RESPONSIBLE AND TRUSTWORTHY AI IN DIGITAL LENDING: FROM PRINCIPLES TO PRACTICES?

AUTHORS

Dvara Research Team: Anubhutie Singh, Beni Chugh & Manvi Khanna

PwC Team: Neeraj Sibal, Nikita George & Vivek Belgavi

This document is currently under review. An updated version will be published after incorporating relevant feedback

DVARA Research

Responsible and Trustworthy AI in Digital Lending: From Principles to Practices[1]

[1] The Whitepaper at present does not delve into the question of *what is AI*? Across frameworks, there is no consensus or universally accepted definition of AI. There is abundant literature answering the question of what AI is. However, it would be crucial to understand what AI/ AI system/AI ecosystem are as well as kinds of AI to understand the perimeters/scope of our work as well as the context (here: digital credit journey). The select papers/frameworks/regulatory updates considered for the present exercise includes those by Economic Advisory Council- PM, World Economic Forum (4), International Monetary Fund (1), G20 (1), Organisation for Economic Co-operation and Development (1), United Nations Educational, Scientific and Cultural Organization i.e. UNESCO (1), Global Partnership for AI (1), Niti Aayog (3), The Dialogues (1), United States (8), European Union- US (1), EU (2), United Kingdom (3), Singapore (2), China (3). It also looks at steps taken by the Ministry of Electronics and Information Technology and financial sector regulators like Reserve Bank of India and the Securities and Exchange Board of India towards implementing RTAI.

* This Whitepaper is a product of joint research endeavours by Dvara Research and PwC. It is currently under industry review and is being peer reviewed at PwC. An updated version will be published after incorporating relevant feedback.

A. Context

The interest and adoption of Artificial Intelligence (AI) in financial services is rising consistently. Latest industry reports suggest that the financial sector, specifically fintech, is leading the adoption of AI in India.[2] It also appears that among the different sectors in the economy, fintech is one of the best poised to make commercial gains from the adoption of AI.[3]

The potential gains from the adoption of AI in financial services have been well documented. To summarize, the gains from AI could accrue on three counts. First, enhanced data-processing abilities, including the ability to process qualitative and audio-based data could lead to a deeper understanding of customers' needs and improve the fit of products. These insights could also significantly improve the customer's journey by sensing their needs and offering customized, relevant, and timely support in a customer-friendly format along the customer journey. In addition to deeper personalization of product design and customer support, AI systems also help financial service providers (FSPs) build stronger defenses against fraudulent activity. Second, AI systems exhibit a high degree of scalability and flexibility.[4] They appear to be capable of solving the wicked dilemma of offering hyper-personalization at scale. Finally, AI systems help realize efficiency gains from process-rationalization,[5] where providers are able to eliminate duplication of tasks by deploying AI. When AI works as intended, it can deepen financial inclusion and enhance the relevance of financial services at population-scale. For instance, Generative AI (GenAI) can significantly enhance the ease of opening accounts for uninitiated customers. It can also nudge them to improve account usage, promote budgeting, and deepen financial literacy through relevant, customized and timely content. In the case of credit, fueled by big data, algorithms could do a better job in predicting creditworthiness of thin-filed customers and credit invisibles.[6]

These gains, however, are tempered by attendant risks. The first set of AI-related risks arises from the advanced processing of rich and personal data. This includes the risk of infringement of privacy, the risk of proliferating bias and discrimination, the AI systems generating 'hallucination' and misinformation, and the reduced robustness i.e., inconsistent accuracy of the AI system. These risks can take away from the gains presented by deeper processing capabilities. Further, these risks are aggravated by the relative scalability of algorithms.[7] Just like the benefits, the risks also get scaled with the model, affecting vast swathes of customers at once. Finally, the difficulties in explaining complex algorithms create anxiety at two levels: On the one hand, it makes the algorithm impermeable and therefore hard to assess for accuracy. It significantly reduces the ability of the customers to question the algorithm, identify mistakes or seek remedies. On the other, it pronounces the philosophical tension between human autonomy and agency and automated decision-making. Inscrutable algorithms chip away at human agency, expecting humans to comply with machines they do not fully understand, nor can completely control. While these risks appear to be esoteric, they can trigger adverse systemic shifts in the financial system and jeopardize customer protection. For instance, reliance on similar, off-the-shelf machine learning tools could encourage herd behaviour among lenders and create procyclical vulnerabilities.[8] Similarly, a less-than-fit algorithm could lend to borrowers who may not have the wherewithal to repay, causing distress and default. It could affect

