

FEAT Fairness Self-Assessment Checklist

A practical tool for financial institutions assessing AI/AIDA systems against MAS FEAT Fairness Principles

This checklist maps directly to the four MAS FEAT Fairness Principles (Principles 1–4) and provides a structured framework for conducting an initial fairness self-assessment. It is designed to be completed by model risk, data analytics, or compliance teams as a first step toward FEAT alignment.

Each section includes specific checks, space for status tracking (Not Started / In Progress / Complete / N/A), and a notes column for documenting evidence and findings. Red flags are highlighted at the end of each section.

Use Case Being Assessed:	
Assessment Date:	
Assessor(s):	
Model / System Name:	
Business Owner:	
Risk Tier (Critical / Significant / Low):	

Note: FEAT Principles are voluntary guidance issued by MAS for Singapore-regulated financial institutions. This checklist focuses on the Fairness principles (1–4). Ethics, Accountability, and Transparency require separate assessment. This tool does not constitute legal or regulatory advice.

Section 1: Principle 1 — Systematic Disadvantage

"Individuals or groups of individuals are not systematically disadvantaged through AIDA-driven decisions, unless these decisions can be justified."

Note: This section uses rate-based screening as an initial diagnostic. Lending contexts may also require credit-policy overlays, marginal effects analysis, matched-pair testing, and reject-inference adjustments depending on jurisdiction and examination practice.

Check	Status	Evidence / Notes
<p>1.1 Identify all demographic groups potentially affected by this AIDA system. Distinguish between: (a) protected/personal attributes where available (e.g., gender, age, ethnicity, nationality), and (b) other relevant customer segments (e.g., low-income, rural, elderly).</p>	<p>■ Not started ■ In progress ■ Complete</p>	List groups identified:
<p>1.2 Calculate approval/selection rates for each identified group.</p>		Attach rate tables or reference document location.
<p>1.3 Compute Impact Ratios (selection rate of each group ÷ selection rate of reference group). Flag material disparities (e.g., IR < 0.80 or other institution-defined threshold) for investigation. Document threshold chosen and rationale.</p>		Reference group used and rationale (e.g., highest-rate group or largest eligible group):
<p>1.4 Conduct statistical significance testing (e.g., two-proportion z-test) for flagged groups. Report confidence intervals alongside p-values. Apply multiple-comparison corrections where testing multiple groups.</p>		Test method used:
<p>1.5 Run intersectional analysis (e.g., gender × age, ethnicity × gender) for groups with sufficient sample size (institution-defined minimum, e.g., ≥30; also consider privacy/k-anonymity thresholds). Document which intersections could not be tested and why; aggregate where needed.</p>		Intersections tested:
<p>1.6 For any group failing the four-fifths screen: document whether the disparity can be justified by legitimate business factors (e.g., genuine risk differences supported by evidence). Note: the four-fifths rule is a screening heuristic, not a legal threshold.</p>		Justification documented? Y / N
<p>1.7 Compare denial/rejection reasons across groups to identify process-level disparities (e.g., one group disproportionately denied for incomplete applications). Confirm denial-reason coding is consistent and completeness is comparable across groups before drawing conclusions.</p>		

Principle 1 Red Flags

Red Flag	What It Indicates	Immediate Action
IR below threshold (e.g., 0.80) for any group with adequate sample size	Potential adverse impact warranting investigation	Document finding; initiate root-cause analysis; escalate per governance thresholds

Red Flag	What It Indicates	Immediate Action
Intersectional failure hidden by single-attribute pass	Compounding disadvantage at group intersections	Expand testing to intersectional combinations; report separately
Denial reasons vary significantly across groups	Possible process-level bias (not just model-level)	Investigate operational/process differences in application handling

Section 2: Principle 2 — Justified Use of Personal Attributes

"Use of personal attributes as input factors for AIDA-driven decisions is justified."

■	Check	Status	Evidence / Notes
■	2.1 List all personal attributes used as direct model inputs (e.g., age, gender, marital status, nationality).		Attributes listed:
■	2.2 For each personal attribute used: document the business justification, necessity, alternatives considered, and governance approval.		Justification document reference:
■	2.3 Identify potential proxy variables: features that may correlate with protected attributes (e.g., geography, occupation, education). Assess proxy risk via correlation analysis and outcome sensitivity testing (e.g., feature removal, counterfactual analysis where feasible).		Proxy candidates identified:
■	2.4 Document whether any personal attributes or proxies were removed or constrained, and the impact on model performance and fairness metrics.		
■	2.5 If protected attributes are not available for testing, document which path is taken: (a) test on attributes you do have (gender, age, nationality) — a valid first-class approach; (b) test on geographic or segment proxies with sensitivity analysis; (c) if permitted by legal counsel and local law, consider inference methods (e.g., BISG, synthetic data). Inference may be restricted in your jurisdiction; do not proceed without legal approval.		Method used:

Principle 2 Red Flags

Red Flag	What It Indicates	Immediate Action
Personal attribute used without documented justification	Governance gap; potential regulatory finding	Document justification or remove attribute; obtain governance approval
High-correlation proxy variable in model with no sensitivity analysis	Model may be indirectly using protected information	Run feature removal / constraint tests; document fairness impact
No protected attribute data and no alternative testing approach	Unable to assess fairness at all	Implement proxy inference or test on available attributes; document coverage gaps

Section 3: Principle 3 — Regular Review and Validation

"Data and models used for AIDA-driven decisions are regularly reviewed and validated for accuracy and relevance, and to minimise unintentional bias."

