

MACHINE LEARNING

Before the Model Goes Live: A Privacy Risk and Data Ethics Checkpoint Framework for Deploying Machine Learning in Nigerian and UK Financial Institutions

Onyinye Lydia Abuo

Published October 2026

Abstract

In Nigerian financial institutions, and across the UK, machine learning (ML) is already part of the fabric of credit applications, fraud detection and customer onboarding processes, but privacy and ethical concerns are only tackled at the back-end, inconsistently or once the machine learning solution is deployed. There is no existing guidance that is consistent or organized around a single decision point that a model should pass before it is allowed to go live. This conceptual review attempts to fill that gap by bringing together literature and regulatory sources in both jurisdictions. It establishes the concept of privacy risk, data ethics, model deployment governance and model risk management, puts the analysis in the context of privacy by design, contextual integrity, accountability and fairness and compares the Nigerian and UK regulatory frameworks. Then it highlights the key privacy and ethical issues that occur during deployment. Based on this, the paper suggests the Pre-Deployment Privacy and Ethics Checkpoint Framework, a gated framework of six stages: data and purpose validation; privacy impact and lawful basis check; fairness and bias testing; explainability and human oversight; security and third-party review; and go-live approval with monitoring hand-off. There is an evidence requirement and a pass/fail gate for each stage, with a named owner for each stage. The paper provides a practical, jurisdictionally flexible structure which turns dispersed obligations into a deployable and auditable trail, and it sets out limitations and testing requirements to move to the point where the framework could be considered valid.

Keywords: machine learning; financial institutions; privacy by design; data ethics; model governance; fairness and bias; regulatory compliance

1 Introduction

Machine learning has been quickly embraced by financial institutions. The [Bank of England and Financial Conduct Authority \(2019\)](#) reported that a significant portion of UK financial firms were already implementing or developing an ML application, with most of these being in fraud detection or credit, and customer engagement. Digital lending, transaction monitoring and remote onboarding in Nigerian banks and fintechs rely on the same principle: the automation of decisions based on huge volumes of personal and behavioural data.

These benefits come with risks. AI models based on their own data can generate sensitive traits, and can predict, perpetuate, or make decisions that customers may not be able to easily interpret or challenge ([Barocas & Selbst, 2016](#); [Mittelstadt et al., 2016](#)). The majority of these risks manifest themselves when the model moves from a controlled development environment to live decisions made for real customers. Regulators have reacted by enacting data protection legislation and issuing

supervisory statements on the risks associated with models and guidelines on artificial intelligence (PRA, 2023; Nigeria Data Protection Act, 2023; Information Commissioner's Office [ICO], 2020).

However, this guidance is disjointed. There are data protection rules, prudential rules, conduct rules, and ethics guidelines (which are largely optional and principle-based) (Jobin et al., 2019). No one instrument asks an institution to demonstrate at one point in time that a model is lawful, fair, explainable and secure prior to going live. Algorithmic accountability has been called for in the research literature in just this sense, namely through end-to-end, stage-based accountability (Raji et al., 2020). The gap is even wider in institutions that have operations both in Nigeria and the UK because they have to be able to balance two systems based on the same basic principles, but with varying levels of development, supervisory approach and enforcement power.

This paper aims to fill this gap by formulating a conceptual framework. It provides a six-stage checkpoint model aligned to the requirements in Nigeria and the UK, which institutions can use as a pre-deployment gate.

2 Conceptual foundations

Privacy risk is the degree of risk to the privacy of the individual that may result from the processing of his or her personal data, which may involve an intrusion upon his or her personal autonomy, a prejudice or a disclosure of the data. It is not the same as organisational risk, as the subject of the risk is the data subject and not the organisation (ICO, 2020). Data ethics goes beyond legal issues to moral questions that arise during the creation, analysis and utilization of data, such as fairness, autonomy and accountability (Floridi & Taddeo, 2016). Model risk management is the identification, validation, monitoring and governance of the risk of the model producing incorrect or misused outputs (Board of Governors of the Federal Reserve System and Office of the Comptroller of the Currency, 2011; PRA, 2023). In this paper we define model deployment governance as decision rights, evidence and controls that govern the entry of a model into production.

