

THE XAI-NEXUS FRAMEWORK: BRIDGING GOVERNANCE GAPS THROUGH EXPLAINABLE AI FOR RESILIENT WATER MANAGEMENT

LE CADRE XAI-NEXUS : COMBLER LES LACUNES DE GOUVERNANCE GRÂCE À L'IA EXPLICABLE POUR UNE GESTION RÉSISTANTE DE L'EAU

Behzad Navidi Nassaj¹ And Narges Zohrabi^{2*}

ABSTRACT

Artificial intelligence is increasingly deployed in water governance to process complex, uncertain, and multi-scalar environmental information. Yet predictive capability alone does not ensure governance relevance. In water management, technical outputs must also be interpretable, communicable, defensible, and institutionally actionable across settings marked by hydroclimatic uncertainty, competing stakeholder priorities, and public accountability. This paper addresses that challenge by proposing XAI-Nexus, a conceptual framework for bridging explainable artificial intelligence, semantic mediation, and human validation in resilient water governance. The framework is developed through an integrative and concept-building research design that draws on literature from explainable AI, responsible AI governance, large language model mediation, and nexus-oriented water decision support. XAI-Nexus is organized around five interdependent layers: evidence input, XAI interpretation, semantic mediation, human validation and governance review, and decision-support packaging. Together, these layers transform technical evidence including hydrological models, scenario outputs, and stakeholder inputs into governance-ready knowledge that is traceable, reviewable, and audience-responsive. To support implementation, the framework is complemented by three operational instruments: a decision documentation protocol, a trust-transparency assessment matrix, and a phased implementation roadmap. The paper argues that the central challenge in contemporary water governance is not only uncertainty or analytical complexity, but a crisis of decidability in which knowledge remains immobilized between analysis and action. XAI-Nexus is therefore advanced not as an automated governor, but as accountable decision infrastructure that enables explanation, translation, validation, and documentation. The study contributes a governance-oriented understanding of AI in which transparency and trust are treated as operational capacities essential to resilient and defensible water management.

Keywords: Artificial intelligence; water governance; water management; stakeholder priorities; analytical complexity.

RESUME

L'intelligence artificielle est de plus en plus déployée dans la gouvernance de l'eau pour traiter des informations environnementales complexes, incertaines à échelle multiple. Cependant, la capacité prédictive à elle seule ne garantit pas la pertinence de la gouvernance. Dans la gestion de l'eau, les résultats techniques doivent également être interprétables, communicables, justifiables et, actionnables sur le plan institutionnel, dans des contextes marqués par l'incertitude hydro climatique, les priorités concurrentes des parties prenantes et la responsabilité publique. Cet article relève ce défi en proposant XAI-Nexus, un cadre conceptuel pour relier l'intelligence artificielle explicable, la médiation sémantique et la validation humaine dans la gouvernance résistante de l'eau. Le cadre est développé à travers une conception de recherche intégrative et conceptuelle qui s'appuie sur la littérature issue de l'IA explicable,

¹ Ph.D. in Water Resources Engineering, Executive Secretary of ICID IRPID-IRN Email: Behzad.Navidi@gmail.com

^{2*} Associate Professor, Department of Water Science and Engineering, Ahv.C., Islamic Azad University, Ahvaz, Iran. Email: NargesZohrabi@gmail.com (Corresponding author)

la gouvernance responsable de l'IA, la médiation des grands modèles de langage et de l'aide à la décision axée sur les liens dans le domaine de l'eau. XAI-Nexus est organisé autour de cinq couches interdépendantes : l'apport de preuves, l'interprétation XAI, la médiation sémantique, la validation humaine et l'examen de la gouvernance, ainsi que le packaging d'aide à la décision. Ensemble, ces couches transforment des preuves techniques, y compris des modèles hydrologiques, des résultats de scénarios et des contributions des parties prenantes en connaissances prêtes pour la gouvernance, facile à suivre, révisables et adaptées au public. Pour soutenir la mise en œuvre, le cadre est complété par trois instruments opérationnels : un protocole de documentation des décisions, une matrice d'évaluation de la confiance et de la transparence et une feuille de route de mise en œuvre progressive. L'article soutient que le défi majeur de la gouvernance contemporaine de l'eau n'est pas seulement l'incertitude ou la complexité analytique, mais une crise de prise de décision dans laquelle les connaissances restent immobilisées entre analyse et action. XAI-Nexus est donc avancé non pas en tant que système de gouvernance automatisé, mais en tant qu'infrastructure de décision responsable permettant l'explication, la traduction, la validation et la documentation. L'étude apporte une compréhension de l'IA axée sur la gouvernance, dans laquelle la transparence et la confiance sont considérées comme des capacités opérationnelles essentielles à une gestion résistante et justifiable de l'eau.

Mots-clés: Intelligence artificielle; Gouvernance de l'eau; Gestion de l'eau; Priorités des parties prenantes; Complexité analytique.

1. INTRODUCTION

Artificial intelligence is increasingly being positioned as a strategic resource in water governance, particularly in contexts that require rapid interpretation of complex, uncertain, and high-volume environmental data (Reichstein et al., 2019; Kratzert et al., 2018; Nearing et al., 2021; Shen et al., 2021). Yet predictive performance alone does not ensure governance usefulness. In institutional settings, model outputs must not only be accurate, but also interpretable, communicable, and defensible across diverse actors involved in planning, regulation, and public accountability. This challenge is especially significant in water governance, where decisions are shaped by hydroclimatic uncertainty, competing priorities, and the need to justify actions within administrative and social frameworks (IPCC, 2023; WMO, 2024a; UNESCO & UN-Water, 2024).

