International Journal of
**INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING**

IJISAE

ISSN:2147-6799 www.ijisae.org Original Research Paper

# Trustworthy AI Through End-to-End Data Lineage and Governance in Utility Data Ecosystems

**Aditya Laljibhai Patel**

**Submitted**: 03/09/2024 **Revised**: 17/10/2024 **Accepted**: 28/10/2024

**Abstract:** Water utilities and power companies using water distribution systems and electricity networks are increasingly turning to artificial intelligence (AI) to decipher data from sensors for various applications like leak detection, load forecasting, or anomaly detection. The integrity of data entering these systems is as important as model accuracy - from field sensing, to governance, provenance recording, model inference, to explainable reporting. This paper combines existing bodies of knowledge on data governance and the FAIR (Findable, Accessible, Interoperable, Reusable) data principles, explainable AI (XAI), and Internet of Things (IoT) utility monitoring to introduce a layered, multi-layered framework that connects end-to-end data lineage with trustworthy AI outcomes in utility environments. It covers centralized, distributed (blockchain-based) and hybrid governance models, discusses the importance of explainability both intrinsically and as it relates to the model, and explores the local vs. global and legacy supervisory control and data acquisition (SCADA) vs. modern analytics platform dimensions as barriers to implementation. The paper makes the case that tracking and governing the origin of AI models must be integrated into their architecture, not as a downstream compliance step, if AI for utilities is to be trusted by regulators and the public. The quantitative comparisons offered are conceptual and illustrative, and are not necessarily measures derived from a specific empirical study, but are provided as a heuristic tool for assessing governance architectures. A hybrid of centralized governance and provenance verification, which brings together centralized policy enforcement with distributed provenance checking, is the most defensible approach between accountability, scalability and cost, and the paper suggests potential avenues for further empirical testing.

***Keywords***: *trustworthy AI, data governance, data lineage, data provenance, explainable AI, FAIR principles, utility data ecosystems, Internet of Things, blockchain, smart grid, water management, algorithmic accountability*

## 1. Introduction

AI is increasingly becoming the backbone of the utility's operations, driving predictive maintenance, demand forecasting, leak detection, and grid balancing. In contrast, many consumer applications of AI have immediate implications for safety, continuity of service, and regulatory compliance, making the bar for trust in AI-based decisions higher in the context of a utility. Trustworthiness in this sense goes beyond the numerical performance of a model and includes the ability of a model to trace back the data it draws from to train and feed the model, the policies that govern how the data is collected and utilized, and the ability of the model to explain its results to the operators, regulators, and affected communities. This paper builds on the conceptual work of three interrelated streams of literature: data governance frameworks, trustworthy and explainable AI, and monitoring utilities with the Internet of Things. It builds an integrated argument that the end-to-end data lineage, understood as the ability to trace a data element from the data point at which it is captured to understanding how it is being used in an AI-driven decision, is a structural precondition to trustworthy AI use in utility settings, rather than an optional auditing feature that could be added after deployment.

*Employer: Logic Explorer LLC End Client: PG&E*

*Email Id / Password: Consult.AdityaPatel@gmail.com*

## 2. Background and Conceptual Foundations

### 2.1 Data Governance as a Precondition for Trust

Data governance is a generally accepted term for a collection of organisational policies, roles and standards that govern the creation, classification, access and retirement of data assets. Abraham, Schneider and vom Brocke (2019) put forward a conceptual approach that views data governance as a three-dimensional continuum of scope, structure and mechanisms, stating that effective data governance is more than just a technical control layer and should involve clearly assigned data governance decision rights. Al-Badi, Tarhini and Khan (2018) also studied governance models that fit the high volume, high velocity data environment, stating that conventional governance models that work well for more modest data assets are stressed by big data environments. Janssen, Brous, Estevez, Barbosa and Janowski (2020) went further in this regard, suggesting that the organisation of data is fundamental to the trustworthiness of a data AI system, a prerequisite for the institutions and public that use it. In utility organisations, governance challenges are complicated by the fact that legacy OT systems – such as SCADA systems from decades ago, not built to be interoperable – must coexist with newer IoT systems delivering sensing infrastructure. This diversity makes it necessary to address the need for governance structures suitable to a utility context that accommodate inconsistent metadata conventions, inconsistent data quality between sensor generations and inconsistent regulatory retention requirements between jurisdictions.

