

Metadata-Guided Diffusion and LLM-Orchestrated Quality Governance for Time Series Imputation

Imane Hocine^{1,*}, Asma Abboura², Abdelaziz Kella^{2,3}, Maria Hanini⁴ and Grégoire Danoy^{1,5}

¹Snt, University of Luxembourg, Esch-sur-Alzette, Luxembourg

²University of Hassiba Benbouali, Chlef, Algeria

³Lasting Dynamics, Las Palmas de Gran Canaria, Spain

⁴University of Sheffield, Sheffield, UK

⁵FSTM/DCS, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Abstract

High-quality time series (TS) data is essential for reliable analytics, forecasting, and knowledge-driven systems. In operational settings, however, TS are frequently degraded by missing values arising from sensor faults, intermittent connectivity, maintenance activities, and extended partial-blackout events. Whilst recent diffusion-based models have improved imputation accuracy, they remain largely signal-centric, make limited use of semantic and operational metadata, and provide little support for data quality considerations.

This vision paper introduces a metadata-governed framework for time series imputation that treats data quality as a first-class system concern. The framework combines diffusion-based generative imputation with a metadata knowledge graph (KG), and large language model (LLM) based agents that coordinate imputation workflows and produce auditable quality narratives. Raw time series are retained in specialised storage systems, whilst metadata, relationships, and quality indicators are maintained explicitly in the knowledge graph. This design supports the dynamic extraction of compact, quality-filtered subgraphs that determine which contextual signals may be used to condition the imputation process. We argue that metadata-aware subgraph conditioning, KG-constrained generative imputation, and LLM-mediated explanation would constitute a practical quality layer for time series imputation. Rather than treating imputation as an isolated preprocessing step, the framework treats it as a governance-aware component within KG-LLM pipelines. The paper outlines core design principles and open challenges for building trustworthy and adaptive data systems.

Keywords

Time series imputation, Knowledge graph, Metadata, Diffusion models, LLM orchestration, Data quality.

1. Introduction

Imagine a city relying on a network of traffic sensors to manage congestion in real time. When multiple sensors fail due to maintenance or network disruptions, the system must reconstruct missing measurements to make safe and efficient decisions. Similarly, satellite time series TS used for vegetation monitoring or climate analysis are frequently incomplete due to cloud cover, sensor outages, or acquisition constraints. In both cases, missing time series values threaten operational reliability, safety, and trust in downstream analytics and decision-making pipelines.

Classical imputation methods use interpolation and state-space models [1]. Neural network approaches apply recurrent architectures and transformers to learn temporal dependencies [2, 3]. Graph-based methods exploit spatial and functional relationships between sensors, propagating information across correlated entities [4, 5, 6]. More recently, diffusion models have been applied to TS imputation [7, 6, 8, 9, 4, 10, 11], generating probabilistic reconstructions through iterative denoising processes. These approaches face a fundamental limitation in controllable generation [11, 12]. Diffusion models learn from observed correlations in training data. Graph-based methods encode relationships through

Published in the Proceedings of the Workshops of the EDBT/ICDT 2026 Joint Conference (March 24-27, 2026), Tampere, Finland

*Corresponding author.

✉ imane.hocine@uni.lu (I. Hocine); a.abboura@univ-chlef.dz (A. Abboura); abdelaziz.kella@lastingdynamics.com (A. Kella); m.hanini@sheffield.ac.uk (M. Hanini); gregoire.danoy@uni.lu (G. Danoy)

🆔 0000-0002-9992-577X (I. Hocine); 0009-0002-1740-088X (A. Abboura); 0000-0003-2872-1996 (A. Kella); 0000-0001-7703-9673 (M. Hanini); 0000-0001-9419-4210 (G. Danoy)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

fixed topologies derived from distance metrics or learned correlations. Neither can systematically access operational constraints that determine which conditioning sources are semantically valid.

Moreover, they operate on raw signal values and remain *signal-centric*. Relevant dependencies are either assumed observable in the data or implicitly learnable from correlations. In practice, many critical relationships are *implicit and external to the signal* [13, 12]. For instance, sensor networks contain structured metadata that is not present in raw measurements, including device specifications, operational logs, and quality assessments. This metadata determines which sensors provide valid conditioning context. Optical and radar satellites measure different physical properties. Using radar to impute optical observations may produce smooth interpolations that violate spectral consistency. A recently recalibrated sensor may have shifted its measurement baseline. Conditioning on its historical values introduces bias. As such, when context is absent or implicit, imputation models may produce statistically plausible but operationally invalid reconstructions. Specifically, diffusion models may condition on unreliable or invalid sources.