Explainable artificial intelligence (XAI) has emerged as one response to this problem by making model behavior more transparent and technically inspectable (Ribeiro et al., 2016; Lundberg & Lee, 2017; Guidotti et al., 2018; Arrieta et al., 2020; Molnar, 2022). However, technical explanation does not automatically produce governance relevance. Feature importance scores, local explanations, or model diagnostics may clarify how a system reaches an output, yet still remain inaccessible or insufficient for decision-makers who must interpret those outputs in institutional, regulatory, and public terms. This gap reflects a broader challenge long identified in the literature on usable science and knowledge-to-action interfaces (Cash et al., 2003; Lemos et al., 2012; Kirchhoff et al., 2013; Meadow et al., 2015; Pahl-Wostl, 2019). At the same time, large language models (LLMs) offer new possibilities for translating technically explained outputs into more intelligible and context-sensitive forms (Bommasani et al., 2021; Brown et al., 2020; Bubeck et al., 2023). Their role, however, should not be understood as autonomous decision-making or epistemic authority, but as supervised semantic mediation that helps render model outputs usable for governance deliberation, while remaining subject to risks of hallucination, bias, and unsupported inference (Bender et al., 2021; Weidinger et al., 2021; Ji et al., 2023; NIST, 2024; Floridi & Cowls, 2019).

Against this background, this article proposes XAI-Nexus, a conceptual framework designed to connect technical explainability with governance-oriented interpretation and human validation. Figure 1 illustrates the conceptual logic of XAI-Nexus as a bridging framework between technical knowledge production and decidable water governance, highlighting how explainability, semantic translation, and human validation address key barriers to institutional usability.

The framework integrates three complementary functions: XAI as a mechanism of transparency, LLMs as instruments of semantic mediation, and human oversight as the basis for institutional review and accountability. Rather than treating AI outputs as self-sufficient recommendations, XAI-Nexus repositions them as components of a traceable decision-support process in which assumptions, uncertainties, and trade-offs can be translated, examined, and justified. This orientation is particularly relevant in water–energy–food–ecosystem (WEFE) nexus settings, where resource decisions generate interdependent social, environmental, and sectoral consequences (Hoff, 2011; Rasul & Sharma, 2016; Albrecht et al., 2018; D’Odorico et al., 2018; Simpson & Jewitt, 2019). In such contexts, the governance challenge is not merely to generate technically robust predictions, but to make those predictions interpretable, contestable, and institutionally usable across interconnected water, food, energy, and ecosystem domains.

The novelty of XAI-Nexus lies in its attempt to move beyond model-centered explainability toward governance usability. Specifically, the framework contributes by: (1) linking XAI outputs to semantically mediated interpretation for different governance audiences; (2) embedding human validation as a non-optional layer of review and responsibility; and (3) structuring AI-assisted outputs as traceable decision-support packages rather than automated decisions. On this basis, the article argues that resilient water governance requires not only transparent models, but institutional mechanisms that transform technical explanation into accountable and practically usable knowledge. In this sense, the framework responds to growing calls for human-centered, transparent, and accountable AI in public-sector decision contexts (OECD, 2019; UNESCO, 2021; NIST, 2023, 2024; OECD, 2024; European Union, 2024).

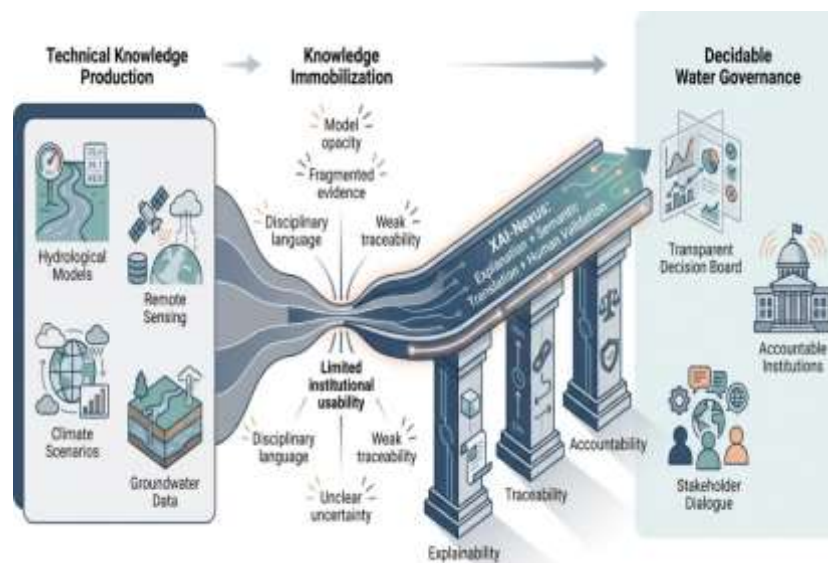


Figure 1. Conceptual architecture of XAI-Nexus for transforming technical WEFE Nexus knowledge into governance-usable and accountable decision support.

