DRN SUBMISSION ON DRAFT NRB AI GUIDELINES FOR BANKS & FINANCIAL INSTITUTIONS

Digital Rights Nepal

30 December 2025

DRN Submission on Draft NRB AI Guidelines for Banks & Financial Institutions

Introduction

Nepal Rastra Bank (NRB) has demonstrated commendable foresight with the issuance of the draft "Artificial Intelligence Guidelines," a proactive initiative announced in the Monetary Policy for fiscal year 2024/25. This measure establishes a foundational governance framework at an early stage of AI adoption, positioning Nepal to strategically balance technological innovation with financial stability and consumer protection.

The purpose of this document is to provide a formal, rights-based, and constructive analysis of these Guidelines. This review identifies specific areas for enhancement to ensure the deployment of artificial intelligence by licensed financial institutions is lawful, accountable, and aligned with Nepal's constitutional values and international best practices. It seeks to strengthen the Guidelines so they effectively govern AI systems that exercise real power over individuals' access to financial services and economic participation.

This review is structured to build upon the solid foundation of the draft. It will first acknowledge the strengths of the current text, then address several fundamental conceptual gaps that could undermine its practical effectiveness. Following this, the review will provide a detailed clause-wise analysis with specific textual and procedural suggestions. It concludes with broader ecosystem-level suggestions to foster a supportive environment for responsible AI innovation.

Positive Aspects of the Draft Guidelines

Recognizing the positive foundations of the draft Guidelines is strategically important, as these strengths provide a solid base upon which to build a more robust and rights-respecting framework. The document demonstrates a clear commitment to responsible governance, incorporating several principles that align with international best practices and are well-suited to Nepal's current level of AI maturity. Key strengths include:

- **Institutional Accountability at the Highest Level:** By placing ultimate responsibility for AI outcomes on the Board of Directors and senior management (Clause 5.1), the Guidelines ensure AI governance is treated as a core corporate responsibility rather than a siloed technical issue.

This top-level oversight creates powerful incentives for licensed institutions to take AI-related risks, ethics, and impacts seriously.

- **Adoption of a Risk-Based Approach:** The framework's adoption of a risk-based approach is a significant strength, appropriately balancing the need for innovation with the imperatives of financial stability and consumer protection. This methodology is well-aligned with Nepal's current AI maturity, allowing for proportionate regulation that does not unduly stifle experimentation with lower-risk applications.

- **Explicit Recognition of Individual Rights:** A noteworthy aspect of the Guidelines is the inclusion of risks to individual rights within the criteria for identifying "high-risk AI systems." Clause 6.1.1.d explicitly lists risks to privacy, fairness, non-discrimination, and equality, acknowledging that AI systems in finance can have profound effects on people's lives and therefore warrant heightened scrutiny.

- **Emphasis on Transparency and Data Protection:** The Guidelines demonstrate a commitment to transparency, explainability (Clause 7), and data protection. Provisions mandating compliance with Nepal's Privacy Act, adopting data minimization principles, requiring explicit consent for data use, and guaranteeing customer opt-out options (Clause 8) embed crucial consumer protection safeguards directly into the AI governance framework.

These foundational elements provide an excellent starting point. For the Guidelines to achieve their full potential, however, it is essential to address several underlying conceptual gaps to enhance their practical implementation and effectiveness.

Analysis of Gaps and Suggestions

a. The Need for an Explicit Rights-Based Framing

A critical enhancement would be to reframe AI governance from a purely institutional risk paradigm to one grounded in the protection of individual rights. The current draft frames AI primarily as a matter of operational, model, and systemic risk to licensed institutions. While Clause 6.1.1.d acknowledges risks to fundamental rights, this recognition is secondary rather than a guiding principle.

