Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Minimum Viable Governance Methodology for Artificial Intelligence in Fragile and Resource Constrained Contexts

Albert Adusei Brobbey

albertab@primemeridianintl.com

California Intercontinental University

Narayan P. Bhosale

California Intercontinental University

James Ackom

Prime Meridian International

Systematic Review

Keywords: responsible AI, fragile contexts, resource-constrained environments, AI governance, public sector AI, humanitarian AI, resilience, human-AI collaboration, uncertainty-aware modelling, accountability

Posted Date: June 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9816076/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Artificial intelligence is increasingly proposed for public service delivery, humanitarian coordination, social protection, and crisis response. Yet many governance frameworks assume stable infrastructure, reliable data pipelines, mature institutions, and sufficient technical capacity. Those assumptions often fail in fragile, conflict-affected, and resource-constrained contexts, where data may be incomplete or politically contested, connectivity may be intermittent, and human oversight capacity may be scarce. This article introduces FRAG-AI, a minimum viable governance methodology for AI systems deployed under constraint. The method translates existing responsible AI principles into operational gates that can be applied before, during, and after deployment: context constraint profiling, use-case triage, data reliability assessment, uncertainty-aware model selection, human-AI collaboration design, accountability logging, resilience stress testing, and adaptive post-deployment monitoring. FRAG-AI contributes a constraint-adjusted scoring model and a deployment decision matrix that distinguish systems that can proceed, proceed only with safeguards, remain in assisted decision-support mode, or should not be deployed. The methodology is demonstrated through three scenario applications: rural public health triage, humanitarian supply allocation, and social protection eligibility support. The article argues that responsible AI in fragile environments cannot be reduced to stronger models or generic ethical principles. It requires governance designs that are small enough to implement, strong enough to prevent foreseeable harm, and resilient enough to remain accountable when institutions and infrastructure are under stress.

1. Introduction

AI research is often evaluated in environments that are data-rich, institutionally stable, and technically resourced. Benchmark datasets are available, infrastructure failures are treated as exceptions, and accountability can be assigned to organizations with formal governance structures. These conditions do not describe many of the contexts in which AI is now being proposed. Public agencies, humanitarian organizations, non-governmental organizations, local health systems, and social protection programs increasingly face pressure to use AI for targeting, prioritization, forecasting, fraud detection, early warning, and operational coordination. In fragile and resource-constrained settings, however, the same AI system can become harmful when it is introduced into weak data environments, fragmented institutional arrangements, and contested political conditions.

The problem is not simply that low-resource settings have less data or weaker infrastructure. The deeper problem is that governance itself becomes fragile. A model may be technically accurate on a limited validation set but still be unsafe because the data-generation process excludes displaced people, because appeal channels do not exist, because the agency deploying the system cannot monitor drift, or because local staff lack authority to override automated recommendations. In these environments, model performance is only one part of system performance. Resilience, accountability, contestability, and the ability to degrade safely become equally important design properties.

Existing responsible AI frameworks provide valuable starting points. NIST's AI Risk Management Framework, the NIST Generative AI Profile, ISO/IEC 42001, the OECD AI Principles, UNESCO's Recommendation on the Ethics of AI, and the EU AI Act all emphasize risk management, transparency, accountability, human oversight, and data governance [3–8]. Broader AI ethics and human-centered AI scholarship similarly stresses reliability, safety, trustworthiness, and the need to move from abstract principles to accountable practice [16, 17]. Humanitarian data responsibility guidance also stresses safe, ethical, and effective management of personal and non-personal data in operational response [9, 10]. Yet these frameworks are often too broad or capacity-intensive for organizations that operate with intermittent electricity, incomplete registries, unstable staffing, limited legal recourse, and urgent operational demands. What is missing is not another high-level list of principles, but a method for deciding the minimum governance controls that must be present before AI can be responsibly used under constraint.

This article addresses that gap by proposing FRAG-AI, a minimum viable governance methodology for AI in fragile and resource-constrained contexts. The method is designed for organizations that cannot implement a full enterprise AI management system at once but still need a disciplined way to determine whether an AI use case should proceed, what safeguards are required, and when deployment should be paused or prohibited. FRAG-AI is not a substitute for mature AI governance. Rather, it is a structured floor: a set of operational gates below which deployment should not occur.