### 2.2 The FAIR Principles as a Data Stewardship Standard

The FAIR guiding principles were articulated by Wilkinson et al (2016) stating that scientific and organisational data should be Findable, Accessible, Interoperable and Reusable with a particular focus on the machine-actionability of the data and not just on manual human interpretation and location. Designed for scientific data stewardship, the FAIR principles readily apply to utility data ecosystems, where thousands of sensor endpoints publish heterogeneous streams that need to be discoverable and interpretable by downstream analytics systems, without having to do special work for each new data source. The use of FAIR principles for the telemetry of utilities suggests that utility streams of sensor data should be identifiable at any time, have a standard metadata schema for describing the units and calibration status, and have documented access protocols that differentiate operational, regulatory and public-facing applications.

### 2.3 Trustworthy and Explainable Artificial Intelligence

Multiple complementary frameworks describe trustworthy AI. Floridi et al. (2018) put forth an ethics framework based on the principles of beneficence, non-maleficence, autonomy, justice, and explicability as the conditions for an AI system to be said to be socially acceptable. Thiebes, Lins, and Sunyaev (2021) distilled the elements of trustworthy AI into dimensions such as fairness, transparency, and robustness, while Li et al. (2023) examined how commitments to trustworthy AI principles can be translated into actions, with many organisations stating their trustworthiness commitments but not including technical safeguards in their system designs. Similarly, Kaur, Uslu, Rittichier and Durresi (2022) performed a trustworthy study of AI, highlighting that trust cannot be determined on the basis of model outputs alone, but also needs to consider the source and quality of training data.

Explainability is another trustworthy AI cornerstone. Barredo Arrieta et al. (2020) present a generally used taxonomy in which they separated intrinsically interpretable models from post-hoc explanation methods used to explain opaque models, and further separated local explanations, which explain a single prediction, from global explanations, which explain the overall behaviour of the model. Rudin (2019) made a stronger case that the utility infrastructure decisions are plausibly part of high-stakes decisions, which should be based on inherently interpretable models and not post-hoc explanations of black-box systems, and that such post-hoc explanations may not accurately capture the reasoning that the model actually used. Linardatos et al. (2021) surveyed the expanding family of interpretability methods, and Chamola et al. (2023) examined trustworthy and explainable AI in general, highlighting that there remains a "conflict between model complexity, predictive performance and explanatory fidelity".

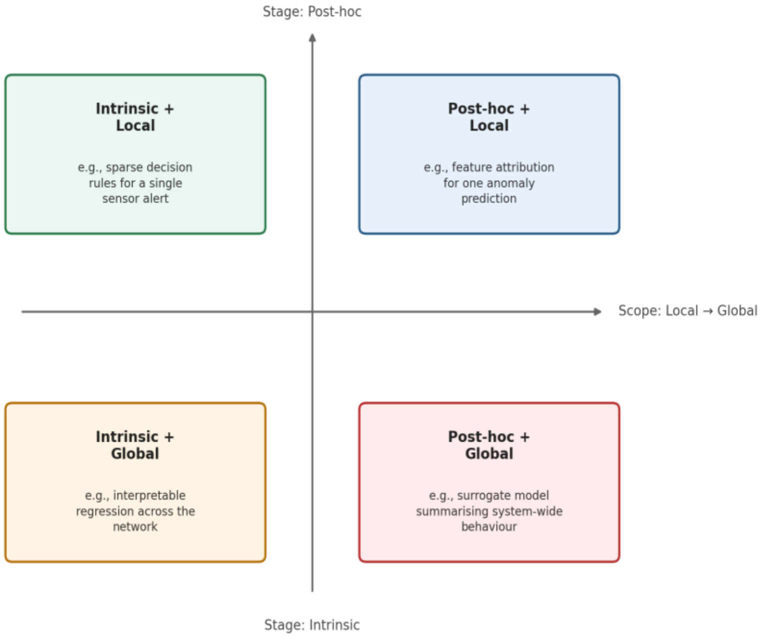


**Figure 1. Conceptual Taxonomy of Explainable AI Approaches by Stage and Scope**

*Note. Quadrant placement follows the intrinsic/post-hoc and local/global distinctions discussed by Barredo Arrieta et al. (2020). Examples are illustrative of utility-sector applications rather than reported findings from a specific cited study.*

## 3. End-to-End Data Lineage in Utility Data Ecosystems

Data lineage refers to the documentation of the history of a data element, including its source, how it has changed, and the processes it is passed to downstream. Simmhan, Plale, and Gannon (2005) introduced the concepts of data provenance in e-science—that is, the metadata about the derivation of a data product, rather than the data product itself. This separation is the essence of utility applications, where an AI prediction can only be as reliable as the list of sensor readings, quality fixes, and aggregation that led to the features that were fed into the model.

