INTERNATIONAL JOURNAL OF INTELLIGENT DATA AND MACHINE LEARNING (IJIDML)

eISSN: 3087-4297

Volume. 02, Issue. 11, pp. 14-24, November 2025"

Assessing The Interoperability and Semantic Readiness of BIM And IFC Data for AI Integration in the Architecture, Engineering, And Construction Industry: A Systematic Review

Bima Satria Nugraha
Faculty of Civil and Environmental Engineering, Bandung Institute of Technology, Bandung, Indonesia

Professor Anindya Larasati
Department of Architecture and Design, National University of Singapore, Singapore, Singapore

Dr. Huỳnh Chí Dũng
College of Construction Engineering, Hanoi University of Civil Engineering, Hanoi, Vietnam

Article received: 10/09/2025, Article Revised: 26/10/2025, Article Published: 01/11/2025

ABSTRACT

Purpose: This systematic review aims to critically assess the current state of Building Information Modeling (BIM) and Industry Foundation Classes (IFC) data interoperability and semantic readiness for scalable integration with Artificial Intelligence (AI) applications across the Architecture, Engineering, and Construction (AEC) industry.
Design/Methodology/Approach: A Systematic Literature Review (SLR) was conducted, adhering to PRISMA guidelines, analyzing key research focused on the intersection of BIM, IFC, and AI. A conceptual framework categorizing AI-ready data into five pillars—Structural Consistency, Semantic Completeness, Geometric Fidelity, Temporal Coherence, and Contextual Richness—was developed to synthesize findings.
Findings: While AI applications, notably in predictive maintenance, risk assessment, and generative design, exhibit clear reliance on BIM/IFC data, the implementation is often impeded by significant data quality challenges. The core issue lies in the semantic gap: IFC, designed primarily for data exchange, frequently lacks the explicit, complete, and consistently structured information required for machine learning algorithms. Current approaches heavily rely on labor-intensive pre-processing, graph-based data transformations, or domain-specific custom property sets, compromising true interoperability. Furthermore, the handling of geometric and topological data within IFC frequently suffers from inaccuracies that render it unsuitable for highly sensitive computational tasks like automated quantity take-off and robot navigation.
Originality/Value: This review introduces a novel framework for assessing AI-ready BIM data and systematically maps the specific data requirements of various AI applications to the current limitations of the IFC schema. It provides a foundational critique, guiding future research toward developing the necessary semantic middleware, robust geometric validation tools, and standardization efforts for achieving seamless BIM-AI integration.

KEYWORDS

Building Information Modeling, Industry Foundation Classes, Artificial Intelligence, Semantic Interoperability, Construction Technology, Data Readiness, Knowledge Graph.

INTRODUCTION

1.1. Contextualizing the Digital Transformation in Construction

The Architecture, Engineering, and Construction (AEC) industry faces immense pressure to enhance its performance metrics, which includes mitigating chronic issues such as cost overruns, project delays, and safety incidents. Historically, the sector has lagged behind other industries in adopting digital technologies, often characterized by fragmented workflows and reliance on traditional, document-centric processes. This landscape, however, is rapidly evolving due to the imperative for innovation driven by global urbanization and sustainability mandates. At the forefront of this shift lies Building Information Modeling (BIM), which fundamentally changes how design, construction, and operation information is created, managed, and shared throughout the entire asset lifecycle. BIM is not merely a geometric modeling tool; it serves as a shared knowledge resource that provides a reliable basis for decisions during the life cycle of a facility, defining an operational paradigm shift for project delivery. The adoption of BIM is associated with improved coordination, enhanced communication, and a higher potential for error reduction throughout the project lifecycle.

1.2. The Ascendance of Artificial Intelligence in AEC

Parallel to the rise of BIM, Artificial Intelligence (AI) has emerged from theoretical concepts to a transformative technology across various sectors. While the foundational ideas of AI trace back to the mid-20th century, modern advancements, specifically in Machine Learning (ML) and Deep Learning (DL), have unlocked unprecedented computational power, driven by algorithmic sophistication and the proliferation of big data. The AEC industry stands to benefit profoundly from AI by moving beyond static models toward predictive, proactive, and generative decision-making capabilities. Current AI applications in AEC are wide-ranging, encompassing sophisticated tasks such as optimizing facility design through generative algorithms, predicting maintenance needs for mechanical, electrical, and plumbing (MEP) components, monitoring construction site safety in real-time using computer vision, and automating quality control processes through integrated sensory data.