The contribution is threefold. First, the article specifies a context-sensitive governance problem: AI systems in fragile contexts fail when model risk, data unreliability, infrastructure instability, and institutional capacity gaps interact. Second, it develops FRAG-AI as a methodology with eight modules that translate responsible AI principles into practical controls. Third, it proposes a constraint-adjusted deployment score and demonstrates how the method can be applied to three public-interest use cases. The result is a practical framework for researchers, public agencies, humanitarian actors, and technology partners seeking to design AI systems that remain accountable when conditions are difficult rather than ideal.

2. Background and Gap

Fragility is multidimensional. The OECD States of Fragility framework evaluates fragility across economic, environmental, political, security, societal, and human dimensions, and identifies contexts where exposure to risk exceeds available coping capacity [2]. Such contexts are not only countries experiencing open conflict. They may include subnational regions, displaced populations, emergency response systems, post-disaster settings, health systems under stress, or public agencies with limited digital maturity. The collection call for this journal rightly treats conflict-affected and resource-constrained contexts broadly, including geographic areas, populations, topics, and forms of research where instability shapes data and decision-making.

AI governance literature has developed rapidly, but much of it assumes organizational capacity that fragile contexts may lack. AI risk management frameworks require inventories, risk registers, model documentation, monitoring processes, incident response procedures, and assigned accountability roles. These are important controls. The challenge is that a local public health office, refugee service provider, or municipal social protection agency may not have the staff, budget, data infrastructure, or legal architecture to implement them fully. A governance approach that assumes capacity can unintentionally exclude the settings where governance is most needed.

At the same time, low-resource AI methodology often focuses on technical strategies such as transfer learning, few-shot learning, active learning, data augmentation, uncertainty estimation, and domain adaptation [18]. AI-for-development research has also emphasized the promise of machine learning in settings where institutional and resource constraints shape both data and deployment [19]. These methods are valuable, but they do not resolve governance questions by themselves. A model that is uncertainty-aware may still be deployed in a process where uncertain outputs are treated as authoritative. A transfer-learning system may still import biases from a source domain that is socially or institutionally unlike the target context. A small-data model may still become part of a high-stakes eligibility process without appeal mechanisms.

The sociotechnical AI literature warns against abstraction traps: the tendency to define a computational problem while stripping away the institutional and social context that gives the problem meaning [13]. Data cascades research similarly shows how early data quality problems can propagate through downstream high-stakes AI systems [12]. These insights are especially important in fragile settings, where missingness, displacement, under-registration, documentation gaps, and security risks are not random inconveniences. They are often structured by conflict, poverty, administrative exclusion, and power asymmetries.

FRAG-AI is therefore motivated by a double gap. Governance frameworks are often too general and resource-intensive, while low-resource AI methods are often too technical and insufficiently institutional. A useful methodology must connect both sides. It must ask whether a model can work with limited data, but also whether the organization can contest, monitor, explain, and repair the decision process if the model fails.

Existing approach	Core strength	Fragile-context limitation	FRAG-AI response
Responsible AI principles	Normative clarity around fairness, transparency, accountability, and safety	Often difficult to translate into deployable controls under severe resource constraints	Turns principles into minimum operational gates and decision thresholds
Enterprise AI management systems	Strong lifecycle governance and continuous improvement	May assume mature inventories, documentation, legal teams, monitoring tools, and stable ownership	Provides a lightweight governance floor for organizations that cannot yet implement full maturity models
Low-resource machine learning	Methods for small datasets, transfer learning, uncertainty, and adaptation	May not address institutional accountability, appeal, oversight, and political vulnerability	Links technical method choice to governance capacity and human-AI collaboration design
Humanitarian data responsibility	Strong emphasis on do-no-harm, data protection, and operational ethics	Does not always specify AI-specific model, drift, automation, and uncertainty controls	Extends data responsibility into AI deployment gates and monitoring routines

3. Design Requirements for AI Governance Under Constraint

A governance methodology for fragile and resource-constrained contexts should satisfy six design requirements. The first requirement is constraint awareness. Governance controls must begin by identifying the constraints that shape deployment: data scarcity, data unreliability, connectivity, compute, staff skills, procurement dependence, legal recourse, security exposure, and community trust. Without this profile, governance defaults to assumptions imported from better-resourced settings.

The second requirement is minimum viability. A method designed for constrained institutions cannot require a full AI office, dedicated model risk team, legal engineering function, and advanced monitoring stack before it offers any guidance. It must define the smallest set of controls that are necessary for safe use, while also showing how those controls can grow as capacity improves. Minimum viable governance is not weak governance; it is a disciplined recognition that some controls are non-negotiable and others can be tiered.

