Deloitte.

Together makes progress

VU VRIJE UNIVERSITEIT AMSTERDAM

Taxonomies for Connecting Technical Components of AI-based Fraud Detection Systems and Fairness Approaches

By: Sebastiaan Berendsen

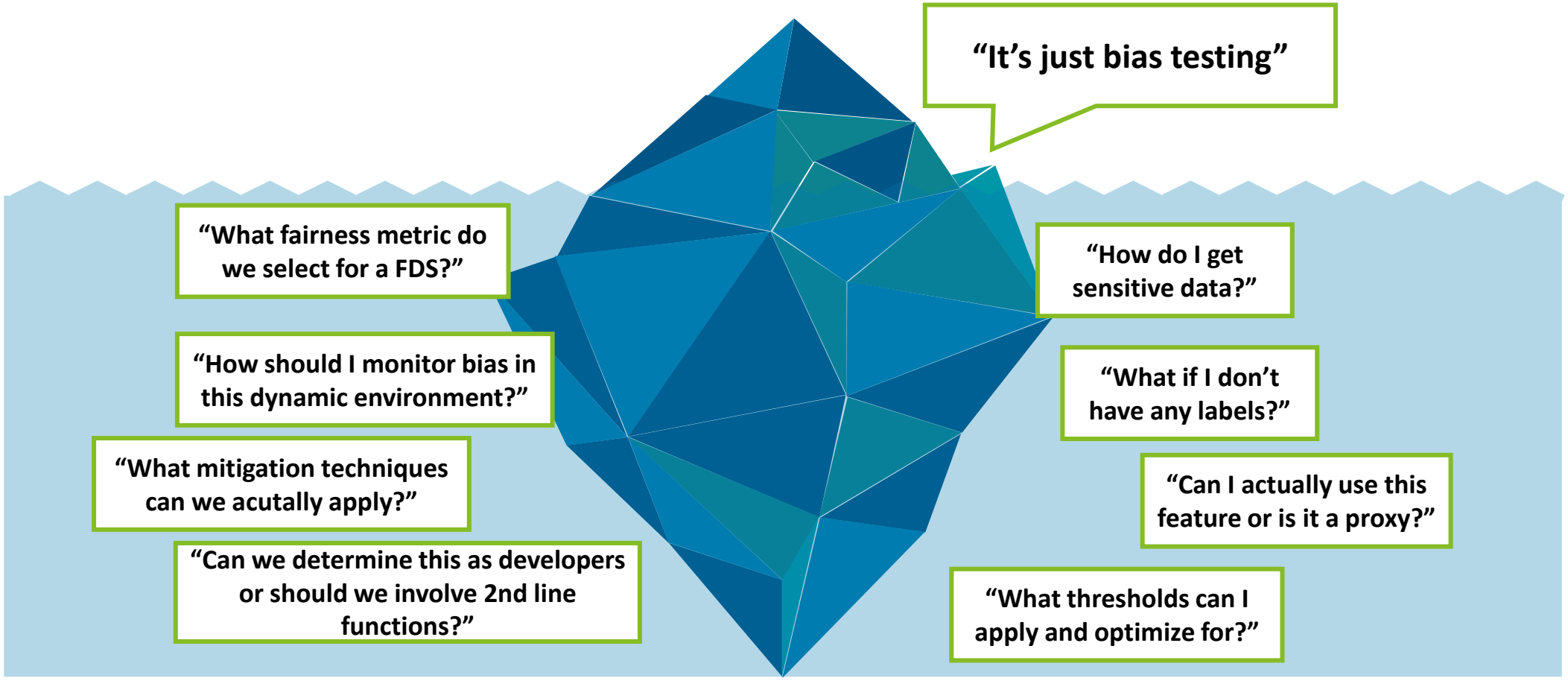
06/05/2026

AI-based fraud detection systems (FDS) and AI fairness are typically treated separately

The world of AI-based FDS

The world of AI fairness

Treating AI-based FDS and AI fairness separately hides key trade-offs that affect real-world outcomes