The ambition of these AI models—to learn, predict, and optimize complex phenomena—places a critical dependence on the quality and structure of the input data. Specifically, AI algorithms require data that is not only voluminous but also high-fidelity, semantically rich, and computationally accessible. In the context of the built environment, this data is primarily intended to be supplied by the integrated models generated within the BIM environment. AI's ability to process vast quantities of data predicts improved risk assessment and more optimized resource allocation across complex construction projects.

1.3. The Critical Role of IFC and Data Standards

The power of BIM is inherently constrained by the challenge of interoperability, the ability for different software applications to exchange and utilize data reliably. This challenge is rooted in the proprietary nature of many BIM authoring tools. To overcome this fragmentation, the Industry Foundation Classes (IFC) standard was developed by buildingSMART International. IFC is a vendor-neutral, non-proprietary data model that describes building and construction industry data, serving as the common language for the digital exchange of information. Theoretically, IFC should function as the key that unlocks the full potential of BIM data for computational applications, including AI. By providing a structured, hierarchical schema that defines elements, properties, and relationships within a building model, IFC offers a pathway to a singular, accessible data source for diverse AI models across different life cycle stages. The standardization offered by IFC is expected to facilitate the smooth handover of information between project stakeholders and disparate software platforms.

1.4. Problem Statement and Research Gap

Despite the foundational potential of BIM and the standardization provided by IFC, the seamless and scalable integration of AI remains a significant hurdle. AI models, particularly those based on advanced ML and DL, demand data that is explicitly complete and semantically unambiguous. The current reality reveals a pervasive semantic gap between the information structure defined by IFC (designed primarily for data exchange/archiving) and the strict data requirements of AI algorithms (designed for pattern recognition and automated reasoning). Models exported via IFC often suffer from structural inconsistencies, missing critical non-geometric data (such as properties necessary for energy analysis or cost estimation), and ambiguous semantic classification. This insufficient readiness necessitates significant, labor-intensive pre-processing, compromising the efficiency and scalability of AI deployment.

Existing literature touches upon specific BIM-AI applications; however, a systematic, overarching synthesis that critically evaluates the readiness of the underlying IFC data structure through the exacting lens of diverse AI requirements is largely absent. Specifically, the literature lacks a unified framework that evaluates both the semantic and geometric/topological fidelity of IFC data required for advanced computational models. This review seeks to address this critical gap by systematically analyzing the current methodologies for preparing BIM/IFC data for AI and identifying the persistent interoperability and data quality challenges that collectively impede scalable, industry-wide AI deployment.

1.5. Research Objectives

This systematic review is structured around the following objectives:

- To systematically map and classify current AI applications that specifically utilize BIM/IFC data, identifying their core data dependencies.

● To critically analyze the state of semantic enrichment and data completeness within the IFC schema necessary for advanced computational models.

● To evaluate the common methodological approaches for pre-processing, converting, and transforming IFC data into a machine-consumable format.

● To identify and articulate the persistent interoperability and data quality challenges that collectively impede the scalable and robust deployment of AI throughout the AEC sector.

1.6. Structure of the Article

Following this introduction, Section 2 details the Systematic Literature Review (SLR) Methodology, including the search strategy, eligibility criteria, and the conceptual framework developed for AI-ready data. Section 3 presents the Results, classifying AI applications, mapping data requirements, and documenting common pre-processing methodologies and identified challenges. Section 4 provides a deep Discussion of the findings, focusing on the limitations of IFC, the role of knowledge graphs, and the critical importance of geometric fidelity. Finally, Section 5 offers the Conclusion and proposes future research directions.

2. METHODS

2.1. Systematic Review Protocol

This study employed a Systematic Literature Review (SLR) methodology, designed to provide a comprehensive, unbiased, and repeatable synthesis of existing research. The protocol adhered to the widely accepted guidelines established by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement, ensuring the highest level of rigor and transparency in the selection and reporting process. The SLR approach was chosen over a narrative review to manage the vast and multidisciplinary body of literature spanning construction technology, computer science, and engineering management, thereby minimizing selection bias and promoting evidence-based conclusions.