The third requirement is decision specificity. Many AI governance documents state that organizations should ensure fairness, accountability, and transparency. Field teams need sharper guidance: can this AI system be used for direct decision-making, only as advisory support, only in pilot mode, or not at all? A fragile-context method should produce deployment decisions that are clear enough for operational use.

The fourth requirement is uncertainty visibility. In fragile settings, missing data and shifting conditions are normal. Governance must therefore require uncertainty to be represented, communicated, and acted

on. A model that hides uncertainty is more dangerous than a less accurate system that honestly signals when it does not know enough.

The fifth requirement is contestability. People affected by AI-supported decisions must have practical mechanisms to challenge errors, especially when systems influence access to food, cash transfers, health services, shelter, documentation, or protection. Where appeal mechanisms cannot be provided, high-stakes automated recommendations should remain non-determinative.

The sixth requirement is resilience. AI systems should be evaluated not only for accuracy in normal operation but also for graceful degradation under stress. The relevant question is: what happens when the network fails, the data are stale, the local office loses staff, a subgroup is missing from records, the model drifts, or a partner organization withdraws access?

4. The FRAG-AI Methodology

FRAG-AI is organized as an eight-module methodology. The modules are sequential enough to support project planning, but iterative enough to support adaptation during deployment. Each module produces a concrete output that can be reviewed by technical staff, program managers, community representatives, and accountable decision-makers.

Module	Purpose	Minimum output	Stop condition
1. Context constraint profile	Identify operational, institutional, data, infrastructure, and security constraints	Constraint profile with severity ratings and evidence notes	Unknown operating context or no responsible owner
2. Use-case triage	Classify the AI use case by harm severity, reversibility, rights impact, and urgency	Use-case risk class and permitted automation level	High-impact use case with no human review or appeal path
3. Data reliability assessment	Evaluate completeness, representativeness, timeliness, consent, provenance, and political sensitivity	Data reliability score and missingness map	Data source cannot be explained, lawfully used, or updated
4. Uncertainty-aware model choice	Select models that fit sample size, domain shift, and explainability needs	Model choice rationale with uncertainty communication plan	Model cannot communicate uncertainty for high-stakes decisions
5. Human-AI collaboration design	Define decision roles, override rights, escalation points, and training needs	Human-AI workflow map and override protocol	Human operators are accountable but lack authority or competence to intervene
6. Accountability and audit ledger	Record data versions, model versions, prompts, recommendations, overrides, and decisions	Minimal audit log and retention rule	No ability to reconstruct materially important decisions
7. Resilience stress testing	Test system behavior under infrastructure failure, stale data, subgroup missingness, drift, and adversarial misuse	Stress-test report and degradation plan	Unsafe output persists during predictable failure conditions
8. Adaptive monitoring	Monitor drift, harm signals, complaints, override rates, and context changes	Monitoring dashboard or low-tech review routine	No assigned review cadence or incident response trigger

4.1 Module 1: Context Constraint Profile

The Context Constraint Profile is the entry point. It requires the deploying organization to rate the deployment environment across five dimensions: data constraint, infrastructure constraint, institutional constraint, human oversight constraint, and protection constraint. Data constraint refers to missingness, quality, provenance, representativeness, and update frequency. Infrastructure constraint refers to connectivity, power, compute, cybersecurity, and tool availability. Institutional constraint refers to legal authority, accountability roles, procurement dependence, staffing, and continuity. Human oversight constraint refers to whether qualified people can interpret, override, and contest the AI system. Protection constraint refers to the sensitivity of affected populations and the possibility that data or outputs could expose people to harm.

The profile should be written in plain language rather than technical shorthand. For example, a data constraint should not be recorded merely as 'low quality data.' It should say whose data are missing, why, how often records are updated, what decisions depend on those records, and whether missingness follows predictable social or geographic lines. This is consistent with the warning from data-cascade research that upstream data problems can become downstream model failures when they are hidden behind aggregate metrics [12].

4.2 Module 2: Use-Case Triage

The same AI method can be acceptable in one use case and unacceptable in another. FRAG-AI triages use cases according to four properties: harm severity, reversibility, decision authority, and rights proximity. Harm severity asks what could happen if the system is wrong. Reversibility asks whether an incorrect recommendation can be corrected before serious harm occurs. Decision authority asks whether the system advises, ranks, allocates, or decides. Rights proximity asks whether the output affects legal status, movement, health, safety, livelihood, eligibility, or access to essential services.