The analysis is based on four principles. Privacy by design is a philosophy that considers privacy from the beginning of the system design process and makes it a key element of its architecture (Cavoukian, 2011). Contextual integrity theory suggests that information flows contain norms that are related to their context, and therefore a flow can be considered an invasion of privacy even if the data is legally recorded (Nissenbaum, 2004). Accountability entails that the responsibility for the outcome of the algorithm can be attributed and traced, which relies on documentation and traceable decision-making processes (Kroll et al., 2017; Mitchell et al., 2019). Fairness requires that models should not produce unjustified disparate outcomes for protected and/or vulnerable groups (Mehrabian et al., 2021). All of these principles imply that a deployment gate should be based, not on technical performance, but on purpose, information flows, responsibility, and outcomes.

3 Regulatory and governance landscape: Nigeria and the UK

The Nigeria Data Protection Act (NDPA) 2023 is the main legislation in Nigeria. It introduces the concept of the Nigeria Data Protection Commission (NDPC) as regulator, introduces infringing powers for the processing, and sets out the rights of data subjects and the duties of data controllers and data processors and requires an impact assessment when processing is likely to cause a high risk to the data subject (Nigeria Data Protection Act, 2023). It is based on the previous Nigeria Data Protection Regulation (NDPR) (National Information Technology Development Agency [NITDA],

2019) and complemented with the NDPC GAID implementing directives. There are expectations on information security governance (Central Bank of Nigeria, 2018) that are imposed on the sector by the Central Bank of Nigeria (CBN) including the risk-based cybersecurity framework. The rules of the game in Nigeria are few and light, and leave the supervision of ML to general data protection duties and internal risk policies.

The UK GDPR and the Data Protection Act 2018 regulate the processing of personal data in the UK, such as the restrictions on making decisions based exclusively on automated means with a legal or similar effect, and the need for an impact assessment of the processing of personal data by organisations if they are carrying out high-risk processing (Data Protection Act, 2018; General Data Protection Regulation, 2016). The ICO's guidance on AI and data protection is an elaboration of these responsibilities into more concrete expectations on fairness, transparency, accuracy and security (ICO, 2020). The Prudential Regulation Authority has published model risk management principles for banks (PRA, 2023) and the Financial Conduct Authority has published the Consumer Duty, which relates directly to outcomes for retail customers, which is relevant to algorithmic fairness (Financial Conduct Authority, 2022).

The regimes have similar key principles – principles of lawfulness, purpose limitation, data minimization, security, accountability and the use of impact assessments. There are three ways in which they differ. The first is that while there are well established prudential model risk regimes in the UK, there are none in Nigeria. Second, the UK has a more mature pool of guidance from regulators relating to the concepts of AI, explainability and fairness. Thirdly, the capacity and maturity of enforcement varies and this impacts the enforcement of obligations. The cross-border institution gets more detailed technical benchmarks from the UK regime and the Nigerian regime imposes binding local obligations that must be fulfilled in isolation. The common principles can then serve as the backbone of the single checkpoint approach, with the different needs as "evidence tests" for each jurisdiction.

4 Privacy risks and ethical risks in ML deployment

In the literature, there are six classes of risks which are repeatedly mentioned. The first one is about "lawful basis" and "lawful purpose". The purposes of data collection by ML projects are often concealed within the data being collected, for example, credit scores are derived from transaction histories, or mobile data is used to assess creditworthiness, which can allow data collection to go beyond the scope of the original purpose and to be used in ways that the data subject might not have expected (ICO, 2020; Nissenbaum, 2004). The second is that of the minimisation of data. Models tend to be optimised for more data and more features; this is counter to the principle of processing only required data as the alternative data sources can also provide inferences about sensitive attributes.