These failures may well occur in reality. In traffic management, imputing flow measurements from a highway sensor using an unrelated arterial sensor can overestimate throughput capacity, triggering unsafe signal timing decisions. In satellite-based crop monitoring, conditioning optical vegetation indices on radar input produces incoherent values that misclassify crop health, propagating errors into yield forecasts. When such imputed values are ingested into knowledge graphs for downstream reasoning, the damage compounds. Invalid upstream reconstructions become trusted facts in the graph, silently degrading every query and inference built on them. These failure modes cannot be resolved by improving model capacity alone. They require access to external metadata to determine which sources are semantically valid for conditioning.

To address this gap, we envision a metadata- and governance-aware framework for TS imputation. The system combines a metadata knowledge graph (KG), subgraph-conditioned diffusion, and large language model (LLM)-based orchestration. This vision directly addresses key challenges in hybrid KG-LLM ecosystems by improving the fidelity of data streams feeding KGs, reducing downstream reasoning errors caused by poor upstream data quality, and supporting human-in-the-loop governance [10].

Hypothesis and Vision

Our central hypothesis is that externalising implicit relationships in a KG and conditioning diffusion models on quality-filtered subgraphs improves imputation under structured missingness, heterogeneous sensors, and evolving operational conditions. Learning-based models capture temporal and cross-variable patterns from observed data. They cannot infer operational semantics, provenance, or quality constraints that are not present in raw signals [4]. Data-driven imputation degrades when correlations are absent, when sensors are heterogeneous, or when operational context changes. We therefore propose treating metadata as a first-class artefact, external to both data and models, and representing it in a KG that is explicit, queryable, and evolvable [14, 15]. Within this paradigm: (i) Diffusion models perform statistical reconstruction, generating candidate imputations conditioned on observed data and selected context [7, 9, 16]. (ii) Knowledge graphs externalise implicit relationships, enforce semantic and quality constraints, and encode provenance signals [14, 15]. (iii) LLM-based agents orchestrate the workflow and generate human-readable explanations for audit and governance [17, 18]. Separating statistical reconstruction from semantic and operational reasoning allows the system to adapt to evolving conditions without retraining core models. The framework does not replace state-of-the-art diffusion methods; it provides a conditioning and governance layer that makes implicit metadata explicit during generation. This reframes TS imputation as a governed, explainable component of data ecosystems rather than an isolated preprocessing step. This paper describes the system architecture and identifies open technical challenges. Implementation is in progress. The design targets scenarios where structured metadata exists and quality governance is required.

2. Related Work

TS imputation has been widely studied. Classical imputation methods include interpolation, Kalman filtering, and state-space models [1]. These methods provide interpretability and computational efficiency. Yet, they fail on high-dimensional data with non-linear dependencies and complex missing patterns. Neural architectures using recurrent networks and Transformers [16, 19] capture non-linear dynamics through learnt representations. These models treat each series independently or learn correlations from data. They do not exploit explicit metadata about sensor properties, operational constraints, or quality indicators. Diffusion models represent a recent advance in generative imputation [7, 6, 8, 9, 16]. These models generate probabilistic reconstructions by modelling temporal and cross-variable dependencies through iterative denoising. Recent work extends diffusion to forecasting and general TS generation [16, 10, 11]. Conditioning remains limited to observed measurements and patterns learnt from training data. External metadata specifying sensor modality, operational constraints, or quality indicators is not incorporated.

Graph-based methods propagate information across correlated sensors using GNNs, hypergraphs, or attention mechanisms [4, 5, 6], capturing spatial, functional, and topological dependencies. Yet, they often rely on fixed graphs derived from heuristics and rarely exploit rich metadata or quality indicators, limiting interpretability and adaptability. General-purpose TS architectures such as TimeMixer++ mainly emphasise representation capacity across TS tasks rather than metadata conditioning [10].

A parallel line of work uses knowledge graphs to encode structured metadata, provenance, and relational constraints [14, 15]. Recent research explores LLMs as orchestrators for KG workflows, including construction from unstructured text, subgraph selection, and explanation generation [17, 18]. Prior work focuses on KG construction, completion, retrieval-augmented generation, and post-hoc reasoning, with few approaches guiding TS imputation or producing auditable quality narratives.