2.2. Search Strategy and Data Sources

The search strategy targeted key electronic databases recognized for their coverage of engineering, computing, and construction literature, including Web of Science, Scopus, and Google Scholar. The primary search query was constructed using a combination of keywords, separated by Boolean operators:

(BIM OR "Building Information Model" OR IFC OR "Industry Foundation Classes") AND ("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR AI OR ML) AND (Construction OR AEC OR "Built Environment")

This combination of terms was designed to capture research that explicitly connects the foundational data models of the built environment (BIM/IFC) with advanced computational methods (AI/ML). The search was refined to include only peer-reviewed journal articles, conference proceedings, and reputable technical reports published up to the cutoff date of the review. The initial search yielded a substantial number of articles, which were then subjected to rigorous screening.

2.3. Eligibility Criteria and Study Selection

The retrieved documents underwent a two-phase selection process based on the following eligibility criteria:

Inclusion Criteria:

1. The study must explicitly involve the application or discussion of an AI/ML technique.

2. The study must utilize or directly discuss the use of BIM or IFC data as the primary or critical input source for the AI model.

3. The article must be published in English and be a full-text peer-reviewed source (journal or conference paper).

4. The article must be available in full text for review.

Exclusion Criteria:

1. Studies focusing on general AI applications without specific reference to BIM/IFC data.

2. Studies focusing purely on BIM implementation without computational AI application.

3. Short abstracts, editorials, and non-academic publications.

In the first phase, titles and abstracts were screened for relevance to the core intersection of BIM, IFC, and AI. In the second phase, the full text of the remaining articles was assessed against the inclusion criteria. Any conflicts regarding inclusion were resolved through consensus among the review team.

2.4. Data Extraction and Synthesis

For each included study, the following data points were systematically extracted and recorded:

- AI Application Domain: (e.g., Safety, Scheduling, Energy Analysis, Generative Design)

- AI Technique Employed: (e.g., Convolutional Neural Networks, Support Vector Machines, Expert Systems)

- BIM/IFC Version and Source: (e.g., IFC4, Revit, custom schema)

- Data Pre-processing/Transformation Method: (e.g., graph conversion, feature engineering, vectorization)

- Identified Data Challenges/Limitations: (The specific issues encountered with the BIM/IFC data).

The data synthesis involved a rigorous thematic analysis. Extracted challenges were clustered into recurring themes to identify the most significant and pervasive interoperability issues. This thematic clustering informed the development of a conceptual framework which is central to the subsequent analysis.

2.5. The Conceptual Framework for AI-Ready Data

To structure the analysis of the results, a conceptual model defining the necessary characteristics of an "AI-Ready" BIM model was established. This framework posits that for BIM/IFC data to be successfully consumed by sophisticated AI algorithms, it must satisfy five critical pillars. The absence or weakness of any one pillar predicts failure or significant friction in the AI integration process.

1. Structural Consistency: Adherence to the standardized IFC schema without model redundancy or inconsistent entity usage, ensuring the computational path is predictable. This relates to the formal compliance of the data with the schema definition.

2. Semantic Completeness: The inclusion of all necessary non-geometric properties (e.g., fire rating, material strength, maintenance date) explicitly required by the AI task, beyond basic geometric representation. This addresses the richness of the information.

3. Geometric Fidelity: The accuracy, precision, and topological correctness of the physical geometry, crucial for tasks like quantity take-off and robot navigation. This addresses the quality of the spatial model.

4. Temporal Coherence: The ability of the model to represent or link to time-dependent data (4D BIM), essential for scheduling, progress monitoring, and asset lifecycle management.

5. Contextual Richness: The seamless integration of external, real-world data sources (e.g., IoT sensor readings, site photographs) with the BIM elements, necessary for predictive and operational AI models, establishing the digital twin connection.

3. RESULTS

3.1. Classification of AI Applications Utilizing BIM/IFC Data

The review of the collected literature revealed a strong concentration of AI research across three primary AEC domains, each with distinct data requirements and vulnerabilities related to the IFC data source.

3.1.1. Design and Analysis (Generative Design, Code Compliance Check)

AI applications in the design phase focus on automating iterative tasks and optimizing performance outcomes. Generative design utilizes algorithms to explore a vast solution space based on performance criteria (e.g., structural load, daylighting, spatial layout), often requiring highly structured parametric input data. Automated code compliance checking relies on rule-based or machine learning systems to flag potential violations against building codes or project specifications by comparing model properties against regulatory requirements.

