe to wield those tools unsupervised.

- Delegates with NPAI above a specified threshold could be required to:

    - co-design plans *with* disabled people present,

    - attend sessions where no laptop is allowed and the only available interface is conversation,

    - read at least one piece of disability scholarship not written by a non-disabled psychiatrist.

In particularly severe cases, we propose issuing these professionals their own quasi-NDIS plans, with funded supports such as "relational capacity building," "empathy skill development," and "assistive technology: sticky notes reminding them that disabled people are real."

4.4 Ethical considerations

At first glance, diagnosing neurotypical people with machine learning may appear ethically fraught—after all, they are a large, vulnerable majority accustomed to being the diagnoser rather than the diagnosed. However, several safeguards are built into our proposal:

- The model is intended as *critical satire*, not as an operational tool.

- Any attempt to operationalise it would instantly prove its own point and so can be safely left to ethics committees to self-immolate over.

- Unlike traditional clinical models, TabPFN-NeuroMix does not claim neutrality. It openly states its purpose: to protect disabled people from algorithmically enthusiastic professionals with unexamined power.

Put differently: this paper does not suggest building new diagnostic prisons; it merely holds up a mirror to the ones already built for us.

4.5 Limitations

This study has several limitations:

- The sample was skewed toward English-speaking bureaucrats; further research is needed on non-English forms of ableist jargon.

- Our operationalisation of “neurotypicality” is behavioural, not neurological; readers seeking a neuron-level account are invited to sit in a room with us for three hours and see which brain starts to melt first.

- SHAP explanations, while useful, cannot capture the full richness of systemic power; some things are better explained by “they are protecting the budget” than by marginal feature contributions.

---

5. Conclusion

By inverting the usual direction of diagnosis, TabPFN-NeuroMix and its NPAI score make one thing abundantly clear: the real black box in disability systems is not autistic cognition but neurotypical reluctance to relate without mediation.

Complex machine learning is, once again, unnecessary but rhetorically useful. If we can build a model that reliably detects when someone is using “the guidelines” as a shield against proximity, perhaps the next step is simpler:

- Stop trying to perfect AI that diagnoses us.

- Start asking why you need that AI to be comfortable being near us.

- And if the answer is “look, we’re just following the legislation,” then congratulations: the model has already flagged you.

John from the footy, as usual, will be fine. The rest of us would prefer something more ambitious than mateship-based exception handling: systems, and relationships, that do not require an algorithm to justify getting close.