

LSDTs: LLM-Augmented Semantic Digital Twins for Adaptive Knowledge-Intensive Infrastructure Planning

Naiyi Li^{1*}, Zihui Ma^{1,2*}, Runlong Yu^{3†}, Lingyao Li^{4†}

¹Department of Civil & Environmental Engineering, University of Maryland, College Park

²Center for Urban Science and Progress, New York University

³Department of Computer Science, University of Alabama

⁴School of Information, University of South Florida

nli5211@umd.edu, zihuima@nyu.edu, ryu5@ua.edu, lingyaol@usf.edu

Abstract

Digital Twins (DTs) offer powerful tools for managing complex infrastructure systems, but their effectiveness is often limited by challenges in integrating unstructured knowledge. Recent advances in Large Language Models (LLMs) bring new potential to address this gap, with strong abilities in extracting and organizing diverse textual information. We therefore propose LSDTs (LLM-Augmented Semantic Digital Twins), a framework that helps LLMs extract planning knowledge from unstructured documents like environmental regulations and technical guidelines, and organize it into a formal ontology. This ontology forms a semantic layer that powers a digital twin—a virtual model of the physical system—allowing it to simulate realistic, regulation-aware planning scenarios. We evaluate LSDTs through a case study of offshore wind farm planning in Maryland, including its application during Hurricane Sandy. Results demonstrate that LSDTs support interpretable, regulation-aware layout optimization, enable high-fidelity simulation, and enhance adaptability in infrastructure planning. This work shows the potential of combining generative AI with digital twins to support complex, knowledge-driven planning tasks.

1 Introduction

Large-scale infrastructure projects (e.g., energy systems, transportation networks) often face significant challenges in project planning (Flyvbjerg 2007, 2009). A key factor underlying management failures is the lack of effective knowledge integration (Huang, Ato, and Li 2022; D’Amico et al. 2020), as critical project information often remains fragmented across unstructured documents—such as contracts specifying environmental impact assessments, regulatory codes mandating safety standards. Conventional project management frameworks (e.g., PMBOK (Institute 2021), PRINCE2 (Bentley 2012)) are not designed to process or reason over such content. Although Digital Twins (DTs) have transformed infrastructure planning through real-time monitoring and predictive modeling (Broo, Bravo-Haro, and Schooling 2022; Broo and Schooling 2023; Deng et al. 2025), they often depend on structured inputs from Internet of Things (IoT) models (Minerva, Lee, and Crespi 2020;

Kaur, Mishra, and Maheshwari 2019; Zhao et al. 2021). As DT technology advances, there is an urgent need to enable DTs to incorporate more advanced language processing capabilities for information retrieval and integration.

To address these challenges, many DT studies have leveraged semantic web technologies (Jiang et al. 2023; Jagathesaperumal et al. 2023) for structured information extraction. In addition, researchers have applied natural language processing (NLP) tools such as rule-based (Su et al. 2022) or BERT-based models (Jiang et al. 2025) for information extraction. However, challenges still remain. Current methods often struggle to transform unstructured, large-scale text into structured, machine-readable knowledge. They also have limited ability to integrate textual content with spatial representations, reason across documents, or adapt dynamically as real-time planning information become available in DTs.

Recent advances in Large Language Models (LLMs) offer a promising path forward (Wang et al. 2024; Li et al. 2023; Chen et al. 2025). Models like GPT-4 and Gemini demonstrate strong capabilities in understanding documents, synthesizing information across sources, and reasoning about complex relationships (Fan et al. 2024; Dagdelen et al. 2024). LLMs can extract key entities (e.g., stakeholders, deadlines, and constraints), identify dependencies between requirements and regulations, and reason semantically across documents (Sun et al. 2025; Zhou and Ma 2025), with minimal need for manual rule creation. These strengths enable a more scalable and context-aware approach to managing complex project information.

Therefore, we propose **LLM-Augmented Semantic Digital Twins (LSDTs)**—an integrated architecture that combines the language understanding of LLMs with the simulation and visualization strengths of DTs. Our work makes two main contributions. First, we develop an **LLM-driven pipeline for knowledge structuring**, where LLMs extract regulatory constraints and planning knowledge from unstructured documents into a structured semantic graph that unifies diverse regulatory, spatial, and technical requirements. Second, we design a **system workflow for DTs** that links LLM-derived knowledge with real-time simulation and scenario analysis. Our developed system provides important capabilities for automatically interpreting planning documents, structuring information, and supporting adaptive, regulation-aware decision-making in infrastructure projects.

*Co-lead authors.

†Co-corresponding authors.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

2 Related Work

2.1 Infrastructure Planning and Digital Twins

Infrastructure planning for large-scale projects relies on structured semantic models, such as IFC-enabled building information modeling (BIM), which captures detailed building-level geometry and semantics, and CityGML, which extends these representations to the urban context (Lam et al. 2024). Despite these advances, incorporating critical but unstructured information, such as environmental regulations, stakeholder requirements, and policy guidelines, remains challenging, especially during early planning stages. Existing ontology-based systems can extract specific regulatory clauses from textual sources, but they often require extensive manual rule definitions and lack adaptability (Peng and Liu 2023).

In parallel, digital twin (DT) technologies have matured around operational applications, such as sensor-driven monitoring and predictive maintenance Abd Wahab et al. (2024), while their use in the early-stage planning are largely experimental and conceptual (Mousavi et al. 2024). Recent advances, such as asset administration shells (AAS) integrated with information containers for linked document delivery (ICDD), demonstrate how modular and interoperable DTs can support project management (Zhang et al. 2025). Similarly, microservice-based DTs pilots in domains such as smart grids and climate-resilient infrastructure highlight the potential of loosely coupled and adaptable architectures (Mchirgui et al. 2024). However, current DTs rarely integrate unstructured textual data seamlessly or accommodate dynamic, stakeholder-specific constraints in planning contexts. This limitation is particularly pronounced in wind farm planning (Ambarita et al. 2024), where planners must reconcile spatial models, environment assessments, and policy documents.