- Data Requirement Analysis: These applications exhibit an extremely high demand for semantic completeness and geometric fidelity. The AI models must accurately and unambiguously identify specific building elements (e.g., walls, doors, stairs) and their related properties (e.g., fire rating, material). IFC's hierarchical structure and property sets are heavily utilized here. However, studies show that models often lack the specific, granular detail needed, compelling researchers to develop complex, project-specific ontologies or extend the IFC schema with custom properties, thus undermining universal interoperability. The success of automated code checking is directly associated with the initial quality of property assignment within the BIM model.

3.1.2. Construction Management (Scheduling, Cost Prediction, Safety Monitoring)

In construction management, AI leverages BIM data to enhance efficiency and mitigate risk. ML models have been developed for predicting construction cost and duration based on historical data extracted from project models. Predictive maintenance planning for MEP systems utilizes BIM data (component location, type) integrated with IoT sensor data to forecast potential failures. Furthermore, AI-driven safety systems often use BIM geometry for collision detection, path planning for unmanned ground vehicles (UGVs), and identifying high-risk zones.

● Data Requirement Analysis: The core requirement here shifts toward Temporal Coherence and Contextual Richness. AI scheduling models require an explicit link between the 3D geometry and the time-based activities (4D BIM). Safety and predictive maintenance models necessitate the seamless fusion of static BIM attributes with dynamic, real-time data streams (e.g., location, sensor readings). Studies reveal significant difficulties in extracting reliable time-related parameters directly from current IFC instances, often requiring the use of middleware or external scheduling software that is manually linked back to the model elements. The accurate prediction of risks is predicted on the seamless integration of geometrical model data with contextual site information.

3.1.3. Facility Management and Operations (Energy Efficiency, Digital Twin)

The operations phase benefits from AI integration into Digital Twins for long-term asset performance. Models for optimizing energy efficiency are prominent, requiring detailed thermal properties, space boundaries, and HVAC system information. Automated guidance systems also rely on indoor path planning capabilities derived from the building's spatial model.

● Data Requirement Analysis: These applications critically depend on the correct and precise representation of space boundary data and semantic completeness regarding building performance properties. A pervasive challenge in this domain is the inaccuracy of space boundary definitions within IFC models, which directly impacts the reliability of energy simulation models. Researchers consistently find they must manually verify or reconstruct the spaces from the 3D geometry, highlighting a significant limitation in the practical utility of current IFC exports for energy-focused AI. The accuracy of energy performance modeling is closely associated with the correct topology of the building elements as represented in the IFC file.

3.2. Mapping Data Requirements to IFC Schema

The review indicates that the most frequently targeted IFC entities are IfcProduct, IfcElement, and the geometric representation items like IfcProductDefinitionShape. AI models generally succeed in classifying basic geometric objects but struggle with the non-geometric, semantic attributes.

A crucial finding is the high reliance on custom property sets (Psets) when attempting to meet the demands of advanced AI. While IFC provides the structure for adding properties, the absence of universally applied standard Psets for specific domain tasks (e.g., a standardized Pset for machine learning-relevant maintenance data) forces project teams to create unique, non-standard attributes. This process ensures the immediate utility of the data for a single project but simultaneously creates a significant interoperability barrier for external AI models, which cannot universally interpret these unique schema extensions. The result is that a substantial portion of the effort in BIM-AI research is devoted to manual data augmentation and mapping post-IFC export, confirming the inadequacy of the native IFC structure for many cutting-edge AI tasks. This manual enrichment predicts a higher friction point for scalable, multi-project AI deployment.

3.3. Methodologies for Interoperability and Data Pre-processing

To bridge the gap between the complex, textual-based IFC file structure and the matrix/vector input required by most AI models, various pre-processing methodologies have been employed:

1. Graph-Based Transformation: This is one of the most promising approaches. Researchers utilize Knowledge Graphs (KGs) to represent BIM data. IFC's structure, which is inherently object-oriented and graph-like (defining objects and their relationships), lends itself well to KG conversion. This transformation allows AI models to utilize the explicit relationships between elements, providing a richer context than simple attribute lists.