Clause 4 of the guideline mentions about the scope of the document which includes but is not limited to credit scoring, fraud detection, customer services, risk management. However, the Guidelines do not subsequently elaborate on how these identified use cases are to be governed, assessed, or regulated in practice. As a result, the scope of application remains insufficiently clear, particularly with respect to the mechanisms and obligations that apply to different types of AI systems. For instance, while credit scoring is explicitly mentioned within the scope, the Guidelines do not provide any follow-up provisions explaining how AI-driven credit scoring should be designed, deployed, assessed for risk, or monitored. There is no guidance on the potential risks associated with automated creditworthiness assessments, including risks of bias, discrimination, exclusion, or lack of explainability. Furthermore, the scope mentions terms such as “not limited to,” suggesting that the use of AI could be extended to services like credit denial, altering pricing, or restricting access to financial services. These are highly sensitive aspects of banking and finance, as they directly impact an individual’s economic liberty, dignity, and equality.

Under Nepal's legal framework, these are not merely private commercial decisions but acts of regulated decision-making with public law significance, demanding adherence to principles of administrative fairness. Without an explicit rights-based framing, subsequent provisions on fairness, transparency, and grievance handling lack sufficient legal and moral force.

Suggestion: The Guidelines should make the scope under Clause 4 operational by clearly specifying governance and compliance requirements for each listed AI use case, particularly customer-impacting systems such as credit denial, credit scoring. AI-driven sensitive/high risk financial decisions should be subject to clear risk assessment, bias mitigation, explainability, human oversight, and monitoring requirements to ensure consistent application and protection of customer rights. Similarly, Amend Clause 3 (Objectives) to explicitly state that AI systems used by licensed institutions are decision-making tools with direct implications for individual rights and must comply with principles of legality, necessity, and proportionality. This aligns with international best practices emphasizing the need to embed ethical and societal considerations into AI governance.

b. Absence of a Necessity and Proportionality Assessment

The Guidelines implicitly assume that once an institution decides to use AI, the regulatory task is simply to manage its risks. For the framework to be effective, it must incorporate a requirement for institutions to first justify *why* AI is necessary for a particular function, especially for high-risk systems (Clause 6.1.1). In a context where data quality may be uneven, excessive automation risks excluding individuals and populations who do not conform to historical data patterns, thereby exacerbating financial exclusion.

Suggestion: Require licensed institutions to conduct and document a "Necessity and Proportionality Assessment" for all high-risk AI systems. This assessment should explain why AI is being used over simpler or human-led processes and how perceived efficiency gains are balanced against the principles of fairness and inclusion.

c. Lack of Distinction Between Internal and Customer-Impacting AI Uses

Clause 4 (Scope) of the Guidelines does not draw a clear conceptual distinction between internal-support AI uses and customer-impacting systems. AI used for internal document processes presents a fundamentally different risk profile from an AI system used to approve loans or flag accounts for fraud. This ambiguity creates uncertainty in interpretation and enforcement, creating a risk that high-impact systems could be mistakenly treated as low-risk operational tools.

Suggestion: Amend Clause 4 to explicitly distinguish between internal-support systems and customer-impacting systems. The amendment should clarify that systems which directly or indirectly influence customer outcomes require heightened safeguards, ensuring that regulatory scrutiny is proportional to the system's potential impact on individuals.

d. Governance and Accountability

The provisions on outsourcing AI services (Clause 5.3) correctly identify the need for Board approval but lack specificity regarding the contractual safeguards required to protect sensitive customer data and prevent regulatory blind spots. There is a significant risk that sensitive data could be exposed to third-party vendors or that vendors could use institutional data to train their own proprietary models without explicit prohibition.

Suggestion: Amend Clause 5.3 to require that all outsourcing contracts for AI services include mandatory data anonymization, strict purpose limitations explicitly prohibiting vendors from using institutional data for their own model training, explicit audit rights for the licensed institution and NRB, and clear exit clauses to prevent vendor lock-in.

e. Risk Management (Clause 6)

The definition of "High-Risk AI Systems" (Clause 6.1.1) is a strong starting point, but the concept of "rights risk" is identified without being operationalized. Furthermore, high-risk systems are defined but do not mention low and medium risk. The mentioned framework would benefit from clearer definitions for low- and medium-risk categories to guide institutions in applying proportionate controls.