2. MATERIALS AND METHODS

2.1. Research Design

This study adopts a conceptual and framework-building research design to develop a governance-oriented architecture for explainable, traceable, and institutionally usable AI-assisted decision support in resilient water management. Rather than testing a single predictive model, the study synthesizes advances in explainable artificial intelligence, responsible AI governance, large language model mediation, and WEFE nexus decision support into a structured methodological framework. The aim is to specify how technical evidence can be transformed into governance-ready outputs while preserving human judgment, institutional accountability, and stakeholder contestability.

The methodological logic is constructive and integrative. It proceeds through four steps: first, identifying knowledge immobilization as a central governance failure in nexus-based water decision settings where analytical outputs remain disconnected from institutional action; second, deriving design requirements from recent literature on XAI, AI governance, public-sector accountability, and environmental decision support; third, integrating these requirements into an operational architecture linking technical explanation with institutional decision workflows; and fourth, operationalizing the architecture through three instruments: a decision documentation protocol, a trust–transparency assessment matrix, and a phased institutional adoption roadmap. This approach aligns with recent calls for socio-technical and human-centered AI systems in public decision domains characterized by uncertainty, multi-actor trade-offs, and accountability requirements (UNESCO, 2024; OECD, 2024; NIST, 2024; FAO, 2024; European Union, 2024).

2.2. Conceptual Boundary and Design Principles

XAI-Nexus is designed for water governance contexts shaped by hydrological uncertainty, cross-sector interdependencies, and institutional scrutiny. It is particularly relevant where decisions cannot be reduced to a single objective function and where analytical outputs must be interpreted by diverse audiences, including policymakers, basin managers, technical agencies, stakeholders, and oversight bodies. Potential applications include drought prioritization, irrigation allocation, groundwater abstraction management, environmental flow protection, flood-risk planning, crop-pattern adjustment, and basin-level WEFE trade-off assessment.

The framework rests on three assumptions. First, technical evidence already exists in the form of model outputs, scenario analyses, monitoring data, expert judgment, or integrated assessments. Second, these inputs are relevant for decision-making but are not yet sufficiently interpretable, translatable, or institutionally documented for direct governance use. Third, final decisions remain human and institutional responsibilities. XAI-Nexus is therefore not an automated governor, but an augmentation layer positioned between technical analysis and formal decision processes.

This distinction is methodologically essential. Recent responsible AI guidance consistently emphasizes that public-sector AI applications should preserve human oversight, role clarity, documentation discipline, and reviewability, particularly where decisions affect livelihoods, public goods, or environmental rights (NIST, 2024; OECD, 2024; UN Global Digital Compact, 2024; European Union, 2024). XAI-Nexus is therefore explicitly structured to support, not replace, administrative and deliberative judgment.

Six principles guide the framework: transparency, traceability, audience-sensitive explainability, human oversight and contestability, uncertainty disclosure, and frugal implementability. These principles reflect converging demands across the responsible AI, public-sector accountability, and environmental decision-support literatures (NIST, 2024; OECD, 2024; UNESCO, 2024; FAO, 2024). Transparency requires that data sources, assumptions, uncertainty sources, and explanation outputs be inspectable at appropriate levels of detail. Traceability links recommendations to evidence generation, model versions, explanation logs, LLM transformations, expert review, and final decision records. Audience-sensitive explainability distinguishes technical model interpretation from governance-oriented explanation. Human oversight ensures that machine-generated interpretations remain reviewable, correctable, and rejectable. Uncertainty disclosure translates confidence levels, scenario divergence, and knowledge gaps into decision-relevant terms. Finally, frugal implementability recognizes the limited financial, technical, and institutional capacity of many water agencies and therefore emphasizes modular, incremental, and low-cost adoption.

2.3. Structural Architecture of XAI-Nexus

The XAI-Nexus framework consists of linked layers that convert technical evidence into governance-ready decision-support packages (Figure 2).

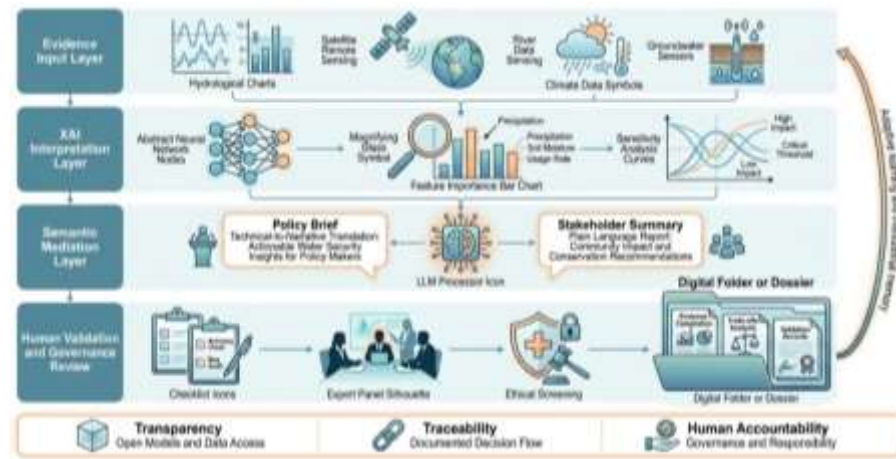


Figure 2. Architecture of the XAI-Nexus framework.