2. Vectorization and Feature Engineering: For tasks like cost prediction or duration estimation, researchers often flatten the IFC data into tabular features. This typically involves manually engineering features (e.g., total wall area, count of windows per floor) that aggregate the geometric and semantic data into numerical vectors suitable for classical ML algorithms. This process, however, risks losing the rich topological and relational context inherent in the model.

3. Visual Programming Tools: The use of visual programming environments (e.g., Dynamo or Grasshopper) is prevalent, enabling researchers to automate the extraction and filtering of specific data subsets from the model before exporting or transforming them for AI input. This addresses model complexity but remains constrained by the initial quality of the authoring model.

3.4. Identified Data Quality and Semantic Challenges

The collective evidence points to three dominant data quality challenges that persistently impede BIM-AI integration:

3.4.1. Structural Inconsistency and Model Redundancy

The translation process from proprietary software to the neutral IFC format is susceptible to errors and inconsistencies. Models exported from different authoring tools, even when ostensibly compliant with the same IFC standard (e.g., IFC4), can exhibit variations in how entities are defined or aggregated. Furthermore, within a single model, redundancy often occurs, such as multiple instances of the same property, or the inconsistent use of IfcSpace versus IfcZone, leading to ambiguity for AI classifiers. AI models that expect a predictable, structurally consistent input often fail when exposed to the natural variability of real-world IFC exports. This structural volatility predicts increased complexity in developing universal AI parsers.

3.4.2. Semantic Ambiguity and Missing Data

This is perhaps the most fundamental challenge. AI requires explicit semantics. If an element in an IFC file is labeled with a generic class (e.g., IfcElementAssembly) without the detailed, required properties, the AI model cannot infer the meaning. Studies on automated code checking consistently highlight the issue of missing properties, such as fire ratings or accessibility information, which are critical for the task but often omitted during the design or modeling stage. Furthermore, the generic nature of many IFC entity names can lead to semantic ambiguity that demands manual human intervention for accurate labeling, a non-starter for large-scale automation. The lack of standardized terminology for non-geometric parameters is associated with higher data pre-processing effort.

3.4.3. Granularity Mismatch

The concept of Level of Detail (LOD) in BIM relates to the geometric fidelity, while Level of Information Need (LOIN) relates to the necessary semantic content. Many AI applications require an LOD/LOIN that exceeds what is typically produced for standard project milestones. For instance, a detailed ML model for predictive equipment failure needs sub-component properties and manufacturer data (high LOIN), which are rarely included in standard IFC exports (low semantic completeness). This mismatch in granularity forces researchers to either simplify the AI task or invest enormous effort in manually enriching the models. The low LOIN in many IFC models predicts limited scope for highly specialized AI applications.

4. DISCUSSION

4.1. Re-evaluating the IFC Standard for AI-Readiness

The findings strongly suggest that while IFC has succeeded as a mechanism for data exchange, its current structure, even in its most recent versions, presents limitations when treated as a directly computable knowledge base for advanced AI. The standard was fundamentally designed to capture and transfer human-readable design intent and geometric representation, not necessarily to optimize data retrieval and relational querying for machine reasoning. The lack of inherent mechanisms to define complex, non-geometric relationships or temporal dependencies within the core IFC schema hinders its application in dynamic AI tasks like scheduling optimization or process simulation. Future evolution of the IFC standard must move beyond entity-property lists toward concepts that support richer, domain-specific ontologies and explicit relationship modeling to facilitate AI integration. This transition is crucial for realizing the vision of a truly "Smart Construction" environment where machine inference is a seamless component of the digital workflow.

4.2. The Interplay of BIM, IFC, and Knowledge Graphs

The prevalence of graph-based methodologies in the reviewed literature validates the hypothesis that an intermediate, semantic layer is necessary to bridge the IFC-AI gap. Knowledge Graphs (KGs) transform the hierarchical IFC structure into a flexible, queryable graph network (nodes and edges), where relationships are explicitly defined and easily traversable by AI agents. This approach successfully addresses the issues of semantic ambiguity and structural inconsistency by normalizing the data and adding an external, custom ontology layer tailored to the AI task.