The result is not simply a high, medium, or low risk label. It is an automation boundary. Low-risk use cases may allow automated processing with human review by exception. Medium-risk use cases should remain human-supervised. High-risk or rights-proximate use cases should remain advisory unless contestability and oversight are strong. Prohibited use cases include systems that make irreversible high-impact decisions without human review, appeal, or auditability.

4.3 Module 3: Data Reliability Assessment

Data reliability in fragile contexts should be treated as a governance issue, not a preprocessing task. FRAG-AI assesses data across six questions. Who or what is missing? How current are the records? How were data collected, consented, and transformed? Can the organization explain provenance? Are there plausible incentives to manipulate or suppress data? Could use of the data increase risk for vulnerable groups?

The assessment produces a Data Reliability Score from 0 to 4. A score of 0 indicates unknown or unverified provenance. A score of 1 indicates severe incompleteness or likely biased missingness. A score of 2 indicates usable but materially incomplete data. A score of 3 indicates adequate quality for advisory use with monitoring. A score of 4 indicates strong provenance, representativeness, update frequency, and documentation. In high-impact contexts, a low Data Reliability Score should reduce the permitted automation level even if model accuracy appears acceptable.

4.4 Module 4: Uncertainty-Aware Model Choice

Resource-constrained AI projects often adopt transfer learning or pretrained models because local datasets are too small to train models from scratch. This can be useful, but FRAG-AI requires the model choice to be justified against the target context. A model imported from a stable administrative environment may not transfer to a displacement camp, remote clinic, or post-disaster registry. The

model choice should therefore include a source-domain comparison, a target-domain uncertainty plan, and a rule for abstention when uncertainty is high.

For many fragile-context deployments, simpler models may be more governable than more complex models. A transparent scoring model with explicit uncertainty bands may be preferable to a high-performing black-box model if local staff cannot interpret errors or explain decisions. The point is not to reject advanced AI. It is to align model complexity with oversight capacity, data reliability, and the stakes of the decision.

4.5 Module 5: Human-AI Collaboration Design

Human oversight is often invoked as a safeguard, but it fails when humans lack time, authority, training, or information. FRAG-AI therefore requires a Human-AI Collaboration Design before deployment. This design specifies who sees the output, what explanation is provided, when the human can override it, when escalation is required, how disagreement is logged, and what happens when staff are overloaded.

The design must avoid accountability dumping. It is not enough to say that a human remains responsible if the human has no realistic ability to challenge the system. In fragile settings, field officers may face pressure to follow automated rankings because the system appears objective or because donors expect efficiency. A valid collaboration design must preserve meaningful human discretion and make override legitimate rather than exceptional.

4.6 Module 6: Accountability and Audit Ledger

Accountability requires reconstruction. If an affected person challenges an AI-supported decision, the organization should be able to identify which data, model version, rule, recommendation, human review, and final decision were involved. FRAG-AI does not require a sophisticated enterprise audit platform. It requires a minimal accountability ledger that records the information necessary to understand material decisions.

The ledger can be implemented with a secure spreadsheet, local database, case management system, or formal logging platform depending on capacity. The minimum record includes case identifier, data snapshot or version, model version, confidence or uncertainty value, recommendation, human reviewer, override decision, reason code, timestamp, and appeal status. For sensitive contexts, the ledger must also minimize exposure by using access controls, pseudonymization, retention limits, and secure storage consistent with humanitarian data responsibility guidance [9, 10].

4.7 Module 7: Resilience Stress Testing

Before deployment, FRAG-AI requires stress tests that reflect likely failure conditions. These include data staleness, connectivity loss, subgroup missingness, sudden demand surge, staff turnover, model drift, adversarial manipulation, and institutional handover. The purpose is not to prove that the system is perfect. It is to identify whether the system degrades safely and whether staff know what to do when conditions deteriorate.

Stress testing should produce a degradation plan. For example, if connectivity fails, the system may shift to local advisory mode with conservative thresholds. If data are stale, recommendations may be suppressed for high-risk cases. If complaint rates rise, the system may pause automated ranking until review. If a subgroup is underrepresented, outputs affecting that subgroup may require manual confirmation or community validation.