### 3.1 IoT-Enabled Data Capture in Water and Energy Utilities

The Internet of Things has moved away from simple sensor applications to internet-scale services that can be used to support complex analytics pipelines, as is evident in the case of utility telemetry, as reported by Georgakopoulos and Jayaraman (2016). In the field of water management, Andrić, Al-Ghamdi, and Koç (2022) discussed IoT solutions for smart water management systems and smart water usage, and Ahmed and Mohammed (2022) presented an IoT-based architecture specifically designed for water leakage detection, claiming that a continuous and granular water usage sensing process can not be achieved by using periodic manual water meter reading. Ryś, Pincheira, Vecchio, Giaffreda, and Kanhere (2021) took this one step further, with a water management system for precision agriculture based on blockchain technology, focusing on the trustworthiness of sensor data as both a function of the device itself and the provenance system that records sensor output. In the electricity domain, Hasan et al. (2022) summarized the applications of blockchain technology in smart grid data and energy trading, highlighting security and provenance issues with high frequency metering interfaces, while Persson, Gorton and Ali (2022) evaluated the potential of blockchain technology for smart grid architectures, showing the pros and cons of transaction throughput versus practical decentralization. Together, these results point to the problem of the IoT sensing infrastructure providing the raw material for utility AI, without on its own ensuring provenance integrity of that AI, which is necessary for trustworthy AI.

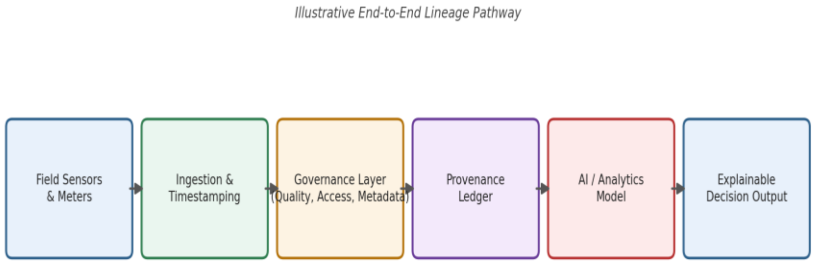


**Figure 2. Illustrative End-to-End Data Lineage Pathway for Utility AI Systems**

*Note. The pipeline stages are synthesized from governance and IoT lineage concepts discussed in the reviewed literature; the diagram is a conceptual model rather than a depiction of any single deployed system.*

### 3.2 Distributed Ledgers and Provenance Verification

Distributed ledger technologies (DLTs) have been suggested as one approach to achieving such a resilient and tamper-evident record of provenance. Since the blockchain allows records to be added to the ledger in a time-ordered sequence and the records are permanently recorded, it provides a technical foundation for confirming that a particular data item came into the pipeline at a particular time and has never been modified after that. However, both Persson et al. (2022) and Hasan et al. (2022) warned that there are also governance concerns with a DLT approach, such as who can write to the DLT and how the consensus process impacts the delay in time-critical grid operations, as well as how the computational load of consensus algorithms is balanced against the available processing power of edge deployed IoT devices.