2.2 LLM Application in Infrastructure Planning

LLMs have demonstrated robust capabilities in entity extraction, relation identification, cross-document reasoning, and multimodal comprehension. These models efficiently transform heterogeneous inputs, including PDFs, HTML pages, and plain text, into structured knowledge without extensive manual engineering. Early built-environment applications exemplify this potential: Hölting et al. (2025) employ LLMs to automatically convert relational construction data into Resource Description Framework (RDF) semantic graphs, significantly streamlining ontology mapping. Madireddy et al. (2025) leverage LLM pipelines to translate building codes into machine-readable rules for BIM compliance checking. Du et al. (2024) introduce Text2BIM, a multiagentframework that generates editable 3D building models from natural language descriptions. Moving beyond single-asset contexts, Gürçan et al. (2025) embeds LLM agents within a Social Digital Twinner platform, enabling natural language querying and steering of agent-based simulations, thus highlighting a novel interaction paradigm for policy-oriented digital twins. A recent survey by Yang et al. (2025) organizes these emerging efforts into a descriptive–predictive–prescriptive taxonomy, identifying crit-

ical challenges including data grounding, continuous model updating, and provenance tracking.

Despite these promising advances, LLM applications in infrastructure planning remain largely limited to isolated linguistic tasks, such as classification, summarization, and question-answering, without deep integration into simulation engines or decision-support workflows. Current prototypes seldom inject LLM-extracted facts into persistent semantic graphs, lack mechanisms for tracking evolving planning constraints, and rarely link outputs directly to scenario analytics or optimization pipelines.

3 Methodology

This section introduces our conceptual framework, LLM-Augmented Semantic Digital Twins (LSDTs), a knowledge-guided planning architecture that integrates LLM-based knowledge extraction, ontology-driven reasoning, and environmental simulation within a unified semantic substrate (Figure 1). LSDTs prioritize structured knowledge construction and guided decision support. Rather than treating compliance, simulation, and optimization as disconnected stages, the framework operates as a closed-loop workflow in which each component iteratively updates, queries, and enriches a shared semantic graph. At the center of this loop is a domain-specific ontology that shapes the structure of extracted content, constrains simulation behaviors, and mediates planning decisions. This knowledge-centered design ensures semantic alignment and interpretability across all modules, enabling explainable and regulation-aware infrastructure planning without resorting to black-box automation.

3.1 Semantic Foundations

At the core of the LSDT framework is a knowledge structure $G = \{(s, p, o)\}$ composed of RDF triples that instantiate a domain ontology \mathcal{O} . The ontology defines key concepts, relationships, and geospatial constraints relevant to infrastructure planning—for example, class hierarchies (e.g., *Wind-Farm* \sqsubseteq *Infrastructure*), property domains and ranges (e.g., *intersects: Geometry* \times *Geometry*), and built-in geospatial functions wrapped around the Java Topology Suite (Vivid Solutions and LocationTech 2025). These axioms provide the shared vocabulary across modules and support schema validation, logical inference, and spatial reasoning. Given a corpus of unstructured planning documents \mathcal{D} , we define a language model function:

$$L : \mathcal{D} \rightarrow \mathcal{I} = \mathcal{I}_A \cup \mathcal{I}_C \quad (1)$$

that extracts two kinds of semantically typed information:

- **Attribute snippets** \mathcal{I}_A : entity-level properties (e.g., turbine spacing, setback distances);
- **Constraint snippets** \mathcal{I}_C : regulation-like statements describing permitted or prohibited configurations.

These outputs are not free-form: they are guided by ontology-aware prompts and post-processed to ensure consistency with \mathcal{O} . Attribute-level statements are converted to RDF triples using a schema-aligned instantiation function $F_{\text{inst}} : \mathcal{I}_A \rightarrow \mathcal{T}^+$, producing an initial graph G_0 . Constraint

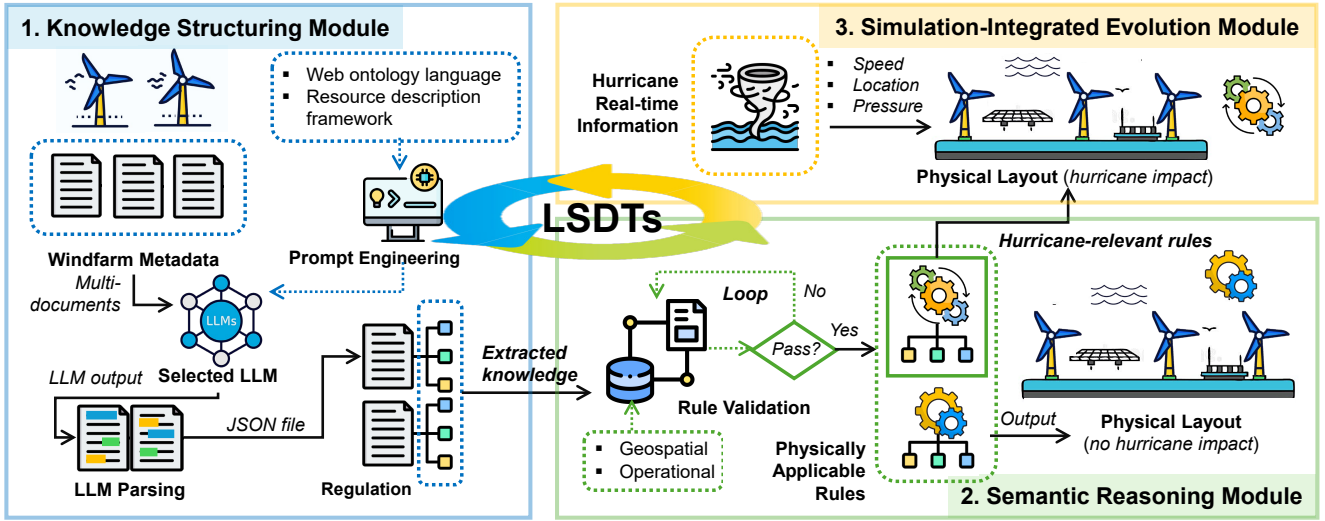


Figure 1: An illustrative framework design of LSDTs.

expressions are formalized as executable logical rules via $F_{\text{rule}} : \mathcal{I}_C \rightarrow \mathcal{R}$, where each rule $r \in \mathcal{R}$ takes the form:

$$\text{Body}(r) \Rightarrow \text{Head}(r) \quad (2)$$

with $\text{Body}(r)$ representing conjunctions of triple patterns and spatial predicates, and $\text{Head}(r)$ asserting derived facts (e.g., :hasConflict true) into the graph.