4.8 Module 8: Adaptive Monitoring

Fragile contexts change quickly. A governance decision made at launch may become invalid after conflict displacement, funding cuts, disease outbreak, political change, or infrastructure failure. Adaptive monitoring therefore tracks both model behavior and context behavior. Relevant indicators include error reports, appeals, override rates, subgroup outcome differences, data update delays, missingness patterns, service denials, staff workload, and incidents of misuse.

Monitoring does not need to be technically elaborate. A monthly review meeting with a structured log may be more realistic than a full dashboard. What matters is that monitoring has an owner, cadence, trigger thresholds, and authority to pause or modify deployment. Without that authority, monitoring becomes documentation without governance force.

5. Constraint-Adjusted Deployment Scoring

FRAG-AI includes a scoring model to support consistent deployment decisions. The model is intentionally simple enough for non-specialist governance teams to use, while still forcing explicit judgment. It has two components: a Governance Control Score and a Context Constraint Penalty.

The Governance Control Score is the average of eight module scores, each rated from 0 to 4. A score of 0 means the control is absent. A score of 1 means the control is informal or incomplete. A score of 2 means the control exists but is weak or not tested. A score of 3 means the control is implemented and usable. A score of 4 means the control is implemented, tested, documented, and assigned to an accountable owner.

The Context Constraint Penalty is the average severity of five constraint dimensions: data, infrastructure, institutional, oversight, and protection. Each is rated from 0 to 3, where 0 means minimal constraint and 3 means severe constraint. The penalty is converted into a multiplier: Constraint Multiplier = 1 - (Penalty / 6). This means severe constraint reduces the deployment readiness score, even when governance controls are present, because controls may be harder to execute under stress.

The final FRAG-AI Readiness Score is calculated as: Readiness Score = (Governance Control Score / 4) x Constraint Multiplier x 100. This score is not a claim of statistical validity. It is a structured decision aid. Its purpose is to prevent organizations from treating generic governance checklists as sufficient when deployment conditions make those controls unreliable.

Readiness score	Deployment decision	Required governance posture
80–100	Proceed with monitoring	Controls are strong and constraints are manageable; maintain audit, monitoring, and incident response.
60–79	Proceed with safeguards	Use limited rollout, human supervision, conservative thresholds, stress testing, and scheduled review.
40–59	Assisted decision-support only	System may support analysis but should not rank, allocate, deny, or trigger high-impact action without strong human review.
Below 40	Do not deploy	Controls and context are insufficient; invest in data, oversight, appeals, and institutional capacity first.
Any score with stop condition	Pause or prohibit	A module stop condition overrides the numeric score.

6. Demonstrative Scenario Applications

This section demonstrates how FRAG-AI can be applied. The scenarios are not empirical case studies and do not report field data. They are structured applications designed to show how the methodology changes deployment decisions when conditions vary.

Scenario	Initial AI proposal	Key constraints	FRAG-AI decision	Safeguards
Rural public health triage	Use AI to prioritize patients for referral where clinicians are scarce	Incomplete records, intermittent connectivity, few trained reviewers, high health impact	Assisted decision-support only	Uncertainty display, clinician confirmation, local fallback protocol, audit ledger, subgroup missingness review
Humanitarian supply allocation	Rank communities for delivery of food or shelter supplies after displacement	Rapid population movement, contested data, security exposure, high political sensitivity	Proceed with safeguards	Community validation, no single-source data reliance, manual appeal pathway, conservative thresholds, harm review
Social protection eligibility support	Flag cases likely eligible for cash assistance or fraud review	Administrative exclusion, outdated registries, appeal limitations, rights-proximate service access	Assisted decision-support only	Positive-use prioritization only, no automated denial, appeal record, data provenance notes, periodic fairness review

The rural health scenario illustrates the difference between advisory and determinative use. AI may help clinicians identify cases that require attention, especially when staff are scarce. However, if records are incomplete and connectivity is unstable, the system should not automatically determine access to referral. The safe use pattern is constrained decision support: the AI output is displayed with uncertainty, clinicians retain authority, and the system shifts to local fallback rules when data are stale.

The humanitarian supply scenario shows why participatory validation matters. A ranking model may improve coordination, but in displacement settings, population data can be incomplete, duplicated, outdated, or politically contested. If the model treats a flawed registry as objective truth, it can reinforce exclusion. FRAG-AI therefore requires data triangulation, community validation, and an explicit rule that automated rankings cannot be the sole basis for exclusion from essential aid.