[2] https://indiaai.gov.in/article/how-ai-is-influencing-the-next-disruption-in-indian-fintech-space
[3] https://web-assets.bcg.com/75/ab/7ec60ba84385ad89321f8739ecaf/bcg-wheres-the-value-in-ai.pdf
[4] https://www.pwc.in/assets/pdfs/research-insights/2022/ai-adoption-in-indian-financial-services-and-related-challenges.pdf
[5] https://www.jpmorgan.com/insights/payments/payments-optimization/ai-payments-efficiency-fraud-reduction
[6] https://www.weforum.org/stories/2024/01/ai-is-driving-the-evolution-of-a-more-inclusive-financial-sector-in-latin-america-here-is-how/
[7] https://www.ecb.europa.eu/press/financial-stability-publications/fsr/special/html/ecb.fsrart202405_02~58c3ce5246.en.html
[8] https://www.fsb.org/2024/06/remarks-by-nellie-liang-on-artificial-intelligence-in-finance/

the borrowers' credit scores, cutting them off formal credit markets and severely erode the lenders' portfolio quality.[9]

An appreciation of the risks and benefits of using AI is imperative for an informed policy stance. This knowledge of the risks and benefits must also be buttressed by an understanding of their underlying causal mechanisms. Collective wisdom[10] now suggests that AI governance must adopt a lifecycle or a value-chain approach. A value-chain approach to AI focuses on improving the visibility over the various components of an AI system and the stakeholders responsible for each. It allows us to unpack the origins of risks and benefits along each link of the chain and allocate proportionate responsibilities to every actor. This visibility over the different components, their potential implications and the respective stewards also offers comfort to the regulator. This approach also allows for the distribution of legal responsibility. A governance framework anchored in the value-chain approach is most likely to be proportionate and both preventive and remedial.

Finally, this value-chain approach to governance is a necessary precondition to craft an AI governance framework that focusses on Responsible and Trustworthy AI. The paradigm of Responsible and Trustworthy AI is fast gaining traction and is poised to guide the way for ethical, safe and inclusive application of AI.[11] A quick review of literature surfaces two essential components of Responsible and Trustworthy AI (RTAI). First, it concerns itself with the AI system as-a-whole and not just the outcomes of the AI. Second, it requires AI systems to align with socially desirable values such as non-discrimination, fairness, that they be technologically robust and safe and amenable to being held to account. It appears that sometimes the terms *responsible* and *trustworthy* are used interchangeably. This Whitepaper, however, differentiates the two attributes by requiring that the processes associated with the design and deployment of AI be *responsible* so that the conduct and the outcomes of the AI system thus designed, are *trustworthy*. Thus, in this Whitepaper, *responsible* describes the processes and procedures put in place to ensure that the conduct, decision and outcomes of the AI systems are *trustworthy*.

This Whitepaper unpacks what RTAI would mean in the context of digital lending. While the term Responsible AI has been adequately conceptualized in academia and elsewhere, it still needs to be coherently contextualized to specific domains. Therefore, building on a systematic review of literature, the first section compiles principles of RTAI along with its essential components. The next section maps relevant tools to each principle. Put another way, these tools are practice-recommendations to digital lenders that can help them perform better on the specific characteristics of RTAI. The unique value addition of this publication is a distance map that allows digital lenders to gauge how far their current AI safeguards are from the desired level and how they might close the gap. The distance map takes the form of a checklist; designed for the technology teams of digital lenders. They should be able to run down the checklist without external supervision and reflect on the overall intensity of their AI safeguards. It serves as a diagnostic tool surfacing relevant areas and offers direction to lenders on how they might further strengthen their AI practices. The distance map is anchored in the understanding that AI in finance needs to be governed not for AI's sake but for the sake of finance. Responsible AI in lending is necessary for responsible lending. Therefore, to AI experts, this list of tools may appear non-exhaustive because it may not capture all the upcoming practices and measures. However, they have been deliberately curated to prioritize financial concerns and to be compatible with the technology being used by digital lenders. By distilling the characteristics of RTAI into specific practices that are amenable to the existing technological capabilities of digital lenders, this Whitepaper contributes to narrowing the principle-practice gap that exists in the research.

