NUT AGI FRAMEWORK: Risk Management File

Version: 1.0 | **Date:** January 30, 2026

1. Risk Identification & Assessment

Risk Category	Likelihood	Impact	Mitigation Strategy	Residual Risk	Owner (CEO Interims where applicable)
Bias in Outputs	Medium	**High**	• Dataset balancing via adversarial debiasing (λ=0.01) • Continuous fairness audits • Human review for high-stakes decisions • Third-party bias testing (planned Q4 2025)	Low	AI Ethics Lead
Cybersecurity Breach	Low	**Critical**	• AES-256 encryption (FIPS 140-3 roadmap) • TLS 1.3 for data in transit • Zero-trust architecture • Regular penetration testing • SHA-256 audit ledgers	Low	CISO
Hallucination / Inaccurate Outputs	Medium	**High**	• Neuro-symbolic integration (15% reduction vs GPT-4) • Symbolic logic constraints (Prolog rules) • 95% confidence threshold (SNP) • Mandatory human oversight for critical decisions	Medium	ML Engineering Lead
Privacy Violation / Data Leakage	Low	**Critical**	• GDPR/PIPL/CCPA compliance • Data minimization protocols • Anonymization/pseudonymization • Access controls (RBAC) • Regular DPIA reviews	Low	Data Protection Officer
Catastrophic Forgetting	Medium	Medium	• Elastic Weight Consolidation (EWC λ=0.1) • Gradient checkpointing (every 10^4 iterations) • <5% forgetting rate in pre-beta tests • Knowledge consolidation protocols	Low	ML Engineering Lead
Misuse by End Users	Medium	Medium	• Clear Terms of Service and Acceptable Use Policy • Abuse detection (Z-score > 3 anomaly detection) • User activity monitoring • Rate limiting and access controls • Content filtering for harmful requests	Low	Trust & Safety Lead

Scalability Failure (>2TB)	Medium	Medium	• Phased scaling (1TB → 5TB → 10TB) • Adaptive batch sizing (1024-4096) • Performance monitoring at scale • Optimization of O(n log n) memory operations • Infrastructure stress testing	Medium	Infrastructure Lead
Regulatory Non-Compliance	Low	**Critical**	• Dedicated compliance team • Regular regulatory monitoring • Pre-launch conformity assessment (EU AI Act) • Third-party audits • Transparent documentation (Annex IV/VIII) • Legal counsel review	Low	Chief Compliance Officer
Unintended ASI Emergence	Very Low	**Catastrophic**	• SNP human moderators (95% confidence override) • Ethical alignment constraints (EU AI Act, ISO 42001) • Kill-switch procedures • Bounded self-evolution (CEE with EWC) • Regular capability audits • No autonomous deployment without human approval	Very Low	CEO / AI Safety Board
Financial Market Manipulation	Low	**High**	• Policy layer with regex-based enforcement • No MNPI (Material Non-Public Information) trading • Market abuse detection (Z-score > 3) • Compliance with SEC/FINRA regulations • Audit trails for all financial decisions	Low	Financial Compliance Officer

2. Risk Monitoring and Review

Frequency: Quarterly risk reviews with monthly monitoring

Escalation Protocol: Medium risks escalate to Risk Committee; High/Critical risks escalate to Executive Leadership

Documentation: All risk assessments, mitigation actions, and incidents logged in centralized repository

3. Post-Market Surveillance

- Continuous monitoring of system performance and outputs
- User complaint tracking and analysis
- Incident response procedures with 24-hour escalation
- Regular bias audits and fairness assessments
- Performance drift detection and correction

_____
Interim Risk Management Owner
Chief Executive Officer, Nrutseab Ltd.