Given a candidate design ΔG (e.g., a proposed wind farm layout), we update the graph $G = G_0 \cup \Delta G$ and apply a forward-chaining reasoner to evaluate the rule set:

$$G^* = \text{Reason}(G, \mathcal{R}) \quad (3)$$

Violations or inferences are explicitly materialized as new triples, enabling transparent, queryable validation.

To incorporate dynamic environmental inputs, such as wind, weather, or hazards, we define a simulation function Ψ that injects time-varying exogenous data S_t into the graph:

$$G_t = \Psi(G_{t-1}, S_t) \quad (4)$$

This produces a sequence $\{G_t\}_{t=0}^T$ representing the evolving semantic state of the infrastructure system under changing conditions. At each time step, \mathcal{R} is reevaluated on G_t to assess regulatory compliance under evolving conditions.

Finally, an external optimization function \mathcal{O}_{pt} queries the graph G^* for constraint violations or performance metrics, and proposes revised candidate designs δG , closing the loop. This architecture supports iterative refinement, where knowledge extraction, reasoning, simulation, and design co-evolve within a shared semantic substrate.

3.2 Modular Implementation

Building upon the core architecture of LSDTs, we now turn to its system-level implementation, which enables real-time planning and adaptive simulation based on the semantic foundation. While the previous section introduced the formalism underlying the framework, this section details how

its components are integrated into a three-module workflow in DTs. Specifically, we demonstrate how LLMs enable the construction of a structured semantic graph, how LLM-derived planning constraints guide scenario simulations, and how environmental inputs are dynamically injected into the semantic substrate to support time-sensitive and regulation-aware decision-making.

Module 1: Knowledge Structuring. This module uses knowledge-guided prompt templates to extract semantic information from unstructured planning documents. Leveraging LLMs, it produces attribute-level facts \mathcal{I}_A and constraint-level expressions \mathcal{I}_C , each anchored to the pre-defined semantic schema. The outputs are post-processed to normalize formats, resolve units, and ensure consistency with the domain ontology. Attribute facts are inserted into the initial RDF graph G_0 via type-safe instantiation, while constraint expressions are compiled into executable rules that form the rule set \mathcal{R} . This automated pipeline enables reliable knowledge acquisition with minimal human supervision.

Module 2: Semantic Reasoning. Once knowledge is structured, it should be used to assess whether candidate designs meet all applicable constraints continuously. Traditional workflows defer compliance checks to post-design phases, often requiring costly revisions. This module embeds compliance logic directly into the semantic graph by executing forward chain inference on $G = G_0 \cup \delta G$ using a set of rules \mathcal{R} . Potential violations are materialized as triples RDF (for example, asserting :hasConflict true on a non-compliant entity) within the updated graph G^* , allowing real-time validation, transparent explanations, and machine-interpretable feedback. This tight coupling between design and reasoning supports iterative layout refinement and speeds up convergence toward regulation-compliant configurations.

Module 3: Simulation-Integrated Evolution. Infrastructure should also withstand uncertain and time-varying environmental factors, such as extreme weather events. How-

ever, resilience assessment is often disconnected from design workflows. This module introduces a simulation function Ψ that injects exogenous, time-indexed inputs S_t (e.g., storm trajectories, sensor readings) into the graph. Simulation results, such as turbine shutdowns or damage estimates, are expressed using the same ontology \mathcal{O} , allowing downstream rules in \mathcal{R} to be re-applied at each timestep. This integration supports compliance-aware adaptation under dynamic conditions, enabling planners to evaluate both feasibility and resilience in a unified reasoning loop.

3.3 Instantiation: Offshore Wind Farm Planning

To demonstrate the generality of LSDTs, we instantiate the framework for offshore wind farm planning, a representative infrastructure task with complex spatial regulations and environmental dependencies. In this case, the ontology $\mathcal{O}_{\text{INFRASTRUCTURE}}$ includes domain-specific classes such as WindFarm, Turbine, and Cable, with geometry-typed properties defined according to geospatial standards. LLMs including GPT-4o-mini (OpenAI 2024), Claude Haiku 3.5 (Anthropic 2024), Gemini-1.5-Pro (Gemini Team, Google 2024), and DeepSeek (DeepSeek-VL Team 2024; DeepSeek AI 2024) are applied to regulatory documents from various documentations, yielding structured RDF outputs \mathcal{I}_A and constraint expressions \mathcal{I}_C . After schema validation, we obtain an initial graph G_0 and a rule set \mathcal{R} encoding constraints such as minimum turbine spacing and exclusion buffers around protected zones.

Candidate layouts are encoded as RDF triples using Well-Known Text (WKT) literals and added to the base graph as δG . We employ the Jena rule engine, extended with geospatial built-ins, to apply the previously defined rule set \mathcal{R} . This module operationalizes the reasoning process described earlier, enabling automated detection of spatial conflicts (e.g., overlap with protected zones) within the semantic graph.

To incorporate environmental stressors, we simulate storm conditions using historical data from NOAA’s HURDAT2 dataset. For each time step t , we compute the wind field using the Holland profile model and apply derived wind speeds for detailed calculations to turbine locations. Each turbine is assigned a status (e.g., *Shutdown*) based on cutout thresholds, and the simulation function Ψ inserts these into the evolving graph state G_t . Constraint rules are re-applied over this state, enabling the planner to identify vulnerable configurations.

An external solver \mathcal{O}_{pt} queries the graph G_t for violations and performance metrics (e.g., power yield), and proposes a revised layout. This optimization loop iterates until a design satisfies all constraints or converges to a fixed point.

4 Experimental Settings

Our experimental setup includes the following components:

Regulatory Documents: We compile a diverse corpus of publicly available planning documents, including Records of Decision (RODs) for existing U.S. wind leases, Environmental Impact Statements (EISs), and technical specifications. To ensure comprehensive coverage, we further incorporate key legislative documents such as the Outer Conti-

ental Shelf Lands Act (OCSLA) (ocs 1953), the National Environmental Policy Act (NEPA) (nep 1969), the Energy Policy Act of 2005 (epa 2005), and the Marine Mammal Protection Act (MMPA) (mmp 1972). These sources provide a mixture of spatial, ecological, and procedural constraints to test LLMs in extracting planning knowledge.