■	Check	Status	Evidence / Notes
■	3.1 Confirm that model accuracy metrics (e.g., AUC, precision, recall) are calculated separately for each demographic group, not just overall.		Metrics by group documented? Y / N
■	3.2 Check for accuracy disparities: does the model perform significantly worse (higher false positive or false negative rate) for any demographic group?		Groups with accuracy gaps:
■	3.3 Review training data for representativeness: are all relevant demographic groups adequately represented? Document any underrepresented groups.		
■	3.4 Verify that a model validation schedule exists and includes fairness metrics (not just performance metrics).		Validation frequency:
■	3.5 Confirm that data quality checks include assessment of missing data patterns across demographic groups (e.g., is data more frequently missing for certain populations?).		
■	3.6 Document any fairness-performance trade-offs: if improving fairness for one group reduces overall accuracy, record the decision made, the rationale, and governance approval per your internal governance process.		Trade-off documented? Y / N

Principle 3 Red Flags

Red Flag	What It Indicates	Immediate Action
Model accuracy varies >10% across groups (e.g., FPR, FNR)	Model is less reliable for some populations	Investigate data quality, feature availability, and representation for affected groups
No fairness metrics in existing validation schedule	Fairness is not being monitored	Add fairness metrics to next validation cycle; document baseline
Training data significantly underrepresents a demographic group	Model may perform poorly for that group	Assess impact; consider data augmentation, reweighting, or model constraints

Section 4: Principle 4 — Outcome Monitoring

"AIDA-driven decisions are regularly reviewed to assess whether they have a systematically discriminatory impact on individuals or groups."

■	Check	Status	Evidence / Notes
■	4.1 Confirm that real-world decision outcomes (not just model scores) are tracked by demographic group over time.		Outcome tracking in place? Y / N
■	4.2 Establish monitoring thresholds: at what IR level or disparity magnitude does a finding trigger escalation? Document thresholds and escalation path.		Threshold defined:
■	4.3 Check for drift: have fairness metrics changed materially since the last assessment? Compare current period to baseline.		Drift detected? Y / N
■	4.4 Review whether human override patterns (where humans overrule the model) differ by demographic group. Identify whether overrides benefit or harm each group (e.g., approve-up vs reject-down).		
■	4.5 Confirm that a remediation process exists: if monitoring detects a new disparity, what happens next? Who is notified? What is the timeline for investigation?		Remediation process documented? Y / N
■	4.6 Confirm that customer appeal/review channels exist for AIDA-driven decisions (per Principle 10) and that appeal outcomes are tracked for demographic patterns.		

Principle 4 Red Flags

Red Flag	What It Indicates	Immediate Action
No ongoing outcome monitoring by demographic group	Cannot detect emerging disparities	Establish quarterly monitoring cadence with demographic breakdowns
Human overrides disproportionately affect one group	Possible bias in override decisions	Audit override patterns; provide decision-maker training
Fairness metrics have drifted materially since last assessment	Model or population characteristics may have changed	Trigger re-assessment; investigate root cause of drift

Assessment Summary

Use this page to record your overall assessment status across all four Fairness Principles.

Principle	Checks Complete	Red Flags Found	Overall Status	Next Action
1. Systematic Disadvantage	__ / 7	Y / N	■ Pass ■ Investigate ■ Fail	
2. Personal Attributes	__ / 5	Y / N	■ Pass ■ Investigate ■ Fail	
3. Review & Validation	__ / 6	Y / N	■ Pass ■ Investigate ■ Fail	
4. Outcome Monitoring	__ / 6	Y / N	■ Pass ■ Investigate ■ Fail	

Assessment completed by:	
Date:	
Reviewed by (governance):	
Governance sign-off date:	
Next assessment due:	

Evidence Pack

Retain the following artifacts alongside this completed checklist:

- Versioned dataset extract (with date and row count)
- Rate tables and Impact Ratio calculations
- Statistical test outputs (z-tests, CIs, corrections applied)
- Test script or spreadsheet (reproducibility)
- Governance sign-off records
- Remediation ticket links (if applicable)
- Model card with fairness section

What This Checklist Is Not

This checklist is a self-assessment tool — not a certification, audit opinion, or legal compliance determination. Completing it demonstrates due diligence, not that a system is free of bias. The four-fifths rule and other screening heuristics used here are flags for investigation, not legal thresholds. This checklist covers Fairness (Principles 1–4) only; Ethics, Accountability, and Transparency require separate assessment with broader organisational controls.

For independent FEAT assessments or help building internal assessment capacity, contact Minerva AI Assurance.

hello@trustminerva.com | trustminerva.com

Minerva Research is a publication by Minerva AI Assurance. This checklist is provided for informational purposes and does not constitute legal or regulatory advice. Financial institutions should consult their legal counsel and compliance teams when implementing AI governance frameworks. MAS FEAT Principles: mas.gov.sg/publications/monographs-or-information-paper/2018/feat. Veritas Toolkit: github.com/veritas-toolkit.