The social protection scenario shows how the direction of automation matters. A model used to identify potentially eligible households for further review may be acceptable with safeguards, while a model used to deny benefits or initiate punitive fraud action would require much stronger governance. Under FRAG-AI, positive-use prioritization can proceed sooner than automated denial because the harm profile and reversibility differ.

7. Discussion

7.1 From Ethical Principles to Operational Floors

The central argument of this article is that fragile contexts require a governance floor rather than only governance aspiration. Principles such as fairness, transparency, accountability, and human oversight remain essential, but they must be expressed as deployable controls. FRAG-AI asks whether there is enough data reliability to support the use case, whether uncertainty is visible, whether a human can meaningfully intervene, whether decisions can be reconstructed, and whether the system can degrade safely.

This approach also helps avoid a common failure in AI-for-development projects: treating pilots as harmless because they are experimental. In fragile settings, pilots can shape access to resources, redirect staff attention, influence donor confidence, and create institutional dependence. Minimum viable governance should therefore apply before pilots, not only before scale-up.

7.2 Implications for Researchers

AI researchers working in low-resource settings should report governance-relevant context alongside technical performance. This includes data provenance, missingness patterns, infrastructure assumptions, target population differences, institutional decision pathways, and monitoring arrangements. Model cards and datasheets are useful tools, but they should be extended to include context constraints, bias and fairness risks, and deployment authority [14, 15, 20].

Researchers should also distinguish technical low-resource performance from institutional readiness. A model that performs well on a small local dataset may still be unready if the organization cannot maintain data updates, explain outputs, handle appeals, or monitor harm. Conversely, a modest model may be valuable if embedded in a transparent workflow with clear human judgment and contestability.

7.3 Implications for Public Agencies and Humanitarian Organizations

For public agencies and humanitarian organizations, FRAG-AI provides a practical intake process. Before procuring or deploying an AI tool, the organization can require a context profile, use-case triage, data reliability assessment, oversight workflow, audit plan, and stress-test report. These outputs can be requested from vendors or produced internally. The method therefore supports procurement governance as well as internal risk management.

The methodology also encourages proportionate governance. Not every use case requires the same burden. A model that clusters non-sensitive operational documents for internal search does not require the same controls as a model that prioritizes households for aid. FRAG-AI allows lighter controls for low-impact uses while preserving strict stop conditions for high-impact uses where oversight and contestability are absent.

7.4 Implications for Donors and Technology Partners

Donors and technology partners often emphasize innovation, scale, and efficiency. In fragile contexts, they should also fund the governance substrate: data stewardship, community validation, appeal mechanisms, local staff training, audit routines, secure infrastructure, and post-deployment monitoring. Without these investments, AI projects may externalize governance burdens onto already constrained institutions.

Technology partners should avoid exporting systems that require cloud connectivity, high data maturity, or specialist oversight when those conditions are absent. They should design for offline operation, low-bandwidth monitoring, local language support, conservative defaults, uncertainty communication, and handover to local institutions. Resilience is not a feature added after deployment; it is a design requirement from the beginning.

8. Limitations and Future Research

This article proposes a methodology and demonstrates it through structured scenarios, but it does not report field validation. Future work should test FRAG-AI with public agencies, humanitarian organizations, and community stakeholders in specific contexts. Validation should examine whether the scoring model improves deployment decisions, whether the stop conditions are usable, and whether the methodology can be integrated into procurement, ethics review, and project management workflows.

A second limitation is that the scoring thresholds are deliberately heuristic. They are designed to structure judgment, not replace it. Future research could calibrate the thresholds through expert elicitation, Delphi studies, retrospective analysis of AI deployment failures, or field pilots. The weighting of constraint dimensions may also need to vary by domain. For example, protection risk may deserve heavier weighting in humanitarian protection systems, while infrastructure resilience may dominate in disaster response logistics.

A third limitation concerns power and participation. FRAG-AI includes community validation and contestability, but a methodology cannot by itself guarantee meaningful participation. Participatory AI can become symbolic if affected communities lack real authority over deployment decisions. Future research should therefore examine how governance methods can transfer decision power, not merely collect feedback.

Finally, FRAG-AI is designed as a minimum viable governance floor. It should not be used to justify permanently low governance maturity. The goal is to enable responsible early action under constraint while creating a path toward stronger AI management systems, better data governance, and more durable institutional accountability.