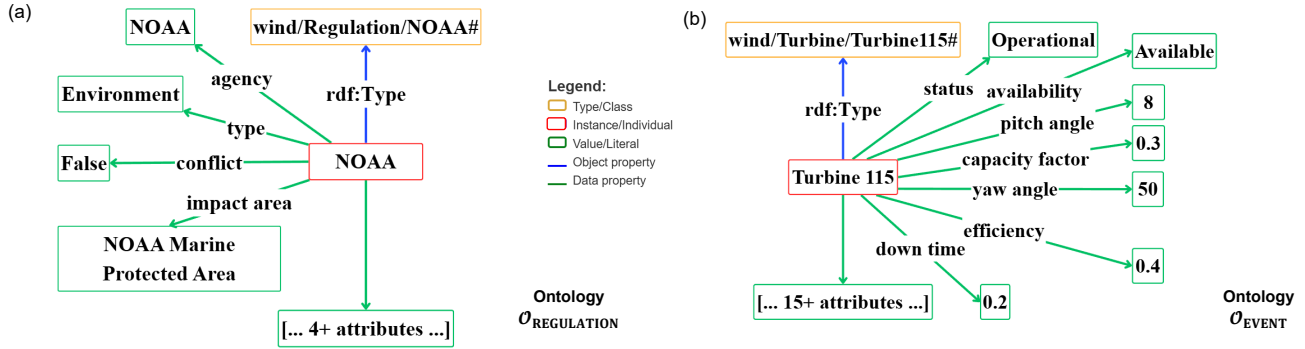
Layout Optimization Setup: We present a pilot study focused on the Maryland Offshore Wind Lease Area (OCS-A 0490), an approved lease zone along the U.S. Atlantic coast (Figure 3(b)). This site is selected due to its regulatory complexity and the fact that construction has not yet begun, making it a timely and realistic testbed for early-stage planning analysis. To provide a broader spatial context, Figure 3(a) illustrates the national distribution of offshore wind development zones, including marine protected areas, military exclusion zones, commercial shipping corridors, bathymetric features, and designated lease boundaries. These layers, sourced from the MarineCadastre.gov GIS database (Bureau of Ocean Energy Management (BOEM) and National Oceanic and Atmospheric Administration (NOAA) 2024), serve as the geospatial foundation for our modeling pipeline.

To encode and reason over planning constraints, we adopt Semantic Web technologies. Regulatory requirements extracted by LLMs are formalized into RDF triples, using OWL ontologies (W3C OWL Working Group 2012) to define spatial concepts, turbine properties, and regulatory constraints. Constraint enforcement is implemented using the Jena rule engine, which enables logical inference over turbine layouts. These structured constraints guide schematic layout generation for 121 turbines arranged in 13 rows, as specified by the Maryland lease design requirements. We use the TopFarm (Larsen and Réthoré 2013) Python package to optimize turbine coordinates, modeling wake interactions via the PyWake (Pedersen et al. 2023) library and the Bastankhah-Gaussian model. The turbine model follows the IEA 15-MW reference specification (240 m rotor diameter, 150 m hub height) (Gaertner et al. 2020). Wind conditions are represented using a 24-sector wind rose, each modeled with a Weibull distribution ($k = 2.5$, $A = 8.0$). The primary objective is to maximize the Annual Energy Production (AEP), a key metric of economic viability. Optimization is subject to two hard constraints: (1) a minimum spacing of five rotor diameters (1,200 m) to mitigate wake losses, and (2) strict adherence to polygonal lease boundaries defined via WKT geometry. We employ Stochastic Gradient Descent (SGD) over 400 iterations to arrive at the final layout.

Hurricane Simulation Setup: We assess hurricane simulations implemented in Python. Hurricane Sandy (2012) serves as a high-impact test case, selected for its status as one of the most destructive storms in recent Maryland history. It made landfall near Brigantine, New Jersey, on October 29, with sustained winds of approximately 80 mph (129 km/h), resulting in extensive flooding and infrastructure disruptions across the Chesapeake Bay region (Zimmerman 2014; New Jersey Department of Environmental Protection 2022; National Weather Service, U.S. Department of Commerce 2012). Figure 4(a) illustrates its trajectory. Key simulation parameters include a hub height of 150 m, a cut-

Document No.	Claude 3.5 Haiku		Gemini 1.5 Pro		GPT-4o-mini		DeepSeek-chat	
	Acc.	No. of Reg.	Acc.	No. of Reg.	Acc.	No. of Reg.	Acc.	No. of Reg.
17	0.800	15	0.692	13	1.000	8	0.692	13
20	0.875	16	0.950	20	0.800	5	0.944	18
21	0.778	18	0.870	23	0.900	10	0.895	19
27	0.778	15	0.913	23	0.750	8	0.625	8

Table 1: Model performance evaluations.



- **Subject (entity)** is the first part of an RDF statement that refers to who or what the RDF statement is about.
- **Predicate (attribute)** is the middle term (the linkage, or "verb") in an RDF statement which describes relationship between subject and object that always goes from subject to object.
- **Object (value)** is the final part of an RDF statement that can be another entity or a plain value.

Figure 2: Example knowledge structures for (a) Regulation and (b) Event domains.

out wind speed of 25 m/s, a Holland parameter of $B = 1.5$, and a radius of maximum winds of 50 km. Turbine response is governed by a shutdown protocol that activates when two conditions are met: (1) the storm enters a defined proximity threshold, and (2) hub-height wind speeds exceed the turbine’s operational cut-out limit. The semantic control model monitors three key turbine-level actions: pitch angle (e.g., blade feathering), yaw angle (e.g., alignment to wind), and operational status (e.g., shutdown). All turbine actions are logged in RDF format, generating a machine-readable operational states. Additional outputs include a time-stamped CSV of wind metrics, turbine states, and storm positions.

5 Experimental Results

Building on our proposed LSDTs framework and experimental design, this section presents results across three key modules of the pipeline.