Existing methods address generative modelling, relational reasoning, or workflow orchestration independently. The proposed paradigm combines diffusion-based imputation, metadata-driven KG, and LLM orchestration in a single framework for governed TS reconstruction.

3. Framework Architecture

The framework comprises three-layer hybrid system operating on different data representations. The metadata knowledge graph maintains sensor specifications, quality indicators, and relationships. Diffusion models generate imputations conditioned on observed measurements and graph-derived context. LLM agents query the graph, configure imputation parameters, and produce audit trails. Figure 1 shows the data flow. When imputation is requested for sensor s at time t , the system would extract a quality-filtered subgraph G_{sub} containing candidate conditioning sensors. A graph neural network computes embeddings injected into the diffusion model’s denoising process. Generated samples are validated against metadata graph constraints. An LLM agent documents which sensors were retained, and why.

3.1. Metadata Knowledge Graph

The metadata knowledge graph would store sensor properties, quality indicators, and operational constraints. Raw TS remain in specialised time-series databases. This separation addresses scalability using a graph database (such as Neo4j) that handles thousands of sensor entities and their relationships, whilst numerical storage manages high-frequency measurements. Sensors are represented as nodes with attributes such as measurement type, physical units, operating range, calibration dates, and reliability indicators. Contextual entities capture geographic regions, infrastructure components, and asset hierarchies. Edges encode relationships such as spatial proximity, network topology, functional similarity, and maintenance dependencies, while partitions summarise similar behaviors or dynamics.

From a technical perspective, the KG uses a lightweight property-graph schema rather than a full ontology. Core node types include *Sensor*, *MeasurementType*, *Location*, *Asset*, and *QualityEvent*.

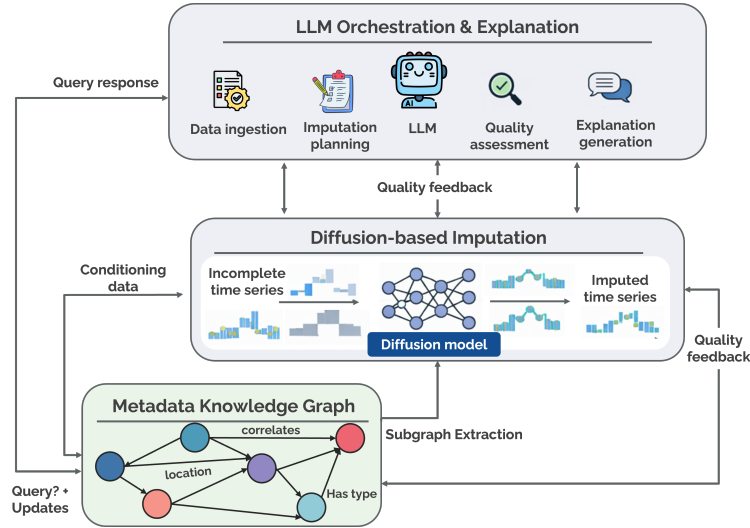


Figure 1: High-level architecture for metadata-guided imputation.

Sensor nodes store attributes such as modality, physical units, operating range, calibration date, and reliability indicators. Edges encode MEASURES, LOCATED_IN, ADJACENT_TO, FUNCTIONALLY_RELATED, and AFFECTED_BY, enabling traversal over specified relationships. Quality constraints are enforced using query-level filtering, allowing the schema to evolve with operational needs.

In doing so, the graph externalises relationships that learning-based models cannot infer from raw signals. A GNN trained on TS may learn that sensors A and B are correlated. It cannot determine whether A measures temperature in Celsius whilst B measures humidity as a percentage, rendering them semantically incompatible for direct conditioning. The graph encodes modality explicitly. Crucially, the system maintains an explicit, queryable, and evolvable representation of metadata without handling high-volume numerical data, enforcing a clean separation between statistical reconstruction and semantic reasoning.

3.2. Quality-Aware Subgraph Extraction

For each imputation task targeting sensor s over time window $[t_1, t_2]$, we extract a task-specific subgraph $G_{\text{sub}} = (V_{\text{sub}}, E_{\text{sub}})$ that defines admissible conditioning context. Candidate neighbors are first identified through graph traversal over spatial, topological, and similarity relations, and then filtered based on quality indicators (e.g., missingness rates, anomaly flags) and semantic compatibility (e.g., sensor modality or measurement units).

