

Bias Mitigation and Model Review Policy

Version: 1.0

Last Updated: October 2025

Approved By: Chief AI Ethics and Compliance Officer (CAIECO)**

1. Purpose

This policy establishes procedures for identifying, evaluating, and mitigating bias in AI and machine learning models developed by Orbitel.ai.

It ensures fairness, inclusivity, and non-discrimination in AI-powered decision-making processes.

2. Scope

Applies to all:

- AI algorithms, data models, and NLP systems used in Orbitel's products.
- Data collection, preprocessing, labeling, and validation processes.

3. Governance and Responsibilities

- **AI Ethics Committee:** Oversees bias mitigation and approves high-risk model deployments.
- **Data Science Teams:** Perform regular bias testing and validation.
- **Compliance and Legal Teams:** Review outcomes against applicable anti-discrimination laws.

4. Policy Statements and Controls

a. Bias Detection and Analysis

- Datasets are assessed for representational balance across gender, ethnicity, geography, and language.
- Statistical and adversarial testing methods are used to detect systemic bias in models.

b. Bias Mitigation Techniques

- Rebalancing and reweighting of underrepresented groups in datasets.
- Differential privacy and data augmentation used to reduce demographic skew.
- Continuous retraining using feedback data to improve fairness metrics.

c. Review and Documentation

- Each model's bias assessment is documented before production release.
- A "Model Card" is maintained for every AI model, summarizing purpose, data sources, risks, and evaluation metrics.

d. Continuous Monitoring

- Model performance and fairness indicators are monitored throughout deployment.
- If bias or drift is detected, immediate retraining or rollback is initiated.

e. Stakeholder Engagement

Orbitel encourages feedback from customers and users regarding potential bias or inaccuracies in its AI systems. Reports can be submitted to aiethics@orbitel.ai.

5. Compliance and Transparency

Orbitel aligns this policy with the **EU AI Act**, **ISO/IEC 42001 (AI Management System)**, and **OECD Principles on Fairness and Transparency**.

Annual fairness reports summarize improvements and findings, published publicly as part of Orbitel's Trust Center.