5.1 Knowledge Structuring Performance

We structure our knowledge base using three core domains—infrastructure ($\mathcal{O}_{\text{INFRASTRUCTURE}}$), event ($\mathcal{O}_{\text{EVENT}}$), and regulation ($\mathcal{O}_{\text{REGULATION}}$). Each domain is supported by a dedicated ontology and follows the same LLM-driven pipeline for schema-compliant knowledge extraction. In this section, we focus on evaluating the performance of our knowledge structuring module in the regulatory domain ($\mathcal{O}_{\text{REGULATION}}$). Figure 2(a) illustrates a representative RDF graph from this domain. It encodes a constraint involving the National Oceanic and Atmospheric Administration

(NOAA), with properties such as *:hasRegulationDescription*, *:hasImpactArea*, and *:hasConflict*. The *:hasConflict* property serves as a dynamic compliance flag, updated during semantic reasoning to indicate whether the wind farm violates this rule. The zoomed-in view highlights the *False* value assigned to this property, confirming that the current layout satisfies the constraint. To assess extraction quality,

Row	Mean X	Std Dev X	Mean Y	Std Dev Y
1	68.43	61.32	136.89	87.76
2	72.31	55.18	58.96	107.90
3	65.14	49.18	35.24	18.17
4	41.63	38.81	49.75	32.86
5	73.29	42.36	44.09	30.72
6	68.01	57.71	26.10	17.50
7	66.73	66.62	28.05	21.64
8	79.44	55.45	38.37	20.53
9	77.18	42.92	27.16	19.89
10	67.15	63.48	33.84	12.69
11	68.70	71.34	47.91	31.09
12	69.40	77.09	70.12	35.50
13	82.87	63.93	109.51	39.85
Overall	67.98	54.83	51.81	54.95

Table 2: Positional deviation (meters).

we tested how well LLMs identify and formalize planning constraints from unstructured regulatory documents. First, we selected four representative documents and asked two

domain experts to manually annotate constraint regulations to serve as ground truth. The inter-annotator agreement, measured using Krippendorff’s Alpha for detailed calculations), achieved an average score of 0.80, indicating strong consistency and reliable reference data for model comparison.

We then evaluated the performance of our four selected LLMs by comparing their extracted regulations against the human-annotated ground truth. Each model received the same schema-constrained prompt. Prompting strategies such as few-shot examples and chain-of-thought reasoning guide the models to generate RDF-compatible triples. We report two metrics: (1) accuracy, defined as the proportion of correctly matched regulations over the total number of expert-annotated regulations, and (2) regulation count, the total number of regulations each model extracted from the document. A regulation is considered a correct match if its extracted structure semantically aligns with a ground-truth rule, regardless of wording differences. The results are summarized in Table 1.

Overall, all four models performed well, achieving relatively high accuracy across the evaluated documents. In particular, Gemini 1.5 Pro demonstrated consistently strong performance, combining both high accuracy and broad coverage. It produced the highest number of extracted regulations and achieved top accuracy in three out of four cases. Claude 3.5 Haiku also showed solid results, with balanced extraction quality and high correctness. We also observed distinct behavioral patterns across the models. For example, GPT-4o-mini reached perfect accuracy (1.000) on one document but extracted fewer regulations overall, suggesting a conservative strategy that favors precision over recall. DeepSeek-chat performed well when regulatory rules are explicitly presented in structured formats but struggled with more implicit or fragmented constraints. This was particularly evident in documents 17 and 22, which focus more on site assessment analysis. In contrast, documents 20 and 21, published by the Bureau of Ocean Energy Management (Bureau of Ocean Energy Management 2024), contain direct regulatory language and are easier for all models to interpret. Across the four models, a total of 289 unique and semantically validated regulations are extracted and encoded into RDF for downstream reasoning. These results confirm the feasibility of automating schema-compliant knowledge acquisition from heterogeneous regulatory corpora.

5.2 LLM-derived Wind Farm Layout

Next, we test whether the structured regulatory knowledge can guide compliant layout generation. Figure 3(c) presents a comparison between the LLM-derived layout (green dots) and the lease-based reference layout (red dots). Both layouts exhibit similar spatial patterns, suggesting that the regulatory graph constructed from text-extracted constraints can effectively guide the generation of realistic turbine configurations. This result demonstrates the feasibility of automating early-stage design workflows using LLMs, even in the absence of direct access to proprietary turbine placement data. To further quantify the spatial alignment between the two layouts, we compute the positional deviation for each

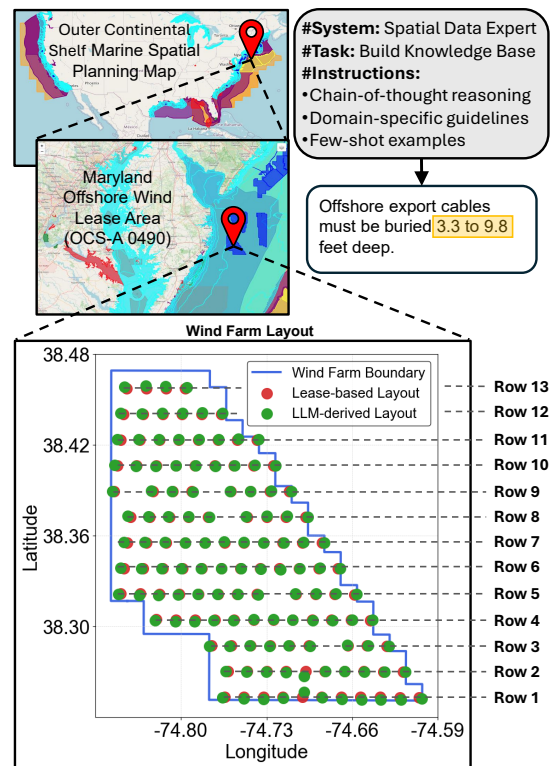


Figure 3: Pilot study of wind farm layout.

row of turbines. Specifically, we calculate the mean and standard deviation of longitude (X) and latitude (Y) differences across corresponding turbines within each row. These statistics capture both the average displacement and the spatial variability between the LLM-optimized and lease-based configurations. Table 2 summarizes the results.

On average, the LLM-derived turbine locations deviate by approximately 67.98 meters in the X direction and 51.81 meters in the Y direction, with standard deviations of 54.83 meters (X) and 54.95 meters (Y). While rows 1 and 2 exhibit slightly larger deviations due to reduced wake interactions and spacing constraints, the overall magnitude of difference remains within acceptable tolerances for schematic-level planning. These findings show that our LLM-driven framework is capable of producing spatially coherent, regulation-compliant layouts under realistic planning constraints.