This approach replaces heuristic neighbor selection (e.g., k -nearest by distance) with metadata-governed decisions. Each edge in G_{sub} has an explicit justification, and operators can inspect, override, or revise these rules through graph queries without retraining the diffusion model. Thresholds and traversal depth d are configurable at query time, enabling transparent and auditable context selection. Using metadata for subgraph extraction, the framework ensures that conditioning context is semantically and operationally valid. This enables flexibility and explainability in the imputation process.

3.3. KG-Conditioned Diffusion Imputation

In standard diffusion-based TS imputation, X_t denotes the noisy latent at diffusion step t and X^{obs} the observed measurements. The model learns the reverse process:

$$p_{\theta}(X_{t-1} | X_t, X^{\text{obs}}).$$

Our framework extends this formulation by conditioning the generation process on the metadata subgraph:

$$p_{\theta}(X_{t-1} | X_t, X^{\text{obs}}, G_{\text{sub}}).$$

Embeddings of G_{sub} are calculated using a graph neural network that encodes relational structure, quality indicators, and semantic attributes. These embeddings are injected into the diffusion model’s denoising network. Implementation options include concatenation with intermediate features, cross-attention mechanisms, or message-passing layers. The choice depends on the specific diffusion architecture and computational constraints.

Constraint enforcement combines soft penalties during training with hard clipping at sampling time.

Let $\mathcal{L}_{\text{diffusion}}$ denote the standard diffusion objective and $\mathcal{L}_{\text{constraint}}$ a penalty for violations of sensor-specific bounds $[v_i^{\min}, v_i^{\max}]$ encoded in the metadata graph. The resulting training objective is

$$\mathcal{L} = \mathcal{L}_{\text{diffusion}} + \lambda \mathcal{L}_{\text{constraint}}.$$

This design allows the knowledge graph and diffusion model to evolve independently. Adding new sensor types requires updating graph schema, not retraining the denoising network. Changing quality thresholds modifies subgraph extraction rules without touching model parameters.

3.4. LLM-Orchestrated Workflow

LLMs orchestrate the imputation pipeline through specialised agents that query the knowledge graph and coordinate workflow steps.

The **KG-construction agent** updates metadata from maintenance logs, calibration records, and operational documentation. It parses unstructured text to extract structured facts and writes them to the graph whilst tracking provenance.

The **imputation-planning agent** receives imputation requests and queries the graph for candidate conditioning sensors. It adapts subgraph extraction based on missingness patterns and metadata quality.

The **quality-assessment agent** evaluates generated reconstructions against metadata-derived constraints. It checks whether values fall within valid operating ranges, assesses temporal consistency with adjacent observations, and compares against ground truth when available. Quality metrics are written back to the graph as provenance records.

The **explanation agent** generates human-readable narratives documenting imputation decisions. For each reconstruction, it describes which sources, constraints, and reliability factors influenced each imputation. Explanations reference only entities and attributes retrieved from the graph to prevent hallucinated claims.

Agent interactions with graph databases use generated Cypher queries. To ground LLM outputs in factual metadata, agents receive graph query results in their prompts and are constrained to reference only retrieved entities. If an agent generates ungrounded claims, the system detects this through entity linking and retries with stricter constraints. Agent interactions follow a retrieve-then-generate pattern. For each task, the relevant agent first issues a structured query to retrieve candidate entities and their attributes from the KG. The query results are then injected into the agent’s prompt as the sole factual context. To prevent hallucination, agent outputs undergo entity-linking verification and every sensor, attribute, or event referenced in the response is checked against the query result set. If ungrounded references are detected, the agent is re-prompted with the specific violation. This retrieve-constrain-verify loop applies uniformly across all agents. The orchestration layer provides transparency for human oversight and structured provenance for downstream reasoning. The modular agent design allows individual components to be updated or replaced as LLM capabilities evolve.

3.5. Operational interaction between components

Components operate at distinct abstraction levels. The KG provides a task-specific quality-filtered subgraph. A lightweight GNN encodes this subgraph into embeddings representing semantic compatibility, quality indicators, and relational structure. LLM agents orchestrate the workflow and generate explanations grounded in KG facts but do not participate in numerical generation.

3.6. Illustrative Example

Consider a satellite-derived TS for a fixed geographic tile observed across five acquisition times. Four sources provide measurements: a primary optical sensor O_1 , an auxiliary optical sensor O_2 , a radar sensor R , and an adjacent tile A observed by O_1 . At time t_3 , the measurement from O_1 is invalid due to cloud contamination, creating structured corruption as shown in Table 1.