[9] https://nation.africa/kenya/business/mobile-money-loans-have-left-us-broke-embarrassed-and-in-ruins-4046776
[10] https://www.iso.org/standard/81118.html
[11] https://rbi.org.in/scripts/BS_ViewBulletin.aspx?Id=22851

B. Principles of RTAI

A deep review of literature indicates that for AI to be Responsible, it must necessarily exhibit the following characteristics:

1. Transparency, Explainability and Contestability:

Transparency: Transparency provides stakeholders with a broad view of the working of the AI system, tailored to their context. It includes the ability to trace the origins of the data and the decision, clearly identify automated decisions and understand the limitations of the AI system. At the same time, the principles allows the developers and deployers of AI systems to safeguard their Intellectual Property and Trade Secrets.[12] For this principle to be effective, it must provide information that is appropriate for the actor seeking it out.[13] While transparency does not assure accuracy, it makes it more likely by enabling an enquiry into the logic of the system and the origin of the underlying data.[14]

Explainability and Contestability: Explainability means enabling people to whom the outcome of an AI system relates, to understand how it was arrived at. This entails providing easy-to-understand information, which empowers those directly affected by the outcome of the AI to challenge it. We imagine the principle to have baked-in reasonable restrictions[15], in that, it only obliges the deployers of AI to make the output intelligible by sharing the underlying factors and logic (*exogenous* explainability) and not lay bare the intricacies of the model itself (*decompositional* explainability).[16] Contestability allows for timely human review and remedy if an automated system fails or produces errors, especially in sensitive contexts.

Transparency, explainability and contestability obligations may differ for the digital lender basis the use case. Globally, the extent of these attributes is being determined by the sensitivity of the use case and the implications of a misjudgement on the part of the AI (*EU AI Act)*. Excessive transparency could create confusion or expose the AI models to exploitation or manipulation. Regulators such as the Monetary Authority of Singapore (MAS) recommend that the sophistication of the explanation should match the expertise of the agent querying it. Similarly, excessive explainability could incentivize developers to reduce the number of variables in the model, reducing its accuracy.[17]

2. Fairness & Non-Discrimination

The principle of Fairness and Non-discrimination mandates that AI systems be designed and implemented to prevent biases and discriminatory outcomes by fostering inclusivity, transparency, and regular monitoring throughout the AI lifecycle, while also complying with legal and ethical standards to protect against any form of discriminatory outcomes. Discrimination here would entail, for example, dissimilar credit terms to individuals who are alike in their creditworthiness. Some regulators such as the MAS have set out practical guidance for implementing fairness. This guidance emphasizes that no two groups or individuals be treated differently without justification and the justifications thus provided should be frequently reviewed for accuracy.[18]

3. Technological Dependability of the AI system (and its ability to respond to realized risks)

This attribute is a composite of three characteristics:

---

[12] https://blogip.garrigues.com/en/intellectual-property/striking-a-balance-between-transparency-and-intellectual-property-rights-in-the-artificial-intelligence-regulation-is-not-an-easy-task
[13] https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf
[14] https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf
[15] https://oecd.ai/en/dashboards/ai-principles/P7
[16] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3974221
[17] https://oecd.ai/en/dashboards/ai-principles/P7
[18] https://www.mas.gov.sg/~/media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf

(i) *Reliability:* Reliability of the output is the ability of an AI system to perform as intended, without failure, for a given time interval, under given conditions.[19] In its essence, reliability ensures that an AI system behaves exactly as its designers intended and anticipated and repeatedly exhibits the same behaviour under similar conditions.[20] Reliability is a goal for overall correctness of the AI system.[21]

(ii) *Robustness and Resilience*: Robustness, in the context of AI systems, refers to the ability of an algorithm or a model to maintain its accuracy and stability under different conditions, including variations in input data, environmental changes, and attempts at adversarial interference. It ensures that the system can withstand unforeseen challenges and continue to function effectively.[22] Resilience of AI refers to its ability to bounce back after disruptions.[23]