### 3.3 Illustrative Comparison of Provenance Mechanisms

Table 1 shows that there are three conceptual approaches in the governance and blockchain literature that seek to record provenance, which are discussed here. The values are heuristic ratings and do not represent actual values or measures from any particular study cited.

| Provenance Mechanism | Tamper Resistance | Latency Suitability for Real-Time Grid Data | Governance Complexity | Relative Implementation Cost |
|---|---|---|---|---|
| Centralized audit log | Moderate | High | Low | Low |
| Permissioned blockchain ledger | High | Moderate | Moderate-High | Moderate |
| Public/permissionless blockchain | Very High | Low | High | High |

**Table 1. Illustrative Comparison of Provenance-Recording Mechanisms for Utility Data**

*Note. Ratings are conceptual heuristics synthesized from discussions in Hasan et al. (2022), Persson et al. (2022), and Rys et al. (2021), not empirical measurements extracted from those studies.*

## 4. Explainability and Trust in AI-Driven Utility Analytics

The explainability needs of utility AI systems vary based on the decision context. A post-hoc, locally scoped explanation that explains the sensor readings that caused the flag to be raised for the current trip can be supported by a model used for flagging a possible pipeline leak and sending out field crews; a model that would alert operators to load-shedding during a crisis with a globally coherent, intrinsically interpretable rationale that can be audited afterward by regulators is more plausible. Rudin (2019) suggested that in high-stakes scenarios, the reliance on post-hoc explanation of a black-box model may result in plausible and unfaithful explanations, which has direct relevance to the utility infrastructure, where a misguided explanation could be a sign of a systematic problem in the sensors or a degradation of the model. Barredo Arrieta et al. (2020) identified a large number of post-hoc explanation methods such as feature attribution and surrogate modelling, but warned that many methods are appearing and there is not yet consensus about how quality of explanations should be assessed. Linardatos et al. (2021) concluded the same, noting that most methods of interpreting models are evaluated on a limited set of benchmark data that are not representative of the complexity of infrastructure systems in operation. When applied to utility contexts, this indicates that methods tested on generic benchmarks should not be taken as a given, and must be validated for a specific domain—a step which is not done in the case of water and energy networks.

### 4.1 Privacy and Trust Trade-offs

Strobel and Shokri (2022) studied the correlation between data privacy and trustworthy machine learning, highlighting that privacy-preserving approaches, like federated learning or differential privacy, may compromise the transparency of how a data-driven model is trained, as well, creating a tension between privacy and transparency of the model training process aimed at above. This is relevant for utility data ecosystems, where data about water or electricity usage at a fine-grained level of the household is also sensitive, even if it will help in identifying anomalies and predicting demand more effectively.