Table 1

Observed TS for a target tile over five acquisitions.

| Source | t_1 | t_2 | t_3 | t_4 | t_5 |
|-----------------------------|-------|-------|-------|-------|-------|
| O_1 (target tile) | 0.62 | 0.63 | ? | 0.66 | 0.67 |
| O_2 (target tile) | 0.60 | 0.61 | 0.62 | 0.64 | 0.65 |
| R (target tile) | 0.41 | 0.40 | 0.39 | 0.38 | 0.38 |
| A (adjacent tile, O_1) | 0.61 | 0.62 | 0.63 | 0.65 | 0.66 |

Let X denote the multivariate series over $\{O_1, O_2, R, A\}$ and X^{obs} the set of all observed entries excluding $O_1(t_3)$. A task-specific metadata subgraph G_{sub} is extracted from the knowledge graph in Figure 2 by retaining only semantically compatible and quality-valid sensors. In this example, O_2 and A constitute admissible conditioning signals, while O_1 is excluded due to cloud invalidation and R due to modality incompatibility with optical vegetation indices. Conditioning the diffusion model on $(X^{\text{obs}}, G_{\text{sub}})$ yields a posterior distribution over the missing value. The imputed value $\hat{y}^{\text{KG}}(t_3)$ is interpreted as a point summary (e.g., conditional mean, MAP) derived under KG-governed context selection. Raw observations alone do not explain why $O_1(t_3)$ is missing nor which alternative sources should be trusted. The metadata KG provides essential context: (i) $O_1(t_3)$ is flagged as invalid due to cloud contamination, (ii) R is a radar sensor and therefore not compatible with optical reflectance, and (iii) the adjacent tile A exhibits historically strong spatial affinity with the target tile. Figure 2 illustrates how these semantic constraints are encoded and operationalised for context selection. A signal-centric approach would average all available sources, including semantically incompatible ones such as R , leading to biased or incoherent estimates. In contrast, the framework excludes invalid sources and conditions only on admissible signals. This estimate respects temporal continuity and is accompanied by explicit, metadata-grounded justification. This example illustrates how *metadata knowledge graphs encode provenance, quality, and semantic constraints that are absent from raw observations*. For illustration, a simplified weighted combination gives: $\hat{y}^{\text{KG}}(t_3) = 0.7 \cdot O_2(t_3) + 0.3 \cdot A(t_3) = 0.623$.

4. Discussion

This framework demonstrates how KGs can govern generative imputation in operational data systems. Three architectural decisions distinguish this approach from signal-only methods. Diffusion models receive explicit semantic constraints rather than inferring them from correlations. This is expected to improve robustness when training correlations are absent, misleading, or outdated. The KG externalises what models cannot learn from measurements alone [13].

Statistical reconstruction, semantic validation, and workflow coordination operate independently. New sensors, updated calibration schedules, or revised quality policies modify graph content without retraining diffusion models. Human operators inspect and override subgraph extraction rules through graph queries. This modularity supports adaptation in evolving deployments. Additionally, LLM agents document which sensors were included or excluded, which constraints were enforced, and which quality measures determined subgraph membership. Explanations reference graph entities, enabling systems to trace provenance.

The framework addresses a gap in hybrid KG-LLM ecosystems by allowing metadata governance at the imputation stage. This reduces the risk of poor TS fidelity propagating errors into reasoning systems. The design pattern generalises beyond TS. Domains where raw measurements require semantic