(iii) *Safety and Security:* The principle of safety refers to reducing 'unintended' behaviour in the functioning of AI. It aims to prevent unwanted harms to human life, health, property or environment.[24] The principle of security would require the system to have implements in place to defend itself against security risks and to respond to and recover from these risks, should they realize. Security risks in AI include security concerns related to the confidentiality, integrity, and availability of the system and its training and output data.[25] Further, a secure system can maintain the integrity of the information that constitutes it. This includes protecting its architecture from the unauthorized modification or damage of any of its parts. A secure system also remains continuously functional and accessible to its authorized users and keeps confidential and private information secure even under hostile or adversarial conditions.[26]

4. Privacy and Data Protection

The principle of Privacy and Data Protection seeks to safeguard individuals' privacy rights by requiring the collection, processing, and storage of data to be conducted with transparency and consent, adhering to relevant laws and ethical standards. It also emphasizes the importance of minimizing data usage, ensuring that data processing does not reveal sensitive information which is not relevant to the context and which the data principal would not have ordinarily shared with the business. Further, it emphasizes ensuring data security, and providing individuals with control over their personal information, including rights to access, correct, and delete their data.

5. Protecting human agency and instituting human oversight

This principle seeks to preserve human autonomy by ensuring that the AI system remains accountable to a human. Further, it emphasizes that individuals be able to make informed and autonomous decisions regarding AI systems. This is achieved through the introduction of the element of human oversight in the functioning of the AI system.

6. Governance and Accountability

Governance: Governance emphasizes the need for effective frameworks for AI systems' development, deployment, and use, with regulation by external bodies and internal governance within organizations.

---

[19] https://airc.nist.gov/AI_RMF_Knowledge_Base/AI_RMF/Foundational_Information/3-sec-characteristics
[20] https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf
[21] https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf
[22] https://www.larksuite.com/en_us/topics/ai-glossary/robustness#what-is-robustness?
[23] https://www.wipro.com/blogs/girish-datar/robust-or-resilient/
[24] https://cset.georgetown.edu/publication/key-concepts-in-ai-safety-an-overview/
[25] https://www.nist.gov/artificial-intelligence/ai-fundamental-research-security
[26] https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf

It includes risk management, regular assessments, and stakeholder involvement, and mandates reporting to allocated boards for responsible AI adoption.

Accountability: Accountability ensures that those responsible are able to demonstrate their adherence to the principles and practices of RTAI. This allows for responsible actors to be held answerable for their roles in the functioning of the AI system.

C. How might RTAI help digital lenders?

As lenders set out to synchronize their AI practices with the RTAI principles, it is worth reflecting on the need for doing so. As briefly alluded to in the first section, the governance of AI in digital lending serves the larger goal of making digital lending safe for customers and lenders. Gains from RTAI include but are not limited to customer protection. This section sets out the key concerns that arise in each phase of designing the AI system and the RTAI practices that might address them.

Unpacking the Model Development Lifecycle for risks and gains for lenders

The broad phases of model development, called the Model Development Lifecycle (MDLC) afford us an understanding of the processes that constitute the AI system.

A typical MDLC comprises three essential phases:[27]

(i) *Solution Design Phase:* This phase involves defining the business requirements, identifying problems, selecting the appropriate methodology, and designing the overall architecture and workflow for the model. It also includes preparing the data required for the model, including identifying the sources of relevant data, cleaning it and structuring it for use.

(ii) *Model Development Phase:* In this phase, data is collected, processed, and analysed to build, train, and validate the model using selected algorithms and techniques.

(iii) *Model Deployment and Monitoring Phase:* This phase focuses on integrating the model into production systems, making it operational, and continuously monitoring its performance for all the RTAI principles.

We juxtapose this universal understanding of the MDLC with the risks and gains associated with the use of AI in lending. This allows us to form a complete impression of the vulnerabilities to be addressed and the gains to be accentuated in each phase, which, in turn helps match the phases to appropriate RTAI attributes. The visualization that follows briefly discusses the key gains and risks that may arise in digital lending from the use of AI. Further, it tags these gains and risks to the phases that they are most likely to realize in. This allows us to contemplate the RTAI attributes most relevant to a specific phase, given these potential gains and risks.