The third one is automated decision making. If decisions are taken on credit and access to accounts without any meaningful human interaction, then data protection law states that there must be safeguards – being able to obtain a human review of the decisions and have the ability to challenge the decision (Goodman & Flaxman, 2017). The fourth one is explainability. There are partial answers for complex models to be accurate yet not explainable (such as, counterfactual explanations and post hoc attribution) and institutions should still be able to explain outcomes to customers, validators and regulators (Bracke et al., 2019; Wachter et al., 2018). The fifth is bias and discrimination. Historical data can carry over past exclusions and proxy variables can code protected characteristics without naming, having disparate impact on certain groups (Barocas & Selbst, 2016; Mehrabi et al., 2021). In

some cases, the models may come to rely more heavily on such proxies (thin file) when credit histories are thin, exacerbating the issue.

The last one is third-party data flows and vendor data flows. Cloud platforms, data brokers and model vendors often are used by institutions, which increases the number of parties involved in managing the personal data, and makes it harder to keep track of accountability, security and cross-border transfers. These risks involve legal, technical and ethical areas which are often owned by different teams, and so these risks can be easily overlooked when reviewing them individually. This is why it is important to have one point in time to test and document each risk before deployment.

5 The proposed checkpoint framework

The Pre-Deployment Privacy and Ethics Checkpoint Framework (PPECF) is a set of six sequential gates through which one passes on the way to deployment. Any model failure will bring the model back to development with remediation noted, until the model passes. The accountability logic of internal algorithmic auditing (Raji et al., 2020) is reflected in the fact that each stage is assigned a named owner, evidence requirement and a pass or fail criterion enabling the tracking and audit of approval. The flow is depicted in Figure 1, and the owner, evidence and gate of each stage are given in Table 1.

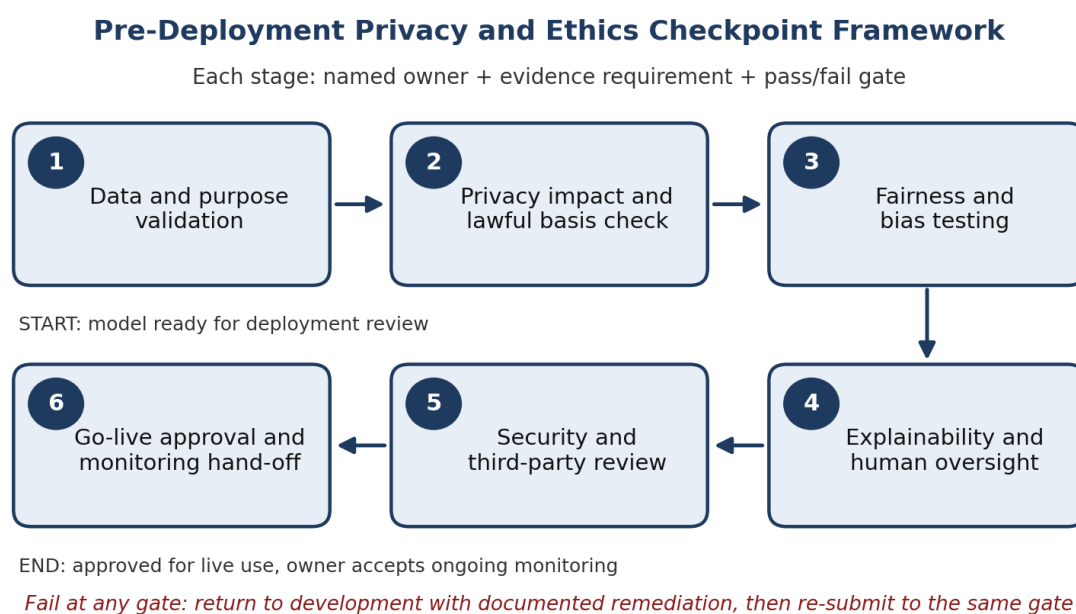


Figure 1. The Pre-Deployment Privacy and Ethics Checkpoint Framework.