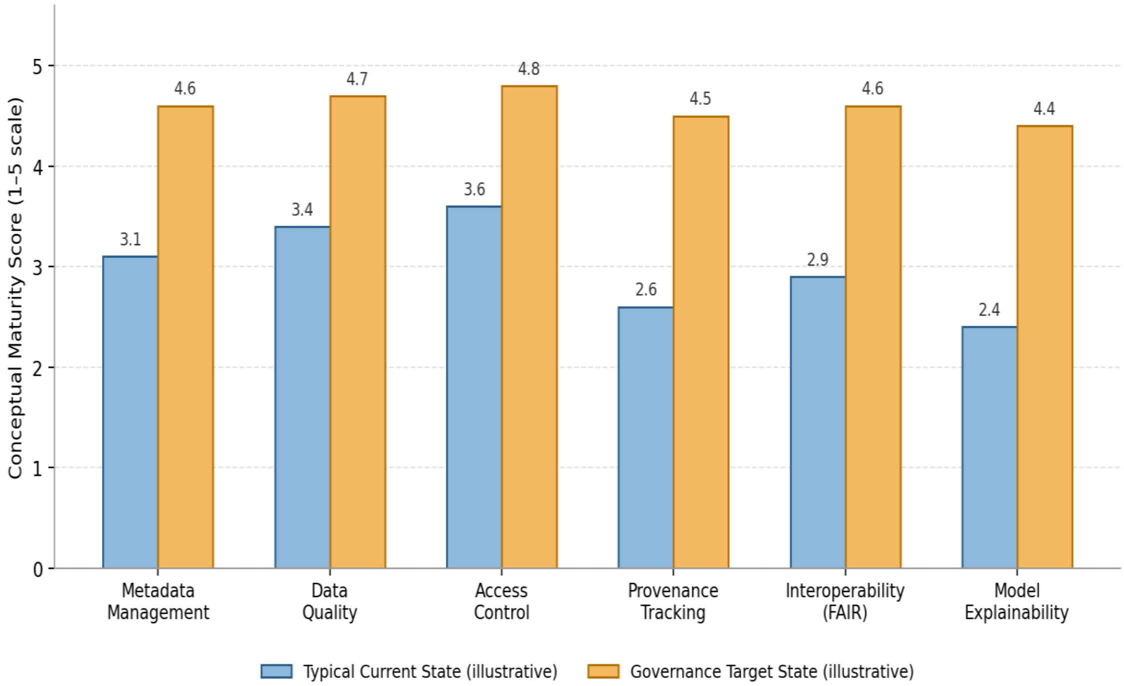


**Figure 3. Illustrative Comparison of Current and Target Governance Maturity Across Dimensions**

*Note. Scores represent an illustrative five-point conceptual maturity scale used for discussion purposes and do not represent measurements from a specific surveyed organization.*

### 4.2 Illustrative Quantitative Framing of Explainability Demand

Table 2 summarizes, in qualitative-to-ordinal terms, how explainability demands plausibly vary across representative utility AI use cases, drawing on the distinctions established in the reviewed XAI literature.

| Utility AI Use Case | Typical Decision Stakes | Explanation Scope Needed | Preferred Model Type |
|---|---|---|---|
| Leak detection alerting | Moderate | Local | Post-hoc or intrinsic |
| Predictive maintenance scheduling | Moderate-High | Local to global | Intrinsic preferred |
| Load forecasting | High | Global | Intrinsic preferred |
| Emergency load-shedding decisions | Very High | Global | Intrinsic required |
| Consumption anomaly flags to customers | Low-Moderate | Local | Post-hoc acceptable |

**Table 2. Illustrative Mapping of Explainability Requirements Across Utility AI Use Cases**

*Note. Mapping is synthesized from the stakes-sensitivity argument advanced by Rudin (2019) and the taxonomy of Barredo Arrieta et al. (2020); it is a conceptual heuristic, not an empirical survey result.*

## 5. An Integrated Governance-Lineage Framework for Trustworthy AI

To summarize the above, a layered architecture is suggested where governance and provenance functions are not only integrated in with, but not added on to, the AI modelling layer. It shows five layers: physical and sensing layer, which includes IoT devices and legacy SCADA endpoints; data governance, which includes implementing FAIR aligned metadata, access policy, and quality rules as described in Abraham et al. (2019) and Wilkinson et al. (2016); provenance and trust layer, which records immutable audit trails based on the distributed ledger discussion in Section 3.2; AI modelling layer, which performs predictive and anomaly-detection tasks; and explainability and accountability layer, which surfaces interpretable outputs to human operators, consistent with the ethical framework proposed in Floridi et al. (2018) and the accountability emphasis of Janssen et al. (2020).