[27] https://www.frgrisk.com/the-model-development-lifecycle-mdlc-model-implementation-and-review/

Figure 1: Gains & Risks from using AI in digital lending and appropriate RTAI attributes to tackle them

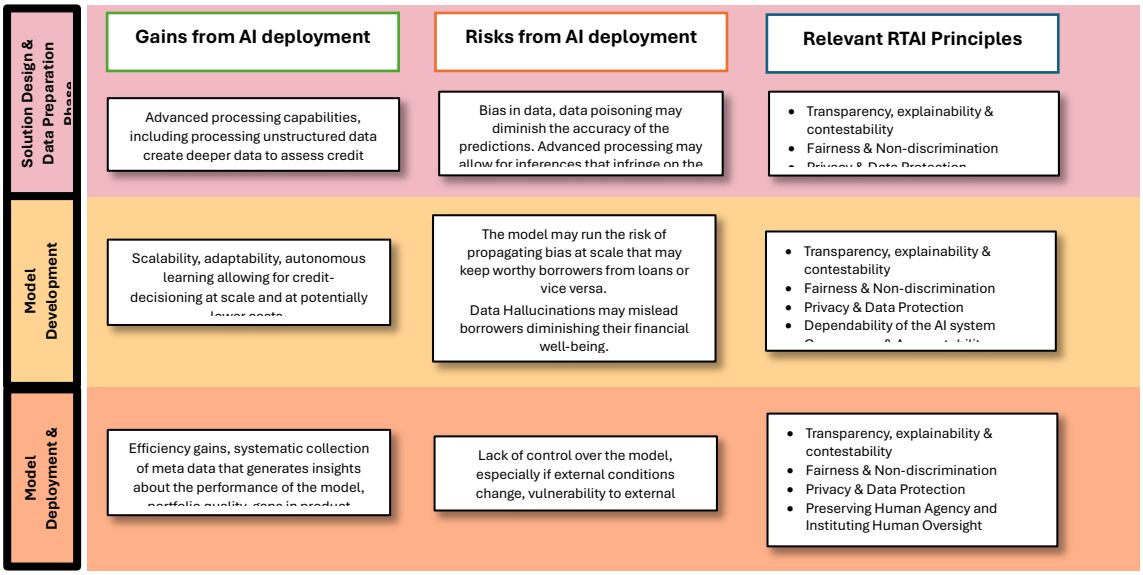

Figure 1 above serves as a compass, alerting AI systems designers and/or deployers and the lenders to the vulnerabilities that all AI systems carry and the implications they may have for lending on the whole.

These benefits and risks are mapped to relevant RTAI principles. Their implementation, however, requires a deeper understanding of the processes encapsulated within each phase of the MDLC and the tools and strategies available to give effect to the relevant RTAI attributes.

Figure 2: Unpacking the processes across MDLC phases and surfacing RTAI attributes relevant to each process