Table 1. Owner, evidence requirement and gate for each checkpoint stage.

Stage	Owner	Evidence requirement	Pass/fail gate
1. Data and purpose validation	Data owner and product lead	Data inventory, documented purpose, feature minimisation review	Purpose documented and each feature justified
2. Privacy impact and lawful basis check	Data protection officer	Impact assessment, lawful basis record, automated decision safeguards	Assessment approved and lawful basis confirmed

Stage	Owner	Evidence requirement	Pass/fail gate
3. Fairness and bias testing	Model validation lead	Subgroup performance and proxy analysis, model documentation	Disparities within agreed thresholds or justified
4. Explainability and human oversight	Model owner and business lead	Explanation method tests, human review and appeal route	Decisions explainable and oversight route operational
5. Security and third-party review	Information security and vendor risk leads	Security testing, vendor due diligence, transfer assessment	Controls verified and vendor terms approved
6. Go-live approval and monitoring hand-off	Model risk committee and senior accountable executive	Sign-off pack, monitoring plan, rollback criteria	All prior gates passed and monitoring owner accepts

5.1 Framework explanation

The stages are arranged to have earlier gates that "block" later gates. It is important to implement data and purpose validation before anything else can be validated later because a model that was created with unsuitable data cannot be repaired later, in line with the 'privacy by design' principle (Cavoukian, 2011). The purpose is converted to legal position in the privacy impact and lawful basis check and this is where the duties of the impact assessment listed in the NDPA and the UK GDPR are fulfilled (Nigeria Data Protection Act, 2023; General Data Protection Regulation, 2016). Fairness testing and explainability come next, as both require a stable model and dataset. Security and third-party review is delayed due to vendor choice and infrastructure design decisions often not being finalized until the model design is finalized. The final gate brings together all the evidence in a coherent decision and passes on the model to the continued monitoring, as per model risk management expectations (PRA, 2023).

The Nigerian and the UK requirements align with the same six gates, and provide different evidence. Both NDPC and ICO expectations are included in the impact assessment at stage two, and additional safeguards of the UK rules on automated decisions are added to the impact assessment at stage two. UK institutions would have an outcome-based fairness test (Financial Conduct Authority, 2022) at stage three while Nigerian institutions might have to look to the principles of data protection and internal policy. The CBN cyber security framework which provides local security expectations (Central Bank of Nigeria, 2018) will be covered in stage five. The need for local adaptation is therefore primarily in the level of gates (not their structure), their form of evidence, and engagement of the regulators.

6 Challenges and limitations

There are four constraints to the framework. Regulatory asymmetry is created when the Nigerian and UK regimes do not have as much detail in benchmarking, thus some of the thresholds will have to be determined by institutional judgement. Lack of skills, tooling and data protection expertise could make things more difficult to operate the gates, especially in smaller fintech businesses. Cost and delay are also issues because if not scaled to the risk of each model, six sequential gates could delay releases. Finally, the framework is an abstract one that has not been empirically tested. Its stage design, owners and thresholds are based on synthesized literature and regulatory sources and changes to regulations in either jurisdiction can impact what is mapped.

7 Recommendations

The checkpoint should be used as a formal deployment gate at the institution; the depth of the checkpoint should be proportionate to the model risk tier; and named owners should be assigned to each checkpoint with documented sign-off. They should retain the evidence in one model record which will enable the internal audit process and regulators to follow decisions. Regulators might facilitate adoption by making expectations clear regarding impact assessments for ML, and through discussions with their supervisors about the evidence that will be shared before deployment, as well as explainability and fairness testing. In general, the NDPC would benefit from more model specific guidance; and in particular, more UK experience. Further studies should validate and pilot the framework in the banks and fintech companies of both countries and conduct the deployment of the framework with case studies, as well as refine the thresholds in each gate and explore the cost and effect on the time to deployment.