2.3.1. Layer I: Evidence Input.

This layer includes hydrological model outputs, agro-hydrological simulations, groundwater assessments, remote-sensing products, drought indicators, allocation scenarios, crop-water productivity analyses, ecosystem indicators, expert elicitation records, and stakeholder submissions. Taken together, these inputs represent not only water-system evidence, but also the broader informational basis required to interpret interdependencies among water allocation, food production, energy demand, and ecosystem conditions within WEFE nexus governance. Its function is not only to aggregate evidence, but also to register minimum metadata, including source, spatial and temporal scale, update frequency, modelling assumptions, quality limitations, and intended decision relevance.

2.3.2. Layer II: XAI Interpretation.

The second layer applies explainability methods to analytical outputs, such as feature importance analysis, local explanations, partial dependence analysis, sensitivity analysis, counterfactual inspection, uncertainty decomposition, and scenario contribution mapping. This layer clarifies which variables, assumptions, and uncertainties shape the model output. At this stage, explanation remains technical and does not yet constitute a governance narrative. This distinction is important because recent literature has warned against equating model interpretability with social intelligibility or policy usability (UNESCO, 2024; OECD, 2024).

2.3.3. Layer III: Semantic Mediation.

The third layer uses LLMs as supervised semantic mediators. Their role is to reformulate validated technical explanations into structured, audience-specific, and traceable narratives rather than generate new evidence or substitute for the underlying model. Typical tasks include producing plain-language policy briefs, summarizing WEFE trade-offs, translating uncertainty into operational cautions, differentiating outputs for various governance audiences, and identifying where expert judgment remains necessary. To reduce hallucination and unsupported inference, LLM use is bounded by controlled inputs, prompt templates, evidence-linked information blocks, and mandatory expert review.

2.3.4. Layer IV: Human Validation and Governance Review.

Outputs from the semantic mediation layer are reviewed by domain experts and institutional actors. Validation assesses fidelity to the original evidence, adequacy of explanation, explicitness of trade-offs, correctness of stakeholder implications, and suitability of the recommendation format. Disagreements are documented rather than erased, since contestation may itself be decision-relevant in complex nexus settings.

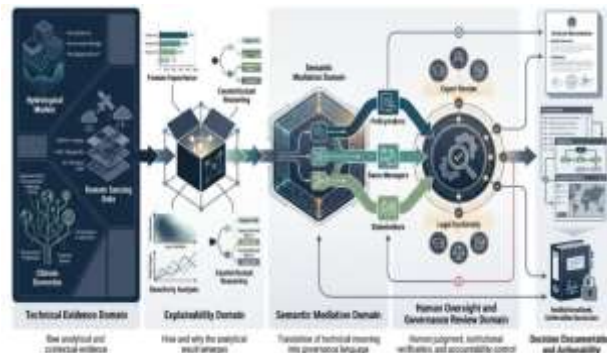
2.3.5. Layer V: Decision-Support Package.

The final output is a traceable decision-support package rather than a single recommendation. It includes the underlying evidence, XAI findings, key assumptions, uncertainty and sensitivity statements, WEFE trade-offs, stakeholder implications, alternative options, translated narratives, expert review notes, and accountability annotations. This package is intended to be operationally useful, reviewable ex post, and more defensible under institutional scrutiny than either raw model output or unstructured expert judgment alone.

2.4. Functional Logic

As illustrated in Figure 3, the operational logic of XAI-Nexus is organized as a staged transformation in which technical evidence is first subjected to explainability extraction, then translated into governance-relevant language, reviewed through human validation, and finally consolidated into decision documentation.

This sequence prevents a common governance failure: moving directly from complex model output to policy recommendation without intermediate explanation, translation, and review. This risk is particularly significant in WEFE nexus decision settings, where seemingly efficient water decisions may produce unintended consequences for agricultural productivity, energy consumption, or ecological integrity if cross-sectoral trade-offs are not explicitly interpreted. In many decision settings, technical teams generate outputs while institutional actors receive only compressed conclusions, weakening trust, obscuring assumptions, and limiting accountability.



XAI-Nexus therefore inserts an explicit interpretive chain between analytics and action. Each transition functions as a governance checkpoint. If, for example, an LLM-generated policy summary cannot be traced back to validated XAI outputs, the recommendation should not proceed in finalized form. In this sense, the framework operates as both a communication architecture and a quality-assurance mechanism.

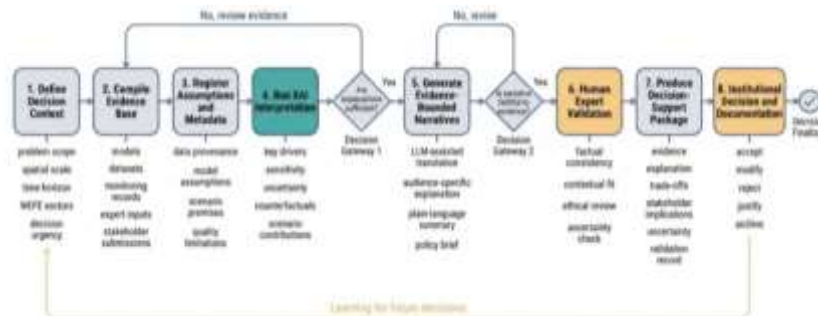


Figure 3. Operational workflow of XAI-Nexus.