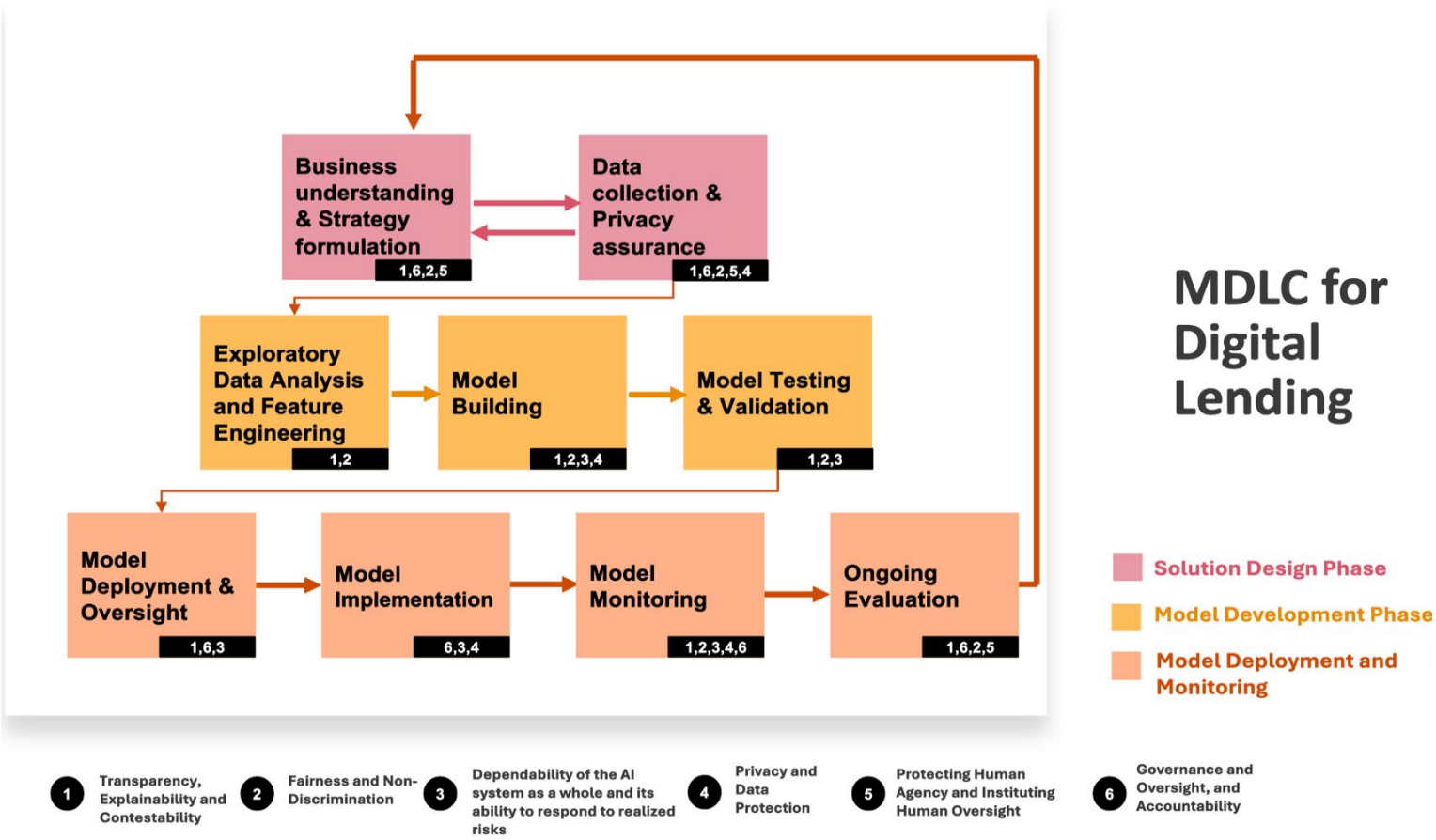

Figure 2 above is a simplified version of the MDLC. We recognize that it is an iterative process, and the steps may not strictly follow a linear direction.

Examining the phases more closely allows us to appreciate the interlinkages and interdependencies between them. It reiterates the need to craft governance frameworks that concern themselves not only with the outcomes of AI but also with how the AI systems themselves came to be. In regulatory parlance, the former is referred to as a product safety approach. There is now a growing appreciation that AI regulations must combine product safety approach that lays out the permissible outcomes and the value chain approach that creates bounds around how AI may be developed to begin with.[28]

RTAI Priorities: What lenders must look out for in each phase

1. Solution Design:

As discussed previously, the solution design phases comprise two essential tasks. First, it focuses on the needs of the business, the requirements of the AI system, identifying the appropriate model etc. Second, it focuses on identifying data sources, procuring the data, cleaning it up and making it fit for use for the AI system. In this phase, risks may arise from a poor fit of the model, disregarding customer concerns when zeroing in on the design of the model. Risks may also arise from using data that may not have been addressed for inheriting bias or data that may have been obsolete etc. Finally, privacy risks may arise if the data were not collected with proper and informed consent or if the collection and processing did not adhere to the principles of data minimization and purpose limitation.

RTAI practices and tools that could help lenders defend themselves against these risks include:

(i) **Instituting Ethical Review Boards and Governance Frameworks:** This practice is anchored in the attribute of Accountability. It emphasizes the need to define roles for model oversight and accountability that are guided by governance frameworks and a policy checklist from stakeholders, applied to the model development process.

