BAHAMAS BUREAU OF STANDARDS AND QUALITY

QUALITY IS OUR STANDARD

DRAFT BAHAMAS NATIONAL STANDARD

RISK MITIGATION GUIDELINES FOR HIGH-RISK AI APPLICATIONS

DBNS 201: 2025

Bahamas Bureau of Standards & Quality (BBSQ)
Source River Centre, 1000 Bacardi Road
P.O. Box N- 4843, Nassau, New Providence, Bahamas
Tel: (242) 362-1748 thru 56
Fax: (242) 362-9172
Email: standards@bbsq.bs
Website: www.bbsq.bs

Website: www.bbsq.bs

ISBN 978-976-8268-13-6

NOTICE

Standards are subjected to periodic review.

The next amendment will be sent without charge if you return the self-addressed label. If we do not receive this label we have no record that you wish to be kept up-to-date. Please note amendments are not exclusive of a revision of the document.

Our address:
Bahamas Bureau of Standards & Quality (BBSQ)
Source River Centre
1000 Bacardi Road
P.O. Box N- 4843
Nassau, New Providence
Bahamas

---------- (□cut along the perforated line□) -

BNS ISO/IEC 22989

NAME: _____

_____

_____

COMPANY/DESIGNATION:

_____

_____

ADDRESS:

_____

_____

_____

AMENDMENTS ISSUED SINCE PUBLICATION

AMENDMENT NO.	DATE OF ISSUE	TYPE OF AMENDMENT	NO. OF TEXT AFFECTED	TEXT OF AMENDMENT

ATTACHMENT PAGE FOR BNS AMENDMENT SHEETS

BBSQ Foreword

This national standard written by The Bahamas' Bureau of Standards and Quality's Artificial Intelligence (AI) Technical Committee member Takoda C. Kemp. The national committee responsible for reviewing this standard is Technical Committee 20 *Artificial Intelligence*. This standard contains requirements that are relevant for The Bahamas.

BBSQ Committee Representation

This ISO International Standard was adopted as a National Standard under the supervision of the National Technical Committee for Conformity Assessment (NTC 20) hosted by the Bahamas Bureau of Standards and Quality which at the time comprised the following members:

Member	Representing
Dr. James Carmichael (**Chairman**)	University of The Bahamas
Mr. Onassis Nottage (**Vice Chairman**)	National Expert
Mr. Joseph Pinder (**Technical Secretary**)	National Expert
Mr. Takoda C. Kemp	Bahamas National Geographic Information Systems Centre
Ms. Mona Nixon	Bahamas Agricultural Industrial Organization
Mr. Andrew Dean	Higgs and Johnson
Mr. Peter Bethell	Treasury Department
Ms. Alphanique Swann-Newbold	Atlantis
Dr. Ancilleno Davis	University of the Bahamas

Contents

BDNS 201: AI Standards – Risk-Mitigation Guidelines for High-Risk AI Applications

1. Guiding Principles

Artificial intelligence (AI) systems should be designed, deployed, and maintained in ways that promote safety, reliability, transparency, and respect for human rights and dignity. AI should support, rather than undermine, social trust, accountability, and the integrity of democratic and community processes.

2. Scope

This document provides guidance to consumers on using artificial intelligence (AI) models to facilitate task completion and workflow management; risk mitigation procedures in the context of the development/deployment of AI tools in high-risk scenarios is the focus of this document. This is also meant to be a supplementary document to BNS ISO 23894 - Information technology — Artificial intelligence — Guidance on risk management.

3. High-Risk Applications

The following classes of AI applications are considered high-risk regarding their development and deployment should be carefully supervised:

- Systems that create or apply social scoring[1] or ranking[2] of individuals.
- Large-scale or indiscriminate biometric identification or surveillance in public or publicly accessible environments.
- The production of synthetic material, which is factually incorrect, and misleading and deliberately meant to deceive/impersonate some individual without due authorisation.
- Emotion-recognition or affect-inference systems used to determine eligibility for the provision of social services including – but not limited to – employment, education, financial services, public services, or other essential opportunities.