However, the current implementation of KGs relies heavily on project-specific manual mapping and ontology creation. The next crucial step involves developing automated tools capable of generating KGs from raw IFC files using industry-standard semantic web technologies (e.g., RDF, OWL) and predefined, globally accepted AEC ontologies. If the industry can standardize the ontology used to interpret and enrich the IFC schema, the resulting KGs could serve as the universal semantic middleware necessary for scalable AI deployment. This semantic layer predicts a substantial reduction in the data preparation time for novel AI applications.

4.3. Addressing Data Completeness: The Custom Pset Dilemma

The recurring reliance on custom property sets highlights a critical tension: the need for project-specific detail versus the mandate for universal interoperability. While custom Psets are necessary to capture the unique information required for a specialized AI model (e.g., an ML classifier for a specific type of connection joint), their unmanaged proliferation erodes the standardization that IFC attempts to enforce.

To resolve this, the AEC industry, in partnership with buildingSMART and AI researchers, must establish and publish a set of Minimum AI Data Requirements (MAIDR). These MAIDR could take the form of standardized, domain-specific IFC Psets (e.g., a "Pset_AI_Energy_Analysis" or "Pset_AI_Safety_Risk") that define the mandatory minimum information required within a BIM model to enable common AI use cases. This shift would provide model authors with a clear target for data completeness, moving beyond general LOD requirements to specific LOIN targets necessary for machine consumption, thereby managing the proliferation of incompatible custom Psets. This standardization effort is expected to improve the reusability of BIM data across different AI platforms.

4.4. A Deeper Dive: The Critical Role of IFC Geometric and Topological Data in AI

While much of the discussion surrounding AI-readiness focuses on the semantic and non-geometric properties, the core geometric representation embedded within the IFC file presents its own set of profound challenges for computational analysis. Geometric data is not merely for visualization; it is the fundamental input for a wide array of AI-driven tasks, including automated quantity take-off, clash detection, robot path planning, visual monitoring, and structural analysis model generation. The limitations in how IFC handles geometric fidelity and topology directly impede the successful execution and scalability of these applications.

4.4.1. The Ambiguity of Boundary Representation (B-Rep) in IFC

IFC primarily utilizes Boundary Representation (B-Rep) to define the 3D geometry of building elements. B-Rep describes a solid object by defining the boundaries that separate its interior from its exterior—typically faces, edges, and vertices. While mathematically sound, the translation of complex B-Rep geometry from proprietary modeling kernels into the neutral IFC format is a common source of error. AI models, particularly those performing automated spatial reasoning or structural analysis, require mathematically perfect solid geometry.

Errors that frequently occur in the B-Rep conversion include:

- Non-Manifold Geometry: A condition where the solid geometry is mathematically ill-defined (e.g., two faces share only an edge, or multiple solids share a single vertex). When an AI algorithm, such as one for Finite Element Model (FEM) generation, attempts to mesh a non-manifold body, the process often fails or produces unreliable results, highlighting the need for clean topological data. The occurrence of non-manifold geometry in IFC models predicts failure in automated structural meshing algorithms.

- Gaps and Overlaps: Slight numerical inaccuracies during the export process often result in small gaps between adjacent components (e.g., a wall and a slab) or slight overlaps. These errors are often invisible to the human eye in a BIM viewer but are catastrophic for automated quantity take-off (QTO) models, which rely on precise boundary intersections for volumetric calculations. An AI system attempting to calculate the required sealant volume, for example, would return an unreliable result if the gap tolerance is not modeled explicitly and correctly in the IFC file. The presence of these micro-gaps is associated with significant overestimation or underestimation in automated QTO processes.

- Face Normal Vectors: For advanced visual AI and robotic applications, the direction of the surface normal vectors for the B-Rep faces is critical. Computer vision algorithms using geometric reconstruction rely on these vectors to understand orientation and visibility. Inconsistencies in IFC exports can lead to "flipped" normal vectors, which confuse these AI systems, leading to inaccurate object recognition or failed robot navigation paths, particularly in complex interior environments. Incorrect normal vector representation predicts errors in visual simultaneous localization and mapping (SLAM) systems.

The challenge is further compounded because the IFC specification allows for multiple ways to define geometry (e.g., B-Rep, Swept Solid, Clipped Half Spaces), and the choice of representation can significantly affect the downstream AI utility. Swept Solids are often simpler and more computationally efficient for basic objects but fail to capture the complexity required for detailed mechanical parts, necessitating the use of the more complex, but error-prone, B-Rep. The lack of a standardized geometric representation across all elements predicts a need for multi-protocol geometric parsers in AI frameworks.