(ii) **Folding in the concerns of diverse stakeholders when determining the details of the AI model:** It is essential that the model prioritize the right matrices with regard to the (often conflicting) interests of all stakeholders. Hypothetically, the model may exhibit accuracy levels of 95% which may bode well for portfolio quality, however when dealing with financially vulnerable customers, an error rate of 5% may cause deep distress to the customers. It is essential that models optimize along the concerns and priorities of all stakeholders for lending as whole to be responsible.[29] Often, inclusivity is included as an afterthought and the customers' concerns are restricted to the challenges they face in the UI/UX of the product. However, there is potential to be more inclusive, purposeful and responsible right from the phase of Solution Design.[30]

(iii) **Following Data Protection Principles and Assuring Privacy:** Data is the mainstay of any AI model. The quality of the data has enduring implications for the functioning of the model. Moreover, not only does data have implications for the loan quality, customer experience, default and distress, but the way the data is used also has direct implications for the privacy of the borrowers and many a times of related cohorts. Data protection

[28] https://www.youtube.com/watch?v=LfHNGtolH3M&t=3842s
[29] https://www.weforum.org/stories/2021/09/companies-how-to-build-more-inclusive-artificial-intelligence/
[30] https://partnershiponai.org/wp-content/uploads/dlm_uploads/2022/07/PAI_whitepaper_making-ai-inclusive.pdf

practices emphasize being mindful of data sources, identifying data integration platforms and opting for privacy-enhancing techniques such as masking etc. before feeding in the data.

Key RTAI practices relevant to the Solution Design Phase are set out in ***Box 1***.

Box 1: RTAI recommendations for Solution Design Phase

Solution Design Phase: **RTAI Recommendations for Digital Lenders**

RTAI recommendations for the Solution Design Phase, encompassing (i) focusing on defining business requirements and formulating strategy as well as (ii) preparing the data essential for the model, include:

(i) Ethical Review Boards and Governance Frameworks:

This defines roles for model oversight and accountability, guided by governance frameworks and a policy checklist from stakeholders, applied to the model development process.

(i) Data Sources:

Data collection strategies are planned before model development, adhering to regulatory policies for ethical handling and minimizing demographic bias, with stakeholder involvement in identifying sensitive attributes.

(ii) Data Integration:

Data integration platforms should be defined in the solutioning phase, The goal is to ingest data in a manner that protects data privacy and is stable in the long run.

(iii) Data Quality:

Data quality criteria have to be established to ensure that the data being used will provide reliable and fair outputs.

(iv) Data Encryption Tools:

Before feeding the data to the model development environment appropriate masking/encryption techniques should be defined.

(v) Data Anonymization Services:

Select suitable anonymization techniques tailored to different features, ensuring the service meets these needs. Masking sensitive attributes helps mitigate bias by preventing the model from correlating demographic groups with outcomes.

Model Development:

This is the phase wherein the model is trained on the dataset. This phase entails data collection, processing, analysis and validating the model. In this phase, RTAI practices emphasize:

(i) **Appropriate Algorithm Selection Framework:** Consider algorithms that are compatible with the limitations of the data at hand, such as, imbalanced datasets. Other considerations when selecting an appropriate algorithm include being mindful of proliferating bias and achieving transparency without necessarily offsetting accuracy.

(ii) **Being mindful of the scope of bias and the working of the model** by deploying fairness assessment tools. An allied measure is to deploy tools to interpret the model and determine leading factors of decisions.

(iii) **Identify parameters of interest and set up systems to track them,** for instance identifying a suite of proxies for fairness, inclusivity and resilience and setting up dashboards that collect the data on the proxies, parse through them and flag discrepancies/divergences.

Key RTAI practices relevant to the Model Development Phase are set out in ***Box 2***.

Box 2: RTAI recommendations for Model Development Phase

Model Development Phase: **RTAI Recommendations for Digital Lenders**

RTAI recommendations for the Model Development Phase, where data is processed and analyzed to build, train and validate the model include:

(i) Feature Store:

Maintain comprehensive documentation for each feature, establish clear policies for data quality, privacy, and ethics, and assign data stewards for governance compliance.