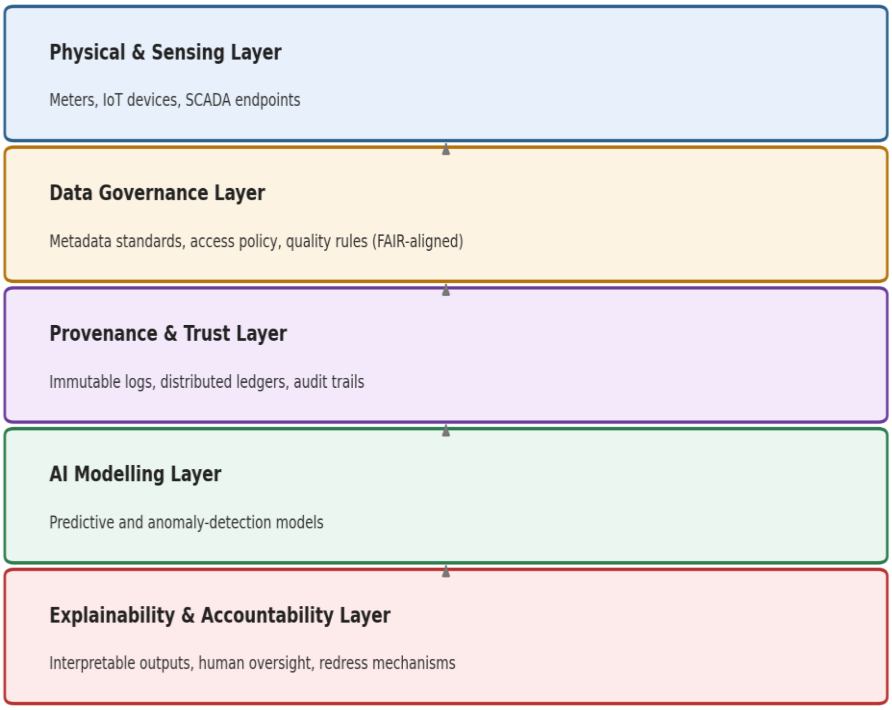


**Figure 4. Proposed Layered Governance-Lineage Architecture for Trustworthy Utility AI**

*Note. The architecture synthesizes governance, provenance, and explainability concepts drawn from the reviewed literature into a single conceptual model; it has not been empirically validated against a deployed utility system.*

One key message of this framework is that the information generated by the sensing and governance layers needs to be available to the explainability layer, enabling an explanation of the model output to be traced back to the specific sensor readings and governance decisions that led to the features that were generated. This lack of traceability can lead to explanations that could focus on the internals of the model, but fail to mention data quality or other issues that might be the actual cause of the incorrect prediction, as hinted at by Simmhan et al (2005) in their provenance-based arguments and explicitly stated in Janssen et al (2020) in their AI-specific governance framework.

### 5.1 Centralized, Distributed, and Hybrid Governance Models

The literature reviewed reveals that there are three general governance models for utility data ecosystems. In a centralized model, the governance authority and the recording of the origin of all the governance processes are focused in a single organizational function, which provides simplicity for administration but also a single point of failure and limited external auditability. In a distributed model, exemplified by blockchain-based provenance (Hasan et al. (2022), Persson et al. (2022) and Rys et al. (2021)), trust is spread between multiple participants of the ledger, resulting in a greater level of tamper resistance coupled with increased levels of complexity for coordination as well as, in some architectures, transaction latency too slow for real-time use in the grid. A hybrid approach is conceptually embedded in Di Vaio, Trujillo, D'Amore, and Palladino's (2021) overview of water governance models for sustainable development goals, in which decentralized governance was advocated, but without losing sight of coordinated policy-setting, which is the responsibility of centralized governance.

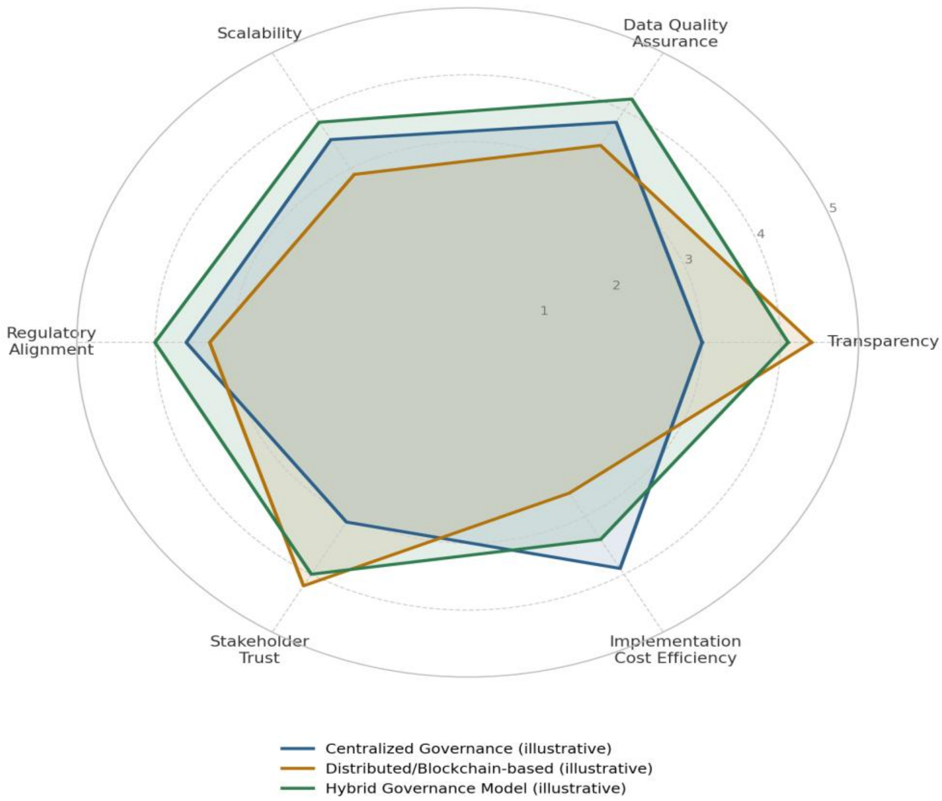


**Figure 5. Illustrative Comparison of Centralized, Distributed, and Hybrid Governance Models**

*Note. Axis scores are illustrative ratings intended to support qualitative comparison across governance archetypes discussed in Section 5.1; they are not derived from a controlled empirical evaluation.*

