NUT AGI FRAMEWORK: Data Governance and Privacy Management

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

1. Data Sources and Collection

Training Data Sources:

- Public datasets: Common Crawl, Wikipedia, GitHub, arXiv, Stack Exchange
- Licensed commercial datasets: financial market data, enterprise corpora
- Synthetic data: 4M symbolic-natural pairs (Wikidata, ConceptNet, WordNet)
- Partner data: aggregated under NDA with consent verification
- User-generated content: with explicit consent and opt-out mechanisms

Operational Data:

- User queries and interactions (logged with pseudonymization)
- System performance metrics and telemetry
- Audit trails and compliance logs (SHA-256 hashed, immutable)

2. Data Processing Legal Basis

Data Type	Legal Basis	Compliance Measures
Public Training Data	Legitimate Interest	• Publicly available information • No personal data in training corpus • Bias mitigation protocols
User Queries (Personal Data)	Consent + Contract	• GDPR Art. 6(1)(a) consent obtained • Privacy Policy + ToS acceptance • Opt-out mechanisms available • Data minimization applied
Partner Data	Contract + Consent	• NDA agreements with all partners • Consent verification required • Data Processing Agreements (DPAs) • PIPL compliance for Chinese partners
Audit Logs	Legal Obligation	• Regulatory compliance requirement • Financial services audit trails (6-10 years) • Immutable SHA-256 ledgers

3. Data Protection Measures

3.1 Anonymization and Pseudonymization

- PII automatically detected and masked using regex patterns and NLP
- User identifiers replaced with cryptographic hashes (SHA-256)
- Location data aggregated to city/region level (not GPS coordinates)
- K-anonymity ($k \geq 5$) applied to aggregated analytics

3.2 Encryption

- Data at Rest: AES-256 encryption (FIPS 140-3 certification roadmap)
- Data in Transit: TLS 1.3 with perfect forward secrecy
- Key Management: Hardware Security Modules (HSMs) for key storage

3.3 Access Controls

- Role-Based Access Control (RBAC): admin, auditor, developer, user roles
- Zero-trust architecture with multi-factor authentication (MFA)
- Principle of least privilege enforced
- Access logs reviewed quarterly by security team

4. Bias Mitigation and Fairness

Training Phase:

- Adversarial debiasing with loss penalty ($\lambda=0.01$)
- Dataset balancing across protected attributes (gender, race, age)
- Pre-training bias assessment using AIF360 metrics

Operational Phase:

- Continuous fairness audits using embedding divergence metrics
- Third-party bias testing (planned Q4 2025)
- Disparate impact analysis for high-risk applications
- Human review for outputs affecting protected groups

5. Data Subject Rights (GDPR Compliance)

Right to Access: Users can request copies of their data via privacy portal

Right to Rectification: Users can correct inaccurate personal data

Right to Erasure: 'Right to be forgotten' with data deletion within 30 days

Right to Data Portability: Export data in machine-readable format (JSON/CSV)

Right to Object: Opt-out of data processing for direct marketing/profiling

Response Time: All requests responded to within 30 days (GDPR Art. 12)

6. Data Retention and Deletion

Data Category	Retention Period	Justification
User Queries (Personal)	30 days (anonymized after)	Service improvement + troubleshooting
Aggregated Analytics	Indefinite (anonymized)	Business intelligence, no personal data
Financial Audit Trails	7 years	Legal/regulatory obligation (SEC, FINRA)
Account Data	Until account deletion + 30 days	Contract performance + grace period
Training Data Snapshots	Model lifetime + 2 years	Model reproducibility, bias audits

7. International Data Transfers

EU to Third Countries:

- Standard Contractual Clauses (SCCs) for data transfers outside EEA
- Transfer Impact Assessments (TIAs) conducted per Schrems II ruling
- EU representative designated for GDPR compliance

China PIPL Compliance:

- Security assessment for cross-border data transfers
- Separate consent for international transfers
- China representative for data protection matters

8. Data Protection Impact Assessment (DPIA)

DPIA conducted for Nut AGI system per GDPR Art. 35 due to:

- Large-scale processing of personal data
- Automated decision-making with legal/significant effects
- Processing of financial and employment data

DPIA Review Cycle: Annual review or when significant changes occur

9. Data Breach Response

Detection: Real-time monitoring with automated alerts for anomalies

Notification Timeline:

- Supervisory Authority: Within 72 hours (GDPR Art. 33)
- Affected Users: Without undue delay if high risk (GDPR Art. 34)
- China CAC: Within required timeframes per PIPL

Breach Log: All incidents documented with date, nature, effects, and remedial actions

10. Governance and Accountability

Data Protection Office (DPO): Contactable at hello@nrutseab.com

Data Governance Committee: Quarterly meetings to review policies and incidents

Staff Training: Annual data protection and privacy training for all employees

Records of Processing: Maintained per GDPR Art. 30 and available for inspection

____

Data Protection Office
Nrutseab Ltd. | hello@nrutseab.com