2.5 Illustrative Use Case: Drought-Based Irrigation Allocation

To clarify the practical relevance of XAI-Nexus, consider a drought-based irrigation allocation scenario in a semi-arid basin where water authorities must prioritize limited water releases among competing agricultural districts while maintaining minimum environmental flows and avoiding excessive energy use for groundwater pumping. This type of allocation problem is emblematic of WEFE nexus governance, because decisions over water distribution simultaneously affect food production, energy use for pumping and delivery, and ecosystem resilience. The technical evidence available to decision-makers may include drought indicators, hydrological forecasts, crop-water productivity estimates, groundwater stress levels, energy requirements for pumping, and stakeholder submissions. Such inputs are consistent with the evidence types considered in the XAI-Nexus architecture, including drought indicators, allocation scenarios, crop-water productivity analyses, ecosystem indicators, and stakeholder records.

In a conventional AI-assisted setting, a predictive model may recommend reducing water allocation to a specific irrigation district by 25 percent because the model estimates lower marginal productivity under projected drought conditions. Without an explanation layer, however, this recommendation remains difficult to justify institutionally. Farmers may perceive the recommendation as arbitrary, basin authorities may be unable to explain why one district is prioritized over another, and environmental agencies may question whether ecological flow requirements have been sufficiently considered. In this case, the AI output may be technically accurate but governance-inusable because it does not provide a transparent account of the evidence, assumptions, trade-offs, or institutional reasoning behind the recommendation. Within XAI-Nexus, the same model output would first be processed through an XAI interpretation layer. For example, feature attribution or counterfactual explanation could show that the recommended allocation reduction is mainly driven by three factors: severe soil-moisture deficit, low crop-water productivity, and high groundwater stress in the affected district. The explanation could also identify conditions under which the recommendation would change, such as improved precipitation forecasts, a shift to less water-intensive crops, or temporary compensation through alternative water sources. This transforms the AI output from a single prescriptive recommendation into an interpretable evidence structure.

The LLM-based semantic mediation layer would then translate this technical explanation into audience-specific governance narratives. For basin managers, the explanation may be framed in terms of drought-risk reduction and system-level allocation efficiency. For farmers, it may emphasize the specific agronomic and hydrological factors affecting the district. For environmental regulators, it may clarify whether minimum ecological flow requirements are protected. For energy planners, it may identify whether reduced surface-water allocation could unintentionally increase groundwater pumping and energy demand. In this way, the LLM does not make the decision; rather, it converts XAI outputs into context-sensitive explanations that can be reviewed, contested, and refined by human actors. A possible governance prompt for the semantic mediation layer could be: “Translate the XAI explanation of the irrigation allocation model into a decision-support note for basin-level governance. Identify the main factors driving the recommended allocation reduction, explain the expected water, food, energy, and environmental trade-offs, distinguish model-based evidence from assumptions, and formulate separate summaries for water managers, farmers, environmental regulators, and energy planners. Do not generate a final decision; provide a transparent explanation for human review.” The resulting decision-support package would include the model output, XAI findings, WEFE trade-off register, stakeholder-specific narratives, validation notes from domain experts, and a final decision record. Compared with a direct AI recommendation, this process makes the basis of the decision more transparent, contestable, and institutionally accountable. The use case therefore illustrates the central claim of XAI-Nexus: the value of explainable AI in water and WEFE governance does not lie only in explaining model behavior, but in converting technical explanations into validated and traceable decision-support knowledge. This distinction is summarized in Figure 4, which contrasts a conventional black-box workflow with the XAI-Nexus pathway. The figure shows that the added value of XAI-Nexus does not lie in replacing technical modelling, but in converting model outputs into transparent, explainable, stakeholder-intelligible, and institutionally accountable decision-support packages.

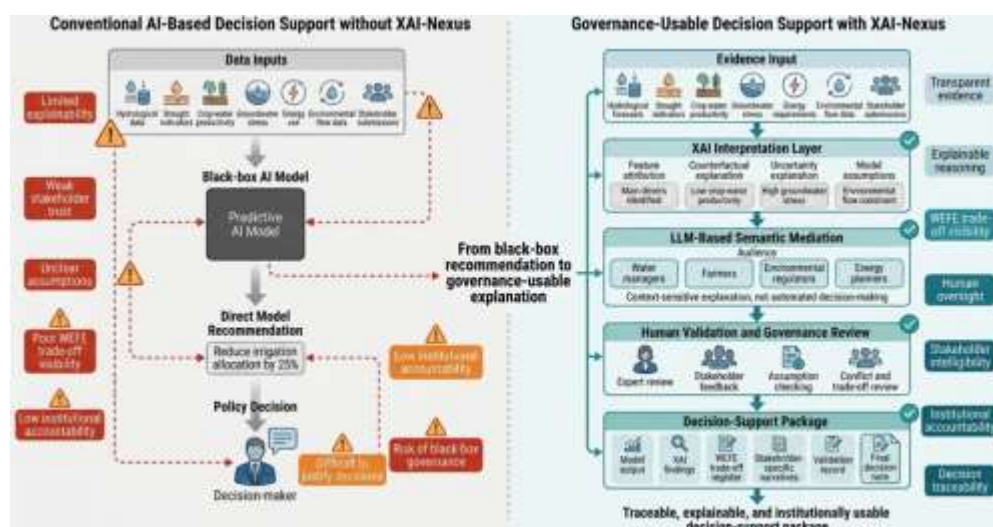


Figure 4. From black-box model to governance-usable explanation: an illustrative comparison of conventional decision support and the XAI-Nexus approach in WEFE Nexus governance.