Certain domains involve significant potential impacts on individuals, institutions, and critical systems. AI used in these areas should only be deployed under enhanced governance, testing, and ongoing oversight. Examples include, but are not limited to AI systems which manage:

[1] *Social scoring* refers to systems that classify individuals based on their social behaviours or personal characteristics into categories such as "good" and "bad." The scores returned by these systems may then be used as a basis for decisions to distribute benefits to some groups and deny such benefits to other groups.

[2] *Social ranking* refers to the organization of individuals or groups into a hierarchy within a society. It signifies an individual's place within the societal structure, influencing how they are perceived and interact with others. [from: https://biologyinsights.com/what-is-social-ranking-and-how-does-it-work/]

- Public safety, security, and any form of adjudication (e.g. determining guilt or liability in a court of law).
- Employment-related decision systems.
- Financial services and financial risk assessment.
- Healthcare, social support, or benefits eligibility.
- Delivery of educational services, including academic assessment and proctoring.
- Management of Critical infrastructure management and transportation systems.
- Border activities, travel, and identity-verification functions.
- Environmental and Natural Resources.

Suggested safeguards for high-risk use include:

- The use of closed-source Large Language Models [3] and environments for sensitive government tasks.
- Risk assessments addressing issues of safety, privacy, bias and potential misuse.
- Independent evaluation of system robustness, accuracy, and fairness.
- Publicly available high-level descriptions of system purpose, capabilities, and limitations.
- Assurance that any automated AI decision-making in high-risk activities will be subject to human review for the purpose of validation.
- The implementation of mechanisms for logging, auditing, and monitoring of high-risk AI applications once they have been put into service.

4. General-Purpose AI Models

A policy of full disclosure and transparency should be adopted by system developers / providers of large or general-purpose AI models and such documentation should disclose central issues including safety practices, alignment approaches, risk-mitigation strategies, and data-governance principles. It is advisable that these general-purpose AI Models incorporate provenance mechanisms such as watermarking or structured content-origin metadata.

Public-sector and institutional users are encouraged to adopt models that meet recognised international standards for safety, security, privacy, and content integrity.

[3] A *closed source large language model* is one where the training data, weights and biases – as well as the hosting and serving of the application – are proprietary. In closed source LLMs, the behaviour and output of the LLM are not accessible or modifiable by a third party.

5. Platforms and Content Integrity

Online platforms and digital service providers should implement measures to support the authenticity, reliability, and traceability of AI-generated content, particularly in contexts that influence public discourse. These measures may include:

- Clear labelling or signalling of AI-generated political or civic-relevant content during sensitive periods such as elections.
- Publicly accessible archives of political and issue-based advertising that includes AI-generated content.
- Swift mitigation or removal processes for malicious synthetic media, including deceptive impersonations of public figures or community leaders.
- The provision of flagger channels[4] to allow users to publish their experiences during interaction with AI applications.

6. Secure AI Environments for Sensitive Data

Organisations handling sensitive, confidential, or high-value information are encouraged to employ secure, controlled AI environments that minimise data-exposure risks. As part of this approach:

- Governments and large institutions may adopt closed-source or locally hosted language models to support tasks involving protected data, internal workflows, or restricted information.
- Such models should operate within isolated, privacy-preserving environments with clearly defined access controls and auditing mechanisms.
- The deployment of these systems should follow recognised best practices for information security, data minimisation[5], and lifecycle governance.
- Where possible, providers should demonstrate that these models meet international standards for secure development, risk management, and content integrity.

This approach supports stronger data-protection guarantees, reduces reliance on external infrastructure for sensitive operations, and ensures appropriate oversight of AI systems used in critical institutional contexts.

[4] A flagger channel is a clearly defined pathway for organizations or individuals to report negative experiences with an AI system under review.

[5] Data minimisation is a privacy and security principle which states that organisations should only collect, store, and use the smallest amount of personally identifiable information necessary to complete a specific, and legitimate task.