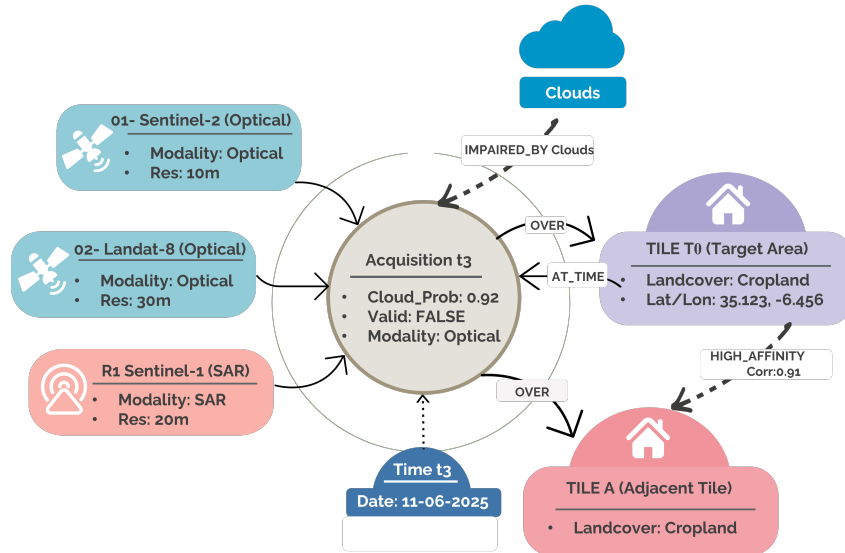


Figure 2: Metadata KG for satellite time series example.

validation (medical sensors, financial data, environmental monitoring) could apply metadata-conditioned generation. The specific technologies (Neo4j, diffusion models, LLM agents) are instantiation choices. The principle is separating what models learn from data, from what systems enforce through metadata.

The proposed approach is designed to minimise training and fine-tuning overhead. Diffusion models are trained or reused independently of the knowledge graph and LLM components. The graph neural network operates on compact subgraphs rather than the full metadata graph to lower inference costs. Since metadata evolves more slowly than raw time series, graph embeddings can be cached or incrementally updated. LLMs are only used at inference time without fine tuning. The approach is feasible in environments where retraining large models is costly or impractical. It is most valuable in domains where structured metadata exists alongside time series measurements, sensors are heterogeneous in their characteristics, and governance is required. For homogeneous univariate time series without operational metadata, the metadata layer adds limited value and standard imputation methods could suffice. The framework’s benefit scales with metadata richness. The more operational context is available in the KG, the greater the improvement over traditional approaches. The specific technology choices are instantiation decisions, and the core principle of separating statistical reconstruction from metadata-governed context selection is architecture-agnostic.

5. Conclusion & Open Challenges

This vision paper opens research directions for diffusion-based time series imputation by explicitly integrating metadata as conditioning context. Existing diffusion models learn temporal dependencies from observed correlations but cannot infer operational semantics, quality constraints, or provenance from raw measurements. We propose externalising this knowledge in queryable graphs and conditioning generation on quality-filtered subgraphs. This reframes imputation as a process constrained by semantic validity and not temporal plausibility alone.

Implementation is underway integrating KGs with diffusion-based imputation models. Evaluation will assess whether metadata and KG-conditioned imputation improves robustness on real-world datasets, including both regular and irregular time series. Envisaged baselines include Yun et al. [8] and Diffusion-TS [16]. Considered metrics for evaluation will include RMSE/MAE, metadata-aware criteria covering constraint violations, structured missingness robustness, and fidelity of LLM-generated explanations.

Notable challenges remain open. Subgraph extraction and graph embedding add computational

overhead. This calls for latency analysis for high-frequency streams, and caching/incremental updates need exploration. The framework assumes reasonably accurate metadata, yet operational systems contain outdated calibration records, incorrect specifications, and stale quality indicators. Degradation under these errors remains an open problem. Potential approaches include uncertainty-aware graph queries, metadata validation pipelines, and human-in-the-loop verification for high-stakes imputations. Similarly, explanation agents must avoid hallucinating facts not present in the knowledge graph. Structured generation and constrained decoding may improve reliability, but formal verification that references only retrieved entities is still unsolved.

Existing imputation benchmarks measure reconstruction error on held-out test data, but do not assess semantic validity, explainability quality, or governance effectiveness. Evaluation frameworks capturing these dimensions are needed.

Additionally, sensor networks lack common vocabularies for quality indicators, calibration procedures, and operational constraints. Domain-specific ontologies exist, but integrating heterogeneous deployments requires schema alignment and entity resolution. KG construction from unstructured logs remains a challenging extraction problem.

These challenges define a research agenda for trustworthy imputation in operational settings. Translating this vision into production systems requires addressing computational efficiency, metadata quality assurance, and evaluation protocols that measure semantic validity and auditability beyond reconstruction error.

6. Acknowledgement

This work was supported by the Luxembourg National Research Fund (FNR) & the National Centre for Research and Development (NCBR) under the SERENITY Project (ref. C22/IS/17395419; POLLUX-XI/15/Serenity/2023)

Declaration on Generative AI

The authors declare that generative AI tools (ChatGPT) were used for language refinement, including Grammar and spelling check, improve writing style, and peer review simulation. All content was subsequently reviewed and edited by the authors, who take full responsibility for the accuracy, originality, and claims presented in this work.