2.6. Operational Instruments

To support implementation, the framework includes three operational instruments.

First, the decision documentation protocol ensures traceability through eight minimum elements: decision context, evidence inventory, analytical assumptions, XAI findings, WEFE trade-off register, audience-specific narrative, validation record, and final decision note. This protocol functions as a governance memory device by making decisions legible not only at the moment of action but also for later auditing, learning, and adaptive revision. In climate-sensitive water governance, where conditions and assumptions may shift quickly, such documentation is critical for institutional learning and resilience. Second, the trust–transparency assessment matrix evaluates whether an AI-assisted decision-support process is sufficiently robust for governance use (Figure 5). As presented in Table 1, the framework assesses five dimensions: evidence transparency, explanation quality, translation fidelity, institutional accountability, and stakeholder intelligibility.

The matrix can be used *ex ante* to assess readiness or *ex post* to review the quality of completed decision-support packages. Its main methodological function is to move trust from an abstract aspiration to an assessable governance property.

	Low Readiness	Moderate Readiness	High Readiness
Evidence Transparency	undocumented data	partial metadata	traceable sources and assumptions
Explanation Quality	black-box output	basic feature explanation	sensitivity, uncertainty, and counterfactual interpretation
Translation Fidelity	unsupported narrative	reviewed summary	evidence-linked, validated, and audience-specific explanation
Institutional Accountability	unclear ownership	partial approval pathway	named reviewers, approval record, and audit trail
Stakeholder Intelligibility	technical jargon	simplified summary	plain-language trade-off and uncertainty explanation

The matrix evaluates whether AI-assisted outputs are sufficiently transparent, explainable, traceable, and usable for governance.

Figure 5. Trust–transparency assessment matrix for evaluating whether AI-assisted decision-support outputs are sufficiently explainable, traceable, accountable, and intelligible for WEFE governance use.

Third, the phased implementation roadmap (Table 2) supports incremental adoption across diverse institutional settings. It begins with diagnostic preparation, followed by minimal viable implementation around a specific use case, controlled integration of LLM mediation, institutionalization of templates and review procedures, and adaptive learning and scaling. This phased approach reflects frugal innovation logic by reducing implementation barriers, avoiding overdependence on advanced technical infrastructure, and making adoption feasible even in agencies with limited computational and financial capacity.

Table 1. Trust–transparency assessment matrix for evaluating the governance readiness of AI-assisted decision-support processes.

Dimension	Key Question	Indicative Criteria
Evidence transparency	Are data sources and model assumptions visible?	Source metadata, assumption log, update history
Explanation quality	Is the recommendation technically interpretable?	XAI outputs, sensitivity clarity, confidence statement
Translation fidelity	Does the governance narrative accurately reflect evidence?	Evidence-linked summary, no unsupported claims, review approval
Institutional accountability	Are roles, approvals, and decision ownership clear?	Named validators, approval pathway, decision note
Stakeholder intelligibility	Can affected actors understand implications and trade-offs?	Plain-language version, trade-off summary, uncertainty explanation

Table 2. Phases of LLM Integration in WEFE Nexus Governance Decision-Making

Phase	Title	Key Activities
Phase 1	Diagnostic Preparation	Identification of governance bottlenecks, existing models, datasets, and decision workflows.
Phase 2	Minimal Viable Implementation	Pilot implementation for a single decision use case using existing analytical tools.
Phase 3	Controlled Integration of LLM Mediation	Introduction of LLM-supported tasks with standardized prompts and expert verification.
Phase 4	Institutionalization	Formalization of templates, review procedures, staff roles, and documentation standards.
Phase 5	Adaptive Learning and Scaling	Refinement of prompts and communication practices based on feedback and expansion to other domains.
Implementation Logic	Frugal Innovation Approach	Low-cost and scalable implementation suitable for institutions with limited resources.

3. DISCUSSION

3.1. From Predictive Performance to Governance Usability

The central contribution of XAI-Nexus is to shift the evaluation of AI in water management from prediction alone toward governance usability. AI systems are commonly assessed through accuracy, error reduction, computational speed, or forecast skill. Although these metrics remain necessary, they are insufficient in public decision environments where outputs must also be justified, communicated, reviewed, and contested. A model may be technically accurate yet institutionally unusable if its reasoning cannot be explained to ministries, basin authorities, irrigation agencies, farmers, or environmental regulators.

XAI-Nexus addresses this gap by treating analytical validity as necessary but not sufficient. Decision support becomes governance-ready only when technical reasoning is inspectable, translated into context-appropriate language, and embedded in a documented review process. This is especially

important in WEFE nexus settings, where decisions generate trade-offs across sectors, institutions, and social groups.

3.2. *XAI as a Governance Instrument*

A second implication of the framework is that XAI should be understood not only as a technical method for model interpretation, but also as a governance instrument. In XAI-Nexus, explanation is linked directly to accountability, contestability, and institutional trust. By showing which variables shaped a recommendation, how sensitive results are to assumptions, and under what conditions alternatives would emerge, XAI expands the space for informed administrative judgment.