Table 3 summarizes this comparison in tabular form, aligning each governance archetype with the dimensions most frequently invoked across the reviewed sources.

| Governance Archetype | Transparency | Coordination Complexity | Suitability for Real-Time Operations | Regulatory Auditability |
|---|---|---|---|---|
| Centralized | Moderate | Low | High | Moderate |
| Distributed (blockchain-based) | High | High | Low-Moderate | High |
| Hybrid | High | Moderate | Moderate-High | High |

**Table 3. Illustrative Comparison of Governance Archetypes for Utility Data Ecosystems**

*Note. Synthesized from Di Vaio et al. (2021), Hasan et al. (2022), and Persson et al. (2022); ratings are qualitative and illustrative.*

## 6. Challenges and Barriers

### 6.1 Interoperability Between Legacy and Modern Systems

There is a common theme throughout the reviewed governance literature in needing to make sense of a mix of legacy SCADA infrastructure built with proprietary protocols and not well documented with metadata, with the more modern IoT and AI platforms which assume a richer, more standardized metadata. Al-Badi et al. (2018) highlighted that many big data governance frameworks tend to minimise the effort needed to adapt and integrate legacy systems with the metadata and access control that current governance standards, such as the FAIR principles, assume.

### 6.2 Regulatory and Ethical Considerations

Given the possibility of automated decisions to impact critical services, Floridi, Cowls, King and Taddeo (2020) identified seven criteria for the design of AI systems for social good: falsifiability, safety, respect for human autonomy. Many regulators around the world are asking for documented reasons for decision making in relation to service continuity and billing, which a good lineage and explainability architecture, as described in Section 5, can help to meet, but the literature has not yet seen a consistent regulatory approach internationally that can be benchmarked against.

### 6.3 Resource Constraints and Scalability

For battery-operated water sensors deployed in remote distribution networks, internet-scale sensing architectures will also need to consider the computational and energy constraints of the edge devices as well as the desire for richer provenance metadata, noted Georgakopoulos and Jayaraman (2016). Similarly, Persson et al. (2022) warned that the computational load associated with consensus protocols within a distributed provenance ledger could be prohibitive for edge devices, which cannot be mediated by architecture, such as by aggregating provenance at gateway nodes before submitting to the ledger.

## 7. Discussion and Future Directions

The synthesis described in this paper suggests that there is a consensus and agreement in the data governance, the trustworthy AI, and the utility IoT literature on a single fact: Trust in AI-based utility decisions will not be achieved by providing the user with explanations of the model, but can only be achieved by ensuring the existence of a documented, auditable chain of custody that can be traced from sensor to decision. The literature review identified a hybrid governance model in which policy coordination occurs at the head while provenance verification is decentralized for high-stakes data as the most balanced one, but a validated generalizable cost benefit analysis of this balance across utility sectors and jurisdictions was not found in the reviewed material. Conceptual synthesis, such as that presented here, would not suffice for validation of the illustrative comparisons presented in Tables 1-3 and Figures 3 and 5, but would require future work based on empirical deployment data. Priority areas include research on longitudinal studies of provenance ledger performance under real-time grid conditions, on creation of standardized metrics for evaluating the fidelity of explanations for utility sensor data, and on comparative regulatory analysis to clarify the extent of provenance documentation that utility regulators in other jurisdictions deem adequate for decisions made using AI assistance for public infrastructure.

## 8. Conclusion

Data governance, FAIR data stewardship, trustworthy and explainable AI, and utility-sector IoT and blockchain literatures have been synthesized into a unified conceptual argument: end-to-end data lineage is a structural (not accidental) prerequisite for trustworthy AI in utility data ecosystems. Incorporating governance, provenance, and explainability as an equal architectural layer alongside AI modelling allows utility companies to begin to build AI systems that produce outputs that are not only correct, but also traceable and accountable to their operators, regulators and communities. The layered structure and illustrative comparisons provided here are offered as a conceptual framework for future empirical work, but are not a validated deployment blueprint.