5.3 LSDTs for Hurricane Sandy Simulation

Finally, to assess system-level adaptability, we simulate the impact of Hurricane Sandy across three phases: (1) **Phase 1:** Sandy tracks northward toward the Mid-Atlantic coast. (2) **Phase 2:** Tropical-storm-force winds reach Maryland’s Eastern Shore. (3) **Phase 3:** The storm moves inland and offshore wind conditions begin to subside.

During Phase 1 (Figure 4(b)), all turbines remain operational. Simulated wind speeds are 69 mph and the distance to the Maryland lease area is approximately 1,219 km, both of which fall below the intervention threshold. As Sandy advances in Phase 2 (Figure 4(c)), hazardous wind fields reach

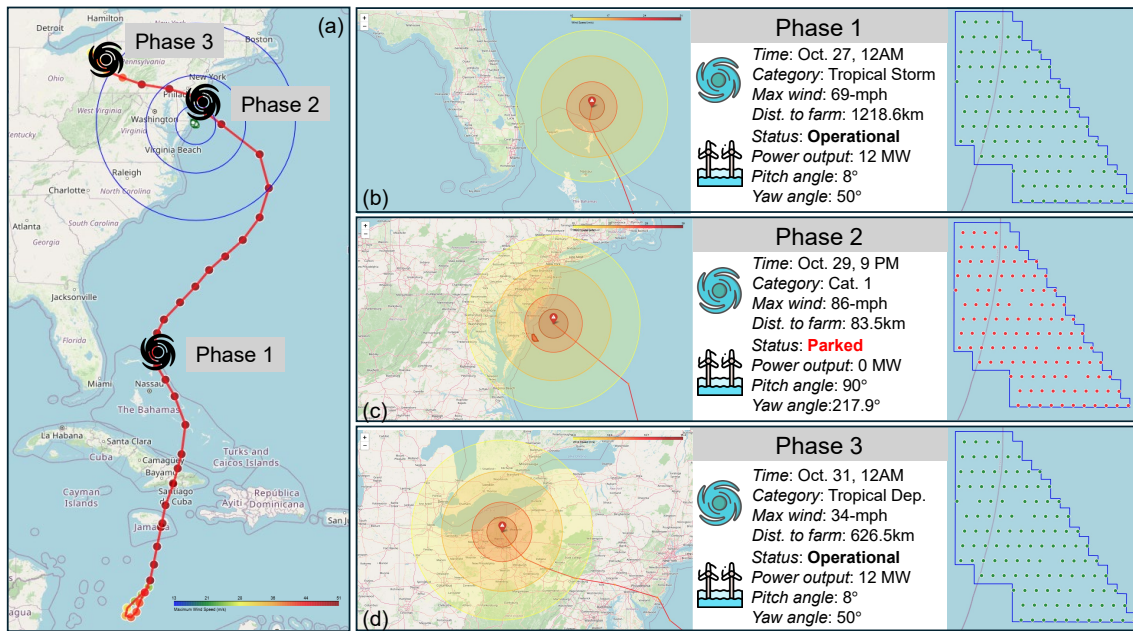


Figure 4: LSDT-driven turbine status changes during 2012 Hurricane Sandy.

the site, prompting the model to initiate turbine shutdowns. In response, turbine blades are pitched to 90 degrees to reduce aerodynamic loads, nacelles are yawed to align with the incoming wind direction (approximately 218 degrees), and the turbine status transitions to *Parked*.

This operational transition is formally captured using RDF triples, as illustrated in our event domain ontology (\mathcal{O}_{EVENT}) (Figure 2(b)). Taking Turbine #115 as an example, the ontology structure records the updated state via the triple: $(:Turbine115 :hasTurbineStatus :Parked)$. The RDF graph represents this turbine as a URI-linked entity, annotated with attributes such as pitch, yaw, wind speed, and drive type. The red arrow highlights the change in operational status, showcasing how simulation outputs are both machine-readable and semantically consistent with the OWL-defined schema. This confirms that turbine state changes are fully traceable within the semantic model and can be leveraged for automated control and monitoring. In Phase 3 (Figure 4(d)), turbines safely recover and return to their operational state as wind conditions stabilize.

6 Discussion and Conclusions

Our study presents LSDTs, a framework that integrates LLMs with DT architectures to address the challenge of retrieving unstructured regulatory and technical information for infrastructure planning. By developing an LLM-driven ontology extraction pipeline, our approach transforms complex planning documents into structured semantic graphs, which enables DTs to support regulation-aware decision-making. Our pilot study using the Maryland offshore wind farm case demonstrates that LSDTs can automate the integration of diverse planning constraints and support efficient layout simulation in different scenarios.

Beyond static validation, our results demonstrate that integrating LLM-extracted knowledge with dynamic simulation and compliance reasoning enables DTs to adapt to evolving scenarios, such as extreme weather events. The hurricane simulation illustrates how the LSDTs workflow can support simulation of infrastructure resilience and generate automated response to emerging hazards. This highlights the system's capacity for adaptivity across both the planning and operational phases even in complex environments.

Looking ahead, the generalizable architecture of LLM-driven DTs systems can support broader application in domains where policies and regulations are originally unstructured but require structured extraction, such as transportation and urban development. Our framework demonstrates that combining LLMs with DTs can accelerate data-driven, regulation-compliant planning and decision-making. However, key challenges remain in ensuring the reliability of LLM outputs and adapting the system to diverse environments. Future work could refine automated rule formalization and engage stakeholders to ensure alignment with societal values and legal standards.

Ethical Statement

The deployment of LSDTs for infrastructure planning presents both positive and negative societal implications. Our work provides key insights for more structured, efficient, regulation-compliant infrastructure design. By automating regulatory knowledge integration, LSDTs may streamline complex planning and reduce the efforts of manual compliance checking, particularly benefiting smaller organizations with limited technical resources. However, several limitations warrant careful consideration. Relying on LLMs to interpret regulations may introduce or amplify mis-

interpretations or omissions originating from training data or source documents. Errors in automated knowledge extraction or rule encoding could lead to flawed decision-making, with significant consequences in high-stakes environments. Moreover, the use of DTs and LLM-driven decision-making may also shift accountability from human planners to automated systems, raising concerns about transparency and responsibility. Therefore, system outputs must be subject to expert oversight, with safeguards in place to detect and correct errors.