9. Conclusion

AI systems are increasingly being proposed for the places where governance is hardest: fragile institutions, unstable infrastructure, resource scarcity, humanitarian crises, and contested data environments. These are also the places where AI failures can be most consequential. Responsible AI in such settings requires more than technical adaptation and more than ethical principle statements. It requires a governance method that recognizes constraint as a core design condition.

FRAG-AI offers such a method. By combining context profiling, use-case triage, data reliability assessment, uncertainty-aware model choice, human-AI collaboration design, accountability logging, resilience stress testing, and adaptive monitoring, it provides a practical governance floor for AI deployment under constraint. Its central message is simple: if an organization cannot explain the context, validate the data, preserve human authority, reconstruct decisions, and degrade safely, the AI system should not be used for high-impact decisions. In fragile and resource-constrained contexts, restraint is not anti-innovation. It is the condition for trustworthy and resilient innovation.

Declarations

Competing interests:

The authors declare no competing interests.

Ethics approval and consent to participate:

Not applicable. This article does not involve human participants, human data, animal subjects, or field experimentation.

Consent to publish:

Not applicable. This manuscript contains no identifiable images or other identifiable personal information.

Funding:

The authors received no specific funding for this work.

Author Contribution:

A.A.B. conceptualized the study, developed the FRAG-AI methodology, drafted the manuscript, and prepared the submission materials. N.P.B. contributed to the governance framing, methodology refinement, and critical review of the manuscript. J.A. contributed to the practical deployment framing, resource-constrained implementation perspective, and manuscript review. All authors reviewed and approved the final manuscript.

Data availability:

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

1. Discover. Artificial Intelligence. Submission guidelines. Springer Nature. https://link.springer.com/journal/44163/submission-guidelines. Accessed 25 May 2026.
2. Publishing OECD. Paris, 2025. https://www.oecd.org/en/publications/states-of-fragility-2025_81982370-en.html
3. National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1, 2023. 10.6028/NIST.AI.100-1
4. National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. NIST AI 600-1, 2024. 10.6028/NIST.AI.600-1
5. International Organization for Standardization. ISO/IEC 42001:2023 Information technology - Artificial intelligence - Management system. Geneva: ISO; 2023.

6. European Union. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence. Official J Eur Union, 2024.
7. UNESCO. Recommendation on the Ethics of Artificial Intelligence. UNESCO, Paris, 2021.
8. OECD AI Principles OECD, Adopted. 2019, updated 2024. https://www.oecd.org/en/topics/ai-principles.html
9. Inter-Agency Standing Committee. Operational Guidance on Data Responsibility in Humanitarian Action. IASC. 2021; updated 2023.
10. United Nations Office for the Coordination of Humanitarian Affairs. OCHA Data Responsibility Guidelines. OCHA Centre for Humanitarian Data; 2021.
11. World Bank. GovTech Maturity Index. World Bank Global Program on GovTech and Public Sector Innovation. https://www.worldbank.org/en/programs/govtech/gtmi
12. Sambasivan N, Kapania S, Highfill H, Akrong D, Paritosh P, Aroyo LM. 'Everyone wants to do the model work, not the data work': Data Cascades in High-Stakes AI. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021. 10.1145/3411764.3445518
13. Selbst AD, boyd, Friedler d, Venkatasubramanian SA, Vertesi S. J. Fairness and Abstraction in Sociotechnical Systems. Proceedings of the Conference on Fairness, Accountability, and Transparency. 2019:59–68. 10.1145/3287560.3287598
14. Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, Spitzer E, Raji ID, Gebru T. Model Cards for Model Reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency. 2019:220–229. 10.1145/3287560.3287596
15. Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, Daume H III, Crawford K. Datasheets for Datasets. Commun ACM. 2021;64(12):86–92. 10.1145/3458723.
16. Shneiderman B. Human-Centered Artificial Intelligence: Reliable, Safe and Trustworthy. Int J Hum Comput Interact. 2020;36(6):495–504. 10.1080/10447318.2020.1741118.
17. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell. 2019;1:389–99. 10.1038/s42256-019-0088-2.
18. Naser MZ. A review of machine learning with small and limited data. J Big Data. 2026;13:18. 10.1186/s40537-025-01346-9.
19. De-Arteaga M, Herlands W, Neill DB, Dubrawski A. Machine learning for the developing world. ACM Trans Manage Inform Syst. 2018;9(2):1–14. 10.1145/3210548.
20. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. ACM-CSUR. 2021;54(6):1–35. 10.1145/3457607.