This reframing matters because many failures in water governance are not caused by insufficient computation, but by insufficiently defensible reasoning. Public agencies often face questions that are both analytical and justificatory: why a basin was prioritized, why a crop restriction was recommended, or why a drought measure was triggered at a particular moment. A traceable XAI layer makes such questions answerable in a more disciplined and reviewable manner.

3.3. *The Added Value and Limits of LLM-Based Semantic Mediation*

The most novel element of XAI-Nexus is the use of LLMs as semantic mediators between technical explanation and governance communication. In multi-actor water systems, the same analytical result must often be communicated differently to policymakers, technical managers, local authorities, civil society groups, and affected users. LLMs can support this differentiation by converting validated technical outputs into audience-specific narratives, provided that their role remains procedurally bounded.

The framework therefore avoids treating LLMs as autonomous decision-makers or neutral interpreters. They are probabilistic language systems that may introduce distortion, oversimplification, or unwarranted confidence. For this reason, semantic mediation in XAI-Nexus is constrained by three safeguards: evidence-bounded input, structured prompting, and mandatory human review. In this way, the framework uses the communicative capacity of LLMs without granting them epistemic authority.

3.4. *Relevance for Low-Capacity and Resource-Constrained Institutions*

A practical strength of XAI-Nexus is its compatibility with frugal institutional innovation. Many water agencies cannot build expensive end-to-end AI ecosystems, but they often already use hydrological models, spreadsheet-based allocation tools, remote-sensing products, and expert assessment routines. The framework does not require replacing these systems. Instead, it adds explainability, translation, and documentation layers that can be implemented incrementally.

This is important because the governance gap in many settings is not the absence of analytics, but the absence of mechanisms that make analytics understandable, reviewable, and usable. A modular architecture is therefore more realistic than a maximalist digital transformation approach, particularly for agencies facing budgetary, staffing, or technical constraints.

3.5. *Practical Outputs and Governance Instruments*

Building on the preceding architectural and operational discussion, the tangible outputs of the XAI-Nexus framework can be summarized in four main components. First, the framework provides a five-layer governance-oriented architecture that connects evidence input, XAI interpretation, semantic mediation, human validation, and decision-support packaging. Second, it offers an operational workflow for transforming technical evidence into institutionally usable and traceable decision support. Third, it specifies a decision documentation protocol that records the decision context, evidence base, analytical assumptions, XAI findings, WEFE trade-offs, audience-specific narratives, validation records, and final decision notes. Fourth, it introduces two supporting governance instruments: a trust–transparency

assessment matrix for evaluating the quality and accountability of AI-assisted interpretation, and a phased implementation roadmap for incremental institutional adoption. Together, these outputs clarify that XAI-Nexus is not presented as an automated decision system, but as a structured decision-support infrastructure for accountable water governance. As such, the outputs of XAI-Nexus are intended not only to support water-sector decision-making, but also to improve the interpretability and accountability of decisions taken under WEFE nexus complexity.

3.6. *Limitations*

This study is conceptual and has several limitations. First, the framework has not yet been tested through real-world basin case studies or comparative institutional pilots. Its feasibility, transaction costs, and user acceptance therefore require empirical validation. Second, XAI-Nexus does not prescribe a universal XAI technique or LLM configuration; implementation will depend on data quality, modelling traditions, sectoral priorities, and regulatory conditions. Third, while the framework can strengthen transparency and traceability, it cannot eliminate structural disagreement. Some WEFE Nexus decisions remain contested not because explanations are weak, but because interests, priorities, or risk tolerances differ.

A further limitation concerns institutional readiness. Agencies with poor record management, unclear mandates, or limited technical staff may struggle even with lightweight implementation unless capacity building is included from the outset. XAI-Nexus should therefore be understood not as a plug-and-play tool, but as an organizational innovation requiring procedural adaptation.

3.7. *Future Research*

Future research should proceed in three directions. First, empirical pilot studies are needed in concrete WEFE decision settings, such as irrigation reallocation, drought emergency planning, groundwater licensing, and environmental flow regulation. Second, comparative studies should evaluate whether different forms of explanation and semantic mediation improve trust, speed, quality, and legitimacy of decision processes across institutional contexts. Third, benchmark protocols are needed to validate LLM-assisted governance narratives against source evidence, particularly in multilingual or high-stakes environments.

Further work is also needed on interface design, stakeholder-facing visualization, uncertainty communication, and integration with digital public infrastructure. Such research would help move XAI-Nexus from a conceptual governance architecture toward a deployable decision-support standard.

Conclusion

This article argued that a central challenge in contemporary water governance is not only scarcity, uncertainty, or modelling complexity, but a deeper crisis of decidability. In WEFE nexus contexts, decision-makers often confront abundant technical evidence without sufficient mechanisms to transform it into explanations that are understandable, defensible, traceable, and actionable. This condition was conceptualized as knowledge immobilization.

In response, the article proposed XAI-Nexus as a human-centered and governance-oriented architecture for resilient water management. The framework integrates explainable AI, LLM-based semantic mediation, and human validation into a structured process that converts analytical outputs into traceable decision-support packages. Its purpose is not to automate authority, but to improve how model-based knowledge enters governance.

The framework contributes to the literature in three ways. First, it reframes the implementation gap in water governance as a problem of decidability and knowledge mobility. Second, it positions XAI as a governance instrument for accountability and trust rather than merely a technical interpretability tool. Third, it introduces a disciplined role for LLMs as semantic interfaces operating under institutional

oversight. Through the decision documentation protocol, the trust–transparency assessment matrix, and the phased implementation roadmap, the article also provides practical instruments for operationalization.