Acknowledgments

This research was partially funded by the Department of Defense (DoD) Minerva Research Initiative, under the project “Semantic Foundations and Formal Methods for Evolutionary System-of-System Architectures,” with support from January to March 2025. We also acknowledge the valuable contributions of Juan Li, a software engineer at Google, who assisted with data cleaning and preprocessing, significantly enhancing the quality of the dataset used in this work.

References

1953. Outer Continental Shelf Lands Act. 43 U.S.C. § 1331 et seq.
1969. National Environmental Policy Act. 42 U.S.C. § 4321 et seq. Codified at 42 U.S.C. § 4321 et seq.
1972. Marine Mammal Protection Act of 1972. 16 U.S.C. § 1361 et seq.
2005. Energy Policy Act of 2005. Pub. L. 109–58, 119 Stat. 594.
- Abd Wahab, N. H.; Hasikin, K.; Lai, K. W.; Xia, K.; Bei, L.; Huang, K.; and Wu, X. 2024. Systematic Review of Predictive Maintenance and Digital Twin Technologies: Challenges, Opportunities, and Best Practices. *PeerJ Computer Science*, 10: e1943.
- Ambarita, E. E.; Karlsen, A.; Scibilia, F.; and Hasan, A. 2024. Industrial digital twins in offshore wind farms. *Energy Informatics*, 7(1): 5.
- Anthropic. 2024. Introducing Claude 3.5 Sonnet. <https://www.anthropic.com/news/claude-3-5-sonnet>.
- Bentley, C. 2012. *Prince2: a practical handbook*. Routledge.
- Broo, D. G.; Bravo-Haro, M.; and Schooling, J. 2022. Design and implementation of a smart infrastructure digital twin. *Automation in construction*, 136: 104171.
- Broo, D. G.; and Schooling, J. 2023. Digital twins in infrastructure: definitions, current practices, challenges and strategies. *International Journal of Construction Management*, 23(7): 1254–1263.
- Bureau of Ocean Energy Management. 2024. Offshore Wind Lease Areas. <https://www.boem.gov/renewable-energy/mapping-and-data/renewable-energy-gis-data>. Accessed July 2025.
- Bureau of Ocean Energy Management (BOEM) and National Oceanic and Atmospheric Administration (NOAA). 2024. MarineCadastre.gov. <https://marinecadastre.gov/>. Accessed: July 31, 2024.
- Chen, C.; Zhao, K.; Leng, J.; Liu, C.; Fan, J.; and Zheng, P. 2025. Integrating large language model and digital twins in the context of industry 5.0: Framework, challenges and opportunities. *Robotics and Computer-Integrated Manufacturing*, 94: 102982.
- Dagdelen, J.; Dunn, A.; Lee, S.; Walker, N.; Rosen, A. S.; Ceder, G.; Persson, K. A.; and Jain, A. 2024. Structured information extraction from scientific text with large language models. *Nature communications*, 15(1): 1418.
- DeepSeek AI. 2024. DeepSeek Coder: When the Large Language Model Meets Programming - Technical Report. Technical report, DeepSeek AI.
- DeepSeek-VL Team. 2024. DeepSeek-VL: Towards Real-World Vision-Language Understanding. <https://arxiv.org/abs/2404.14510>.
- Deng, M.; Fu, B.; Li, L.; and Wang, X. 2025. Integrating LLMs and Digital Twins for Adaptive Multi-Robot Task Allocation in Construction. *arXiv preprint arXiv:2506.18178*.
- Du, C.; Esser, S.; Nousias, S.; and Borrmann, A. 2024. Text2BIM: Generating Building Models Using a Large Language Model-Based Multi-Agent Framework. *arXiv preprint*.
- D’Amico, F.; Calvi, A.; Schiattarella, E.; Di Prete, M.; and Veraldi, V. 2020. BIM and GIS data integration: a novel approach of technical/environmental decision-making process in transport infrastructure design. *Transportation Research Procedia*, 45: 803–810.
- Fan, L.; Hua, W.; Li, L.; Ling, H.; and Zhang, Y. 2024. NPHardEval: Dynamic Benchmark on Reasoning Ability of Large Language Models via Complexity Classes. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*. Bangkok, Thailand: Association for Computational Linguistics.
- Flyvbjerg, B. 2007. Policy and planning for large-infrastructure projects: problems, causes, cures. *Environment and Planning B: planning and design*, 34(4): 578–597.
- Flyvbjerg, B. 2009. Survival of the unfittest: why the worst infrastructure gets built—and what we can do about it. *Oxford review of economic policy*, 25(3): 344–367.
- Gaertner, E.; Rinker, J.; Sethuraman, L.; Zahle, F.; Anderson, B.; Barter, G.; Abbas, N.; Meng, F.; Bortolotti, P.; Skrzypinski, W.; Scott, G.; Feil, R.; Bredmose, H.; Dykes, K.; Shields, M.; Allen, C.; and Viselli, A. 2020. Definition of the IEA 15-Megawatt Offshore Reference Wind Turbine. Technical Report NREL/TP-5000-75698, National Renewable Energy Laboratory (NREL).
- Gemini Team, Google. 2024. Gemini 1.5: Unlocking multimodal understanding across long contexts. Technical report, Google.
- Gürçan, Ö.; Falck, V.; Rousseau, M. G.; and Lima, L. L. 2025. Towards an LLM-Powered Social Digital Twinning Platform. *arXiv preprint*.
- Huang, Y.; Ato, W. X.; and Li, C. 2022. Information integration framework for a public–private partnership system of an urban railway transit project (Part A: System architecture). *Journal of Industrial Information Integration*, 25: 100244.