Suggestion: To operationalize "rights risk" and provide the clear classification framework sought by industry stakeholders, NRB should mandate a documented, pre-deployment "Rights Impact Assessment" (RIA) for all systems deemed potentially high-risk. The output of the RIA would serve as the primary justification for a system's final risk classification (high, medium, or low), thereby creating a direct, auditable link between potential human impact and the corresponding level of regulatory scrutiny.

Regarding Human Oversight, the Guidelines mention it as a risk-mitigation factor but do not sufficiently address the danger of "automation bias," where human staff uncritically defer to AI-generated outputs.

Suggestion: Amend Clause 6 to clarify that meaningful human oversight for high-risk systems includes the authority and responsibility to question, override, or suspend AI-driven decisions. This ensures that final legal and operational accountability for the decision remains with a designated human, not an algorithm.

f. Transparency and Explainability (Clause 7)

The requirement for AI systems to be explainable (Clause 7.1) is laudable but may be technically impractical if applied universally. Mandating strict explainability for all models, including complex systems like deep learning models used in low-risk applications, could stifle innovation without a corresponding benefit.

Suggestion: Link explainability requirements to the system's risk level. The obligation for full, technical explainability should be highest for high-risk decisions that directly impact customers. For low-risk applications, functional transparency (i.e., explaining what the system does and its limitations) should be sufficient. This proportional approach is crucial not only for managing different types of AI models but also for accommodating the diverse capacities of licensed institutions, from large commercial banks to smaller microfinance institutions.

g. Data Privacy and Protection (Clause 8)

The provision for consent withdrawal (**Clause 8.3**) presents a significant technical challenge, as retraining a complex AI model every time a single customer withdraws consent is often operationally infeasible. Additionally, the clause lacks specificity on data-use limitations, failing to define permissible data categories or prohibit the secondary use of sensitive financial data for AI training and profiling.

Suggestions:

- Clarify that customer opt-outs under Clause 8.3 apply prospectively, preventing their data from being used in *future model training or retraining cycles*, while acknowledging the technical infeasibility of removing an individual's data from an already-trained model's aggregated patterns.
- Strengthen Clause 8 by requiring licensed institutions to document the specific data categories used for each AI application, enforce strict purpose limitations, and prohibit or tightly control the repurposing of core banking data for secondary AI-driven analytics or profiling, in line with the requirements of Nepal's Privacy Act.

h. Customer Awareness and Grievance Handling (Clause 12)

The requirement to establish grievance mechanisms is essential, but Clause 12 lacks the procedural detail needed to be effective. It does not specify timelines for resolution, escalation paths, or the nature of remedies. This risks rendering the mechanism a formality with limited practical effect.

Suggestion: Amend Clause 12 to mandate clear, actionable standards for grievance handling. Customers who challenge an AI-influenced decision must be entitled to a response within a defined timeframe, a reasoned explanation

for the outcome, and, for high-impact decisions (e.g., credit denial), a review conducted by personnel with the authority to alter the outcome.

These clause-level enhancements would move the Guidelines from a statement of principles to a practical, enforceable regulatory framework. The next section addresses broader structural elements needed to support this framework.

i. Structural and Ecosystem-Level Suggestions

A robust AI governance framework requires more than just well-defined clauses; it necessitates a supportive regulatory and innovation ecosystem. For the Guidelines to be effective, they must be complemented by structural components that foster responsible AI adoption while ensuring a level playing field.

Suggestion:

- **Establish a Regulatory Sandbox:** To secure Nepal's competitive edge in financial innovation, the establishment of a formal regulatory sandbox is not merely beneficial but essential. Such a controlled environment de-risks experimentation for both incumbents and startups, accelerating the development of locally-relevant AI solutions that might otherwise be stifled by premature compliance burdens.
- **Promote Proportional Applicability:** A "one-size-fits-all" approach risks imposing a heavy compliance burden on smaller entities like Microfinance Institutions (MFIs), which have limited technical and financial resources. For the Guidelines to promote inclusive innovation, they must incorporate a tiered compliance system based on institutional size, risk profile, and the complexity of AI use-cases.
- **Implement a Defined Regulatory Approval Process:** To foster a predictable and encouraging environment for investment, the current ambiguity surrounding the approval process for new AI deployments must be resolved. A clear, time-bound, and staged pathway, such as notification, pilot approval, and full deployment, is a strategic necessity for effective business planning and maintaining innovation momentum.
- **Strengthen Data Sovereignty Mandates:** To safeguard national security and ensure robust regulatory oversight, the Guidelines must go beyond privacy principles and explicitly mandate the localization of critical financial and customer data processed by AI systems. This is vital to prevent sensitive national data from being subject to foreign laws and surveillance, thereby reinforcing Nepal's digital and economic sovereignty.

- **Foster the Domestic AI Ecosystem:** For Nepal to build sovereign capability and reduce dependence on foreign technology, the framework must actively promote Nepali AI developers and entrepreneurs. The current approach implicitly favors large international vendors; strategic measures are required to cultivate in-country expertise, create local economic opportunities, and ensure the development of AI solutions tailored to Nepal's unique context.

Incorporating these ecosystem-level components will ensure that the regulatory framework not only manages risk but also actively cultivates a vibrant, competitive, and responsible domestic AI industry.

g. Proposed Revisions to Annex A: Enabling Risk-Based Supervision

The current reporting template in Annex A serves as a basic inventory of AI systems but is insufficient for effective, risk-based supervision. To enable NRB to identify and prioritize risks, the annex must be transformed from a simple list into a dynamic supervisory tool that captures granular data on risk classification, customer impact, and model validation.

The following revisions to the Annex A reporting template are proposed to capture the necessary information for robust, risk-aware oversight:

Risk Classification: Mandate that for each AI system listed, the institution must declare its risk classification (e.g., low, medium, or high-risk) based on the criteria in Clause 6.1.1 and provide a brief justification.

Customer Impact Disclosure: Add a field requiring a clear description of how each system affects customers' rights or their access to services, including quantitative reporting on key outcomes like the number of AI-related service denials (e.g., loan rejections).

Model Validation Details: Require disclosure of who validated the model (i.e., an internal team or an independent third party) and the date of the last validation to provide insight into the independence and timeliness of risk management.

AI-Related Complaints: Include a new field to report the number and general nature of customer complaints received that are specifically related to each AI system, serving as a crucial feedback loop for identifying consumer harm.

These targeted changes will transform Annex A from a static inventory into a powerful diagnostic tool, allowing NRB to conduct more effective, data-driven, and risk-based supervision.

Conclusion

The Nepal Rastra Bank's draft "Artificial Intelligence Guidelines" represent a vital and forward-thinking first step toward governing the use of AI in Nepal's financial sector. However, for this framework to be truly effective, it requires significant strengthening to fully operationalize rights protections, ensure meaningful accountability, and foster a sustainable innovation ecosystem. While the Guidelines recognize individual rights, they do not yet translate this recognition into clear, decision-level obligations for financial institutions.

This review has put forward a series of constructive suggestions aimed at closing this gap. The key proposals center on: adopting an explicit rights-based framing; operationalizing risk management through mandatory Rights Impact Assessments; clarifying the non-negotiable role of human oversight; strengthening data protection with strict purpose limitations; and building a supportive ecosystem that encourages local innovation.

By incorporating these suggestions, Nepal Rastra Bank can create a regulatory framework that is not only technically sound but also socially legitimate and constitutionally grounded. Such a framework will foster responsible innovation, enhance efficiency, and promote financial inclusion while simultaneously upholding the fundamental principles of fairness, dignity, and public trust in Nepal’s rapidly evolving financial sector.

Bibliography

Nepal Rastra Bank, *Artificial Intelligence Guidelines* (Banks & Financial Institutions Regulation Department, 2025).

Irving Fisher Committee on Central Bank Statistics, *Governance and implementation of artificial intelligence in central banks*, IFC Report No. 18 (Bank for International Settlements, April 2025).

Government of Nepal, *Privacy Act, 2075* (2018).