The broader implication is that resilient water governance in the AI era will depend less on generating more predictions and more on the capacity to explain, translate, validate, and document the knowledge institutions already possess. XAI-Nexus is proposed as a step in that direction: not AI as a replacement for governance, but AI as accountable infrastructure for explanation, justification, and informed collective action.

REFERENCES

- Albrecht, T. R., Crootof, A., & Scott, C. A. (2018). The water–energy–food nexus: A systematic review of methods for nexus assessment. *Environmental Research Letters*, 13(4), 043002. <https://doi.org/10.1088/1748-9326/aaa9c6>
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence: Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). <https://doi.org/10.1145/3442188.3445922>
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., et al. (2021). On the opportunities and risks of foundation models. *arXiv*. <https://arxiv.org/abs/2108.07258>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., et al. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. *arXiv*. <https://arxiv.org/abs/2303.12712>
- Cash, D. W., Clark, W. C., Alcock, F., Dickson, N. M., Eckley, N., Guston, D. H., Jäger, J., & Mitchell, R. B. (2003). Knowledge systems for sustainable development. *Proceedings of the National Academy of Sciences*, 100(14), 8086–8091. <https://doi.org/10.1073/pnas.1231332100>
- D’Odorico, P., Davis, K. F., Rosa, L., Carr, J. A., Chiarelli, D., Dell’Angelo, J., Gephart, J., MacDonald, G. K., Seekell, D. A., Suweis, S., & Rulli, M. C. (2018). The global food–energy–water nexus. *Reviews of Geophysics*, 56(3), 456–531. <https://doi.org/10.1029/2017RG000591>
- European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*.
- Floridi, L., & Cows, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.8cd550d1>
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining bla
- Hoff, H. (2011). Understanding the nexus: Background paper for the Bonn 2011 Conference: The Water, Energy and Food Security Nexus. Stockholm Environment Institute.
- IPCC. (2023). Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC. <https://doi.org/10.59327/IPCC/AR6-9789291691647>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1–38. <https://doi.org/10.1145/3571730>
- Kirchhoff, C. J., Lemos, M. C., & Dessai, S. (2013). Actionable knowledge for environmental decision making: Broadening the usability of climate science. *Annual Review of Environment and Resources*, 38, 393–414. <https://doi.org/10.1146/annurev-enviro-022112-112828>

- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory networks. *Hydrology and Earth System Sciences*, 22, 6005–6022. <https://doi.org/10.5194/hess-22-6005-2018>
- Lemos, M. C., Kirchhoff, C. J., & Ramprasad, V. (2012). Narrowing the climate information usability gap. *Nature Climate Change*, 2, 789–794. <https://doi.org/10.1038/nclimate1614>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, 30.
- Meadow, A. M., Ferguson, D. B., Guido, Z., Horangic, A., Owen, G., & Wall, T. (2015). Moving toward the deliberate coproduction of climate science knowledge. *Weather, Climate, and Society*, 7(2), 179–191. <https://doi.org/10.1175/WCAS-D-14-00050.1>
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1–21. <https://doi.org/10.1177/2053951716679679>
- Molnar, C. (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (2nd ed.). <https://christophm.github.io/interpretable-ml-book/>
- Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., & Gupta, H. V. (2021). What role does hydrological science play in the age of machine learning? *Water Resources Research*, 57(3), e2020WR028091. <https://doi.org/10.1029/2020WR028091>
- NIST. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-1>
- NIST. (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.600-1>
- OECD. (2019). *Recommendation of the Council on Artificial Intelligence*. OECD Legal Instruments, OECD/LEGAL/0449.
- OECD. (2024). *Recommendation of the Council on Artificial Intelligence [updated]*. OECD Legal Instruments, OECD/LEGAL/0449.
- Pahl-Wostl, C. (2019). The role of governance modes and meta-governance in the transformation towards sustainable water governance. *Environmental Science & Policy*, 91, 6–16. <https://doi.org/10.1016/j.envsci.2018.10.008>
- Rasul, G., & Sharma, B. (2016). The nexus approach to water–energy–food security: An option for adaptation to climate change. *Climate Policy*, 16(6), 682–702. <https://doi.org/10.1080/14693062.2015.1029865>
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566, 195–204. <https://doi.org/10.1038/s41586-019-0912-1>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). <https://doi.org/10.1145/2939672.2939778>
- Shen, C., Lawson, K., & Feng, D. (2021). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, 57(9), e2020WR029509. <https://doi.org/10.1029/2020WR029509>
- Simpson, G. B., & Jewitt, G. P. W. (2019). The development of the water–energy–food nexus as a framework for achieving resource security: A review. *Frontiers in Environmental Science*, 7, 8. <https://doi.org/10.3389/fenvs.2019.00008>
- UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. UNESCO.
- UNESCO & UN-Water. (2024). *United Nations World Water Development Report 2024: Water for Prosperity and Peace*. UNESCO.
- United Nations. (2024). *Global Digital Compact*. United Nations.
- Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., et al. (2021). Ethical and social risks of harm from language models. *arXiv*. <https://arxiv.org/abs/2112.04359>
- WMO. (2024a). *State of Global Water Resources 2023*. World Meteorological Organization.