- Höltgen, L.; Zentgraf, S.; Hagedorn, P.; and König, M. 2025. Utilizing Large Language Models for Semantic Enrichment of Infrastructure Condition Data: A Comparative Study of GPT and Llama Models. *AI in Civil Engineering*, 4(14): 1–18.
- Institute, P. M. 2021. A Guide to the Project Management Body of Knowledge (PMBOK® Guide)—Seventh Edition and The Standard for Project Management. Project Management Institute.
- Jagatheesaperumal, S. K.; Yang, Z.; Yang, Q.; Huang, C.; Xu, W.; Shikh-Bahaei, M.; and Zhang, Z. 2023. Semantic-aware digital twin for metaverse: A comprehensive review. *IEEE Wireless Communications*, 30(4): 38–46.
- Jiang, L.; Shi, J.; Wang, C.; and Pan, Z. 2023. Intelligent control of building fire protection system using digital twins and semantic web technologies. *Automation in construction*, 147: 104728.
- Jiang, Y.; Tan, H.; Park, Y.; Zhao, S.; Li, M.; Yin, M.; and Gao, O. H. 2025. Automated Semantic-Aware Paradigm Extraction for Ubiquitous Digital Twins of Heterogeneous Subsystems in Smart Ports. Available at SSRN 5344488.
- Kaur, M. J.; Mishra, V. P.; and Maheshwari, P. 2019. The convergence of digital twin, IoT, and machine learning: transforming data into action. In *Digital twin technologies and smart cities*, 3–17. Springer.
- Lam, P.-D.; Gu, B.-H.; Lam, H.-K.; Ok, S.-Y.; and Lee, S.-H. 2024. Digital Twin Smart City: Integrating IFC and CityGML with Semantic Graph for Advanced 3D City Model Visualization. *Sensors*, 24(12): 3761.
- Larsen, G. C.; and Réthoré, P.-E. 2013. TOPFARM – A Tool for Wind Farm Optimization. *Energy Procedia*, 35: 317–324.
- Li, M.; Wang, R.; Zhou, X.; Zhu, Z.; Wen, Y.; and Tan, R. 2023. Chattwin: Toward automated digital twin generation for data center via large language models. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, 208–211.
- Madireddy, S.; Gao, L.; Din, Z. U.; Kim, K.; Senouci, A.; Han, Z.; and Zhang, Y. 2025. Large Language Model-Driven Code Compliance Checking in Building Information Modeling. *Electronics*, 14(11): 2146.
- Mchirgui, N.; Quadar, N.; Kraiem, H.; and Lakhssassi, A. 2024. The Applications and Challenges of Digital Twin Technology in Smart Grids: A Comprehensive Review. *Applied Sciences*, 14(23): 10933.
- Minerva, R.; Lee, G. M.; and Crespi, N. 2020. Digital twin in the IoT context: A survey on technical features, scenarios, and architectural models. *Proceedings of the IEEE*, 108(10): 1785–1824.
- Mousavi, Y.; Gharineiat, Z.; Karimi, A. A.; McDougall, K.; Rossi, A.; and Gonizzi Barsanti, S. 2024. Digital Twin Technology in the Built Environment: A Review of Applications, Capabilities and Challenges. *Smart Cities*, 7(5): 2594–2615.
- National Weather Service, U.S. Department of Commerce. 2012. Hurricane Sandy – 5 Year Anniversary. National Weather Service – Weather Forecast Office, New York, NY. Page does not list a publication or last-updated date; accessed 1 August 2025.
- New Jersey Department of Environmental Protection. 2022. Remembering Superstorm Sandy: Sandy 10 Years. <https://dep.nj.gov/sandy-10/>. Site metadata indicated “Published: 2.8 years ago” (≈ Feb 2023); accessed 1 Aug 2025.
- OpenAI. 2024. GPT-4o. <https://arxiv.org/abs/2405.07542>.
- Pedersen, M. M.; Forsting, A. M.; van der Laan, P.; Riva, R.; et al. 2023. PyWake: An open-source wind farm simulation tool. <https://doi.org/10.5281/zenodo.2562662>.
- Peng, J.; and Liu, X. 2023. Automated Code Compliance Checking Research Based on BIM and Knowledge Graph. *Scientific Reports*, 13: 7065.
- Su, Z.; Wang, Y.; Zhang, T.; and Zhang, G. 2022. Research on digital twin architecture of rule-based intelligent diagnosis of reactor system. In *2022 Global Reliability and Prognostics and Health Management (PHM-Yantai)*, 1–5. IEEE.
- Sun, Q.; Luo, Y.; Zhang, W.; Li, S.; Li, J.; Niu, K.; Kong, X.; and Liu, W. 2025. Docs2KG: A Human-LLM Collaborative Approach to Unified Knowledge Graph Construction from Heterogeneous Documents. In *Companion Proceedings of the ACM on Web Conference 2025*, 801–804.
- Vivid Solutions; and LocationTech. 2025. JTS Topology Suite. <https://locationtech.github.io/jts/>. Originally developed by Vivid Solutions. Accessed: Aug 1, 2025.
- W3C OWL Working Group. 2012. OWL 2 Web Ontology Language Document Overview (Second Edition). W3C Recommendation.
- Wang, Y.; Fu, T.; Xu, Y.; Ma, Z.; Xu, H.; Du, B.; Lu, Y.; Gao, H.; Wu, J.; and Chen, J. 2024. TWIN-GPT: digital twins for clinical trials via large language model. *ACM Transactions on Multimedia Computing, Communications and Applications*.
- Yang, L.; Luo, S.; Cheng, X.; and Yu, L. 2025. Leveraging Large Language Models for Enhanced Digital Twin Modeling: Trends, Methods, and Challenges. *arXiv preprint*.
- Zhang, J.; Ellwein, C.; Heithoff, M.; Michael, J.; and Wortmann, A. 2025. Digital Twin and the Asset Administration Shell: An Analysis of the Three Types of AASs and Their Feasibility for Digital Twin Engineering. *Software and Systems Modeling*, 24(3): 771–793.
- Zhao, Z.; Shen, L.; Yang, C.; Wu, W.; Zhang, M.; and Huang, G. Q. 2021. IoT and digital twin enabled smart tracking for safety management. *Computers & Operations Research*, 128: 105183.
- Zhou, J.; and Ma, Z. 2025. Named entity recognition for construction documents based on fine-tuning of large language models with low-quality datasets. *Automation in Construction*, 174: 106151.
- Zimmerman, R. 2014. Planning restoration of vital infrastructure services following Hurricane Sandy: lessons learned for energy and transportation. *Journal of Extreme Events*, 1(01): 1450004.