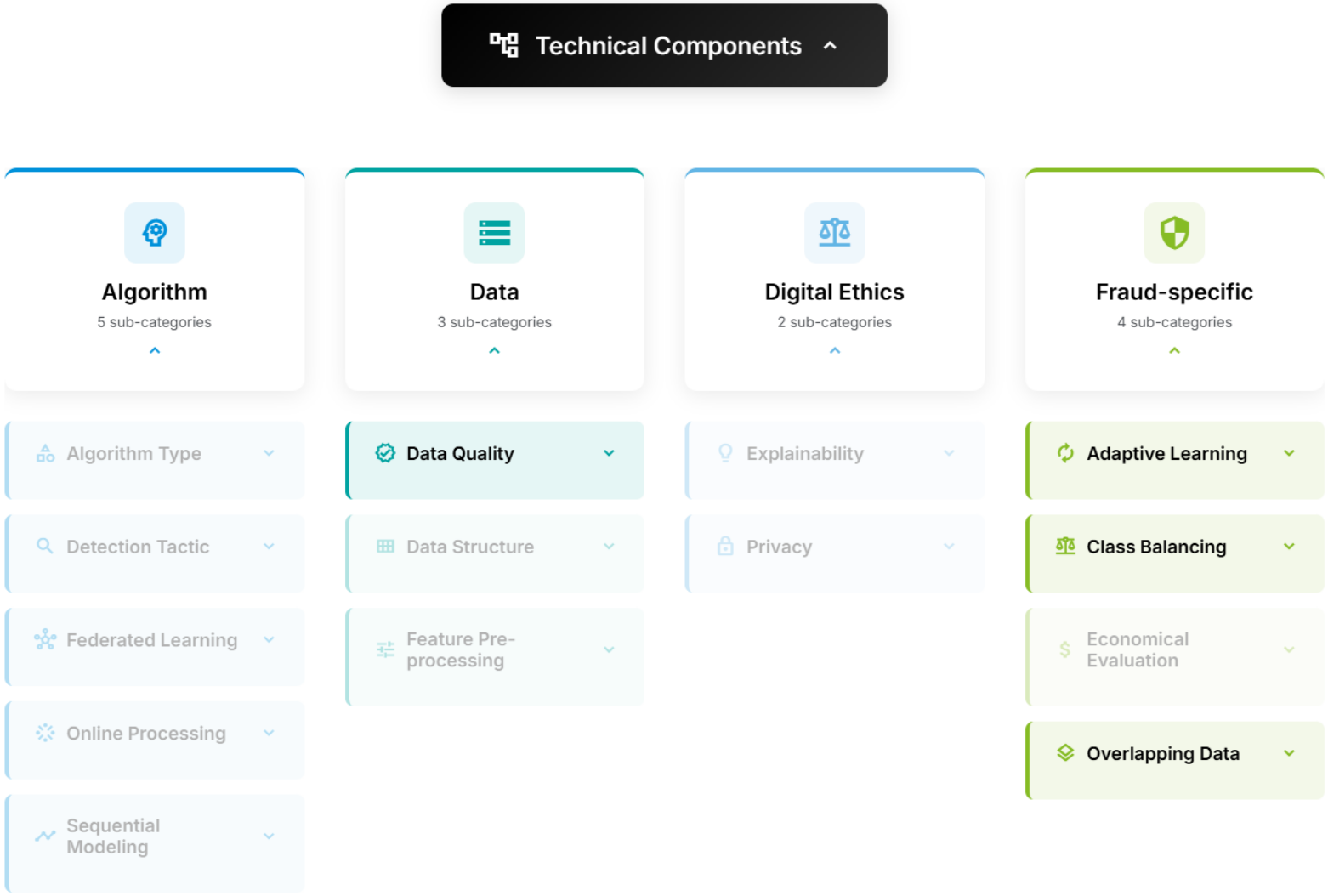
Can we identify how technical components of fraud detection interact with fairness?

1 What technical components of FDS can be identified in the literature, and categorized in a taxonomy?

2 What characteristics of fairness approaches can be identified in the literature, and categorized in a taxonomy?

3 What inter-dependencies can be identified between technical components and fairness approaches when the two taxonomies are applied?

Our proposed Taxonomy of Technical Components (TOTC)*



*TOTC is an iteration on an initial taxonomy proposed by Matschak et al. (2022) – A Taxonomy of Machine Learning Based Fraud Detection Systems

Four technical components directly relate to bias patterns found in AI-based FDS

Data Quality

Label bias: FDS often face challenges on label quality due to inaccurate and inconsistent (historical/selective) labeling processes.

Adaptive learning

Dynamic bias: Changing fraudulent behavior coincides with demographic behaviour.

Class balancing

Imbalanced group and class size bias: Class imbalance is aggregated demographic imbalances.

Overlapping data distributions

Separability bias: Differences in class separability and model features among demographic groups.

These biases impose (statistical) limitations on popular fairness metrics

Label bias

Fairness metrics rely on available and accurate labels, for example, to calculate false positive rate parity.

Dynamic bias

The efficacy of fairness mitigation methods can become affected.

Imbalanced group and class size bias

Brzezinski et al. (2024) showed fairness metrics are easier to satisfy in case of severe class imbalance.

Separability bias

No direct impact on fairness metrics.

Brzezinski et al. (2024) – Properties of Fairness Measures in the Context of Varying Class Imbalance and Protected Group Ratios

What’s next?

1. Let’s bring the theory to practice!
2. TOTC is based on theoretical objects. Can we also classify real-world systems using this method?
3. Our study showed inter-dependencies between technical components and fairness exist. How severe is the impact of these inter-dependencies on achieving fair and accurate AI-based systems?

Get in contact!

Sebastiaan Berendsen
Sberendsen@deloitte.nl

Deloitte.

MAKING AN IMPACT THAT MATTERS since 1845