8 Conclusion

While privacy and ethical issues focus on the deployment of ML, they are subject to multiple rules that are spread out. This paper examined the conceptual research on the subject, the regulations and risks in Nigeria and the UK, suggesting a six-stage audit framework that converts all the obligations into a single auditable pre-deployment routine. It primarily has a structural aspect, tying together common principles to gated decisions with owners, evidence and pass or fail measures, but also allowing for jurisdiction-specific requirements. The framework has not yet been tested and the utility of it will be determined by the application of it. If it is validated, it would assist institutions to convince their clients and regulators that a model was legal, fair, explainable and secure prior to deployment.

References

- [1] Bank of England and Financial Conduct Authority (2019). Machine learning in UK financial services. *London: Bank of England*.
- [2] Barocas S, Selbst AD (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732. <https://doi.org/10.15779/Z38BG31>
- [3] Board of Governors of the Federal Reserve System and Office of the Comptroller of the Currency (2011). Supervisory guidance on model risk management (SR 11-7). *Washington, DC: Federal Reserve*.
- [4] Bracke P, Datta A, Jung C, Sen S (2019). Machine learning explainability in finance: An application to default risk analysis (Staff Working Paper No. 816). Bank of England.
- [5] Cavoukian A (2011). Privacy by design: The 7 foundational principles. *Information and Privacy Commissioner of Ontario*.
- [6] Central Bank of Nigeria (2018). Risk-based cybersecurity framework and guidelines for deposit money banks and payment service providers. *Central Bank of Nigeria*.
- [7] Data Protection Act (2018). Data Protection Act 2018, c. 12. <https://www.legislation.gov.uk/ukpga/2018/12/contents>
- [8] Financial Conduct Authority (2022). A new Consumer Duty (Policy Statement PS22/9). *Financial Conduct Authority*.
- [9] Floridi L, Taddeo M (2016). What is data ethics? *Philosophical Transactions of the Royal Society A*, 374(2083), 20160360. <https://doi.org/10.1098/rsta.2016.0360>
- [10] General Data Protection Regulation (2016). Regulation (EU) 2016/679 (General Data Protection Regulation), as retained in UK law.

-
- [11] Goodman B, Flaxman S (2017). European Union regulations on algorithmic decision-making and a "right to explanation." *AI Magazine*, 38(3), 50–57. <https://doi.org/10.1609/aimag.v38i3.2741>
 - [12] Information Commissioner's Office (2020). Guidance on AI and data protection. *Information Commissioner's Office*.
 - [13] Jobin A, Ienca M, Vayena E (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
 - [14] Kroll JA, Huey J, Barocas S, Felten EW, Reidenberg JR, Robinson DG, Yu H (2017). Accountable algorithms. *University of Pennsylvania Law Review*, 165(3), 633–705.
 - [15] Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 115. <https://doi.org/10.1145/3457607>
 - [16] Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, Spitzer E, Raji ID, Gebru T (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220–229). ACM. <https://doi.org/10.1145/3287560.3287596>
 - [17] Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716679679>
 - [18] National Information Technology Development Agency (2019). Nigeria Data Protection Regulation 2019. *NITDA*.
 - [19] Nigeria Data Protection Act (2023). Nigeria Data Protection Act, 2023. *Abuja: Federal Republic of Nigeria*.
 - [20] Nissenbaum H (2004). Privacy as contextual integrity. *Washington Law Review*, 79(1), 119–157.
 - [21] Prudential Regulation Authority (2023). Model risk management principles for banks (Supervisory Statement SS1/23). *Bank of England*.
 - [22] Raji ID, Smart A, White RN, Mitchell M, Gebru T, Hutchinson B, Smith-Loud J, Theron D, Barnes P (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44). ACM. <https://doi.org/10.1145/351095.3372873>
 - [23] Wachter S, Mittelstadt B, Russell C (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31(2), 841–887.