## References


[1] Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. International Journal of Information

Management, 49, 424–438. https://doi.org/10.1016/j.ijinfomgt.2019.03.014

[2] Al-Badi, A., Tarhini, A., & Khan, A. I. (2018). Exploring big data governance frameworks. Procedia Computer Science, 141, 271–277. https://doi.org/10.1016/j.procs.2018.10.181

[3] Andrić, I., Al-Ghamdi, S. G., & Koç, M. (2022). IoT approach towards smart water usage. Journal of Cleaner Production, 367, Article 133065. https://doi.org/10.1016/j.jclepro.2022.133065

[4] Ahmed, S. A., & Mohammed, N. (2022). A solution for water management and leakage detection problems using IoTs based approach. Internet of Things, 18, Article 100504. https://doi.org/10.1016/j.iot.2022.100504

[5] Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[6] Chamola, V., Hassija, V., Sulthana, A. R., Ghosh, D., Dhingra, D., & Sikdar, B. (2023). A review of trustworthy and explainable artificial intelligence (XAI). IEEE Access, 11, 78994–79015. https://doi.org/10.1109/ACCESS.2023.3294569

[7] Di Vaio, A., Trujillo, L., D'Amore, G., & Palladino, R. (2021). Water governance models for meeting sustainable development goals: A structured literature review. Utilities Policy, 72, Article 101255. https://doi.org/10.1016/j.jup.2021.101255

[8] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

[9] Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to design AI for social good: Seven essential factors. Science and Engineering Ethics, 26(3), 1771–1796. https://doi.org/10.1007/s11948-020-00213-5

[10] Georgakopoulos, D., & Jayaraman, P. P. (2016). Internet of things: From internet scale sensing to smart services. Computing, 98(10), 1041–1058. https://doi.org/10.1007/s00607-016-0510-0

[11] Hasan, M. K., Alkhalifah, A., Islam, S., Babiker, N. B. M., Habib, A. K. M. A., Aman, A. H. M., & Hossain, M. A. (2022). Blockchain technology on smart grid, energy trading, and big data: Security issues, challenges, and recommendations. Wireless Communications and Mobile Computing, 2022, Article 9065768. https://doi.org/10.1155/2022/9065768

[12] Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy artificial intelligence. Government Information Quarterly, 37(3), Article 101493. https://doi.org/10.1016/j.giq.2020.101493

[13] Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy artificial intelligence: A review. ACM Computing Surveys, 55(2), Article 39. https://doi.org/10.1145/3491209

[14] Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., & Zhou, B. (2023). Trustworthy AI: From principles to practices. ACM Computing Surveys, 55(9), Article 177. https://doi.org/10.1145/3555803

[15] Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2021). Explainable AI: A review of machine learning interpretability methods. Entropy, 23(1), Article 18. https://doi.org/10.3390/e23010018

[16] Persson, S., Gorton, I., & Ali, S. (2022). Impact of blockchain technology on smart grids. Energies, 15(19), Article 7189. https://doi.org/10.3390/en15197189

[17] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

[18] Ryś, M., Pincheira, M., Vecchio, M., Giaffreda, R., & Kanhere, S. S. (2021). Cost-effective IoT devices as trustworthy data sources for a

blockchain-based water management system in precision agriculture. Computers and Electronics in Agriculture, 180, Article 105889. https://doi.org/10.1016/j.compag.2020.105889

[19] Simmhan, Y. L., Plale, B., & Gannon, D. (2005). A survey of data provenance in e-science. ACM SIGMOD Record, 34(3), 31–36. https://doi.org/10.1145/1084805.1084812

[20] Strobel, M., & Shokri, R. (2022). Data privacy and trustworthy machine learning. IEEE Security & Privacy, 20(5), 44–49. https://doi.org/10.1109/MSEC.2022.3178187

[21] Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. Electronic Markets, 31(2), 447–464. https://doi.org/10.1007/s12525-020-00441-4

[22] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. Scientific Data, 3, Article 160018. https://doi.org/10.1038/sdata.2016.18