References

- [1] J. Hagiwara, et al., Time series analysis for the state-space model with R/Stan, Springer, 2021.
- [2] W. Cao, L. Wang, J. Li, et al., Brits: Bidirectional recurrent imputation for time series, NeurIPS (2018).
- [3] Z. Che, S. Purushotham, et al., Gru-d: Handling missing data with recurrent neural networks, ICML (2018).
- [4] B. Li, Y. Luo, Z. Liu, J. Zheng, J. Lv, Q. Ma, Hyperimts: Hypergraph neural network for irregular multivariate time series forecasting, in: International Conference on Machine Learning (ICML), 2025. ArXiv:2505.17431.
- [5] X. Wang, H. Si, F. Zhang, X. Zhou, D. Sun, W. Lyu, Q. Yang, J. Tang, Hgts-former: Hierarchical hypergraph transformer for multivariate time series analysis, arXiv preprint (2025). ArXiv:2508.02411.
- [6] M. Liu, H. Huang, H. Feng, L. Sun, B. Du, Y. Fu, Pristi: A conditional diffusion framework for spatiotemporal imputation, in: 2023 IEEE 39th international conference on data engineering (ICDE), IEEE, 2023, pp. 1927–1939.
- [7] Y. Tashiro, J. Song, Y. Song, S. Ermon, Csd: Conditional score-based diffusion models for probabilistic time series imputation, NeurIPS 34 (2021) 24804–24816.

- [8] T. Yun, H. Jung, J. Son, Imputation as inpainting: Diffusion models for spatiotemporal data imputation (2023).
- [9] M. R. U. Islam, P. Tadepalli, A. Fern, Self-attention-based diffusion model for time-series imputation in partial blackout scenarios, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 39, 2025, pp. 17564–17572.
- [10] S. Wang, J. Li, X. Shi, Z. Ye, B. Mo, W. Lin, S. Ju, Z. Chu, M. Jin, Timemixer++: A general time series pattern machine for universal predictive analysis, arXiv preprint (2024). ArXiv:2410.16032.
- [11] R. Jiang, G.-C. Zheng, T. Li, T.-R. Yang, J.-D. Wang, X. Li, A survey of multimodal controllable diffusion models, Journal of Computer Science and Technology 39 (2024) 509–541.
- [12] J. Hoelscher-Obermaier, E. Stevinson, V. Stauber, I. Zhelev, V. Botev, R. Wu, J. Minton, Leveraging knowledge graphs to update scientific word embeddings using latent semantic imputation, in: Proceedings of the first Workshop on Information Extraction from Scientific Publications, 2022, pp. 43–53.
- [13] Y. Liu, G. Shen, N. Liu, X. Han, Z. Xu, J. Zhou, X. Kong, Traffic data imputation via knowledge graph-enhanced generative adversarial network, PeerJ Computer Science 10 (2024) e2408.
- [14] A. Hogan, E. Blomqvist, M. Cochez, C. d’Amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, et al., Knowledge graphs, ACM Computing Surveys (Csur) 54 (2021) 1–37.
- [15] S. Ji, S. Pan, E. Cambria, P. Marttinen, P. S. Yu, A survey on knowledge graphs: Representation, acquisition, and applications, IEEE Transactions on Neural Networks and Learning Systems 33 (2022) 494–514.
- [16] X. Yuan, Y. Qiao, Diffusion-ts: Interpretable diffusion for general time series generation, in: International Conference on Learning Representations (ICLR), 2024.
- [17] Y. Zhu, X. Wang, J. Chen, S. Qiao, Y. Ou, Y. Yao, S. Deng, H. Chen, N. Zhang, Llms for knowledge graph construction and reasoning: Recent capabilities and future opportunities, arXiv preprint arXiv:2305.13168 (2024). V4, Dec 2024.
- [18] D. Edge, H. Trinh, N. Cheng, J. Bradley, A. Chao, A. Mody, S. Truitt, D. Metropolitanansky, R. O. Ness, J. Larson, From local to global: A graph rag approach to query-focused summarization, arXiv preprint arXiv:2404.16130 (2024).
- [19] H. Wu, J. Xu, J. Wang, M. Long, Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting, NeurIPS 34 (2021) 22419–22430.