(ii) Algorithm Selection Framework:

Consider algorithms that can handle imbalanced datasets and reduce biases. To assure transparency without compromising on accuracy, adding in a proxy/surrogate interpretable model that mimics your main model's outputs.

(iii) Model Training Pipelines:

Incorporate techniques like re-sampling, re-weighting, or synthetic data generation in your training pipeline. Use techniques like adversarial debiasing and data masking to further address any potential bias.

(iv) Bias and Fairness Assessment tools:

Analyze outcome distributions across demographics using metrics like disparate impact ratio, false rejection rate, and statistical parity to detect bias. Implement corrective measures when biases are found.

(v) Model Interpretability tools:

Techniques like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) help in understanding the contribution of different features to model predictions.

(vi) Performance Monitoring Tools:

Create real-time performance dashboards that include model performance and fairness metrics and implement mechanisms to flag any discrepancies.

2. Model Deployment and Monitoring

This phase focuses on integrating the model into the processes, deploying it at full scale and closely monitoring its performance. The RTAI recommendations in this phase relate to establishing protocols for the management of external stakeholders that may either be offering services essential to the smooth running of the AI system or may need levers to scrutinize the system (such as the regulator).

(i) **Management of external counterparties:** The model deployment would likely require several parties—external or in-house to interoperate with the AI systems. They would require differentiated access controls and proportionate responsibilities. This consideration also has a legal aspect to it; it may be prudent to design contracts that highlight the roles, responsibilities, liabilities and indemnities of each party to properly allocate responsibilities.

(ii) **Data management protocols:** In this phase, the deployers need in place protocols and processes to take care of data that may no longer be needed.

Key RTAI practices relevant to the Model Development Phase are set out in ***Box 3***

Box 3: RTAI recommendations for Model Deployment and Monitoring Phase

Model Deployment and Monitoring Phase: **RTAI Recommendations for Digital Lenders**

RTAI recommendations for the Model Deployment and Monitoring Phase, which focuses on model integration into operation and continuous performance monitoring, include:

(i) Cloud Services or On-Premises Servers:

Use containerization tools like Docker to ensure that the development environment can be replicated. Use encryption for data at rest and in transit. Employ secure access controls and regular audits

(ii) API Management Systems:

Implement rate limiting and quotas to prevent abuse and ensure fair usage.

(iii) Monitoring and Logging Systems:

Set up alerting systems to notify administrators of any anomalies or deviations from expected behavior.

(iv) Access Management Systems:

Implement RBAC and MFAs to authenticate and manage who uses the system.

(v) Regulatory Compliance Management Systems:

Conduct regular compliance audits to ensure ongoing adherence to relevant laws and regulations.

(vi) User Support Channels:

Create self-service portals for loan information and AI criteria, enabling data modification with logs. Provide processes for individuals to opt out of data collection and to appeal process for AI decision review.

(vii) Data Retention/Disposal Protocols:

Ensure data is encrypted and securely deleted using industry-standard practices.

D. The distance to the RTAI Frontier: How might lenders start implementing RTAI

These principles lend themselves to a suite of practices that can help digital lenders implement RTAI in their operations. The checklist, as discussed earlier, does not include all possible and especially upcoming practices of RTAI. However, it recommends those practices that have come to be recognised as standard and in their absence, it may be difficult to achieve RAI. Thus, the checklist may not be sufficient to implement RAI but is necessary.

Editor's Note: The checklist is under industry review and therefore, has not published.. An updated version of the Whitepaper will be published after the industry consultations.

Table 1: Checklist of RTAI Practices for Digital Lenders (an excerpt)

Key Factor	Principle	Question	Score 1-4	Level of Criticality
Policy for Fairness in models	Fairness & Non discrimination	To what degree is there a governance policy to define the fairness rules in the models? Have you addressed any implicit bias in your models.	0: No policies in place; 1: Policies catering to only identifying known sources of discrimination; 2: Policies catering to only identifying underlying sources of discrimination; 3: Policies catering to only identifying and addressing known sources of discrimination; 4: Policies catering to only identifying and addressing underlying sources of discrimination.	High