4.4.2. Geometric Abstraction for Computational Efficiency

AI models, especially those operating at scale or in real-time environments (like on-site robotic systems), cannot afford to process the full, complex geometry of a high-LOD BIM model. The sheer number of polygons in a large construction project IFC file is computationally prohibitive. This necessitates a process of geometric abstraction, where the IFC data must be simplified without losing the critical topological relationships.

A key methodological challenge for researchers involves automatically converting the dense 3D B-Rep into a sparse, graph-based topological model. This transformation focuses on representing the connectivity and adjacency of building elements rather than their precise physical shape. For example, a Directed Representative Graph (DRG) can be generated from MEP systems using BIM data, allowing an AI model to quickly analyze flow, connectivity, and system paths without having to perform complex geometric intersection calculations. The nodes in this graph represent components (e.g., a pipe, a valve), and the edges represent their physical connections. This shift from geometric fidelity to topological fidelity is crucial for the efficient execution of path planning, interference checking, and structural network analysis AI models. Topological models are predicted to significantly enhance the computational speed of network-based AI analyses.

However, the AI must first reliably extract the correct connectivity from the IFC file. Errors in the IFC model, such as a component's connection port not being perfectly aligned in the B-Rep, can lead to a false negative (a missing connection) in the generated topological graph, rendering the subsequent AI analysis useless. Researchers are increasingly turning to advanced graph-based methodologies that incorporate tolerance limits to infer connection intent, even when the underlying IFC geometry is slightly flawed. This inference, however, introduces a level of uncertainty into the AI result, as the model is making an assumption based on proximity rather than explicitly modeled connectivity. This reliance on tolerance-based inference suggests that the semantic completeness of the connectivity data in the source model is often lacking.

4.4.3. The Challenge of Spatial Reasoning and Boundary Definition

For AI applications in facility management, energy analysis, and emergency response, the accurate definition of spatial boundaries is paramount. The IFC standard defines entities like IfcSpace and uses relationship entities like IfcSpatialStructureElement to organize the model hierarchically. However, a common practical failure point is the generation of the IfcSpaceBoundary.

Energy analysis AI models rely on precisely defined external and internal boundaries to calculate heat transfer and air exchange. If the space boundary generated from the architectural model is not topologically closed or if the boundary is incorrectly mapped to an adjacent wall or slab element, the energy simulation will produce incorrect results. Reviewing case studies consistently shows that:

1. Missing or Incorrect Boundary Mapping: The link between the abstract IfcSpace entity and the specific faces of the surrounding geometric elements (walls, ceilings, floors) is often missing or incorrectly established during IFC export. This forces AI pre-processing routines to recalculate the space boundaries from scratch, which is computationally expensive and prone to its own set of geometric tolerance errors. The manual or automated recalculation of space boundaries predicts significant time expenditure in the energy modeling workflow.

2. Handling of Openings and Penetrations: AI models for emergency path planning need to reliably identify openings (doors, windows, portals) and their connectivity to other spaces. If an IfcOpeningElement is not correctly referenced to the wall it penetrates, the path-finding AI will fail to recognize the traversable space. The complexity of dealing with non-rectangular openings or penetrations further strains the reliability of the underlying IFC geometry for automated spatial reasoning. The robustness of AI-driven path planning is strongly associated with the accurate, topological representation of traversable openings.

The difficulty in reliably extracting geometric and topological data for computational tasks mandates the development of more robust IFC validation and repair tools. These tools, ideally powered by AI themselves, would preprocess the IFC file to automatically detect and correct topological inconsistencies before the data is fed into a high-value application like robotic control or FEM generation. Until such tools are widely adopted, the scalability of geometrically-dependent AI applications remains severely constrained by the unpredictable quality of the source IFC data. The current state of IFC geometric fidelity is seen to predict high technical overhead for integration with advanced computational systems.

4.5. Limitations of the Current Review and Future Research Directions

The systematic review, while comprehensive, is subject to certain limitations that inherently constrain the generalizability of the findings. The search strategy, despite being broad, is constrained by the available, peer-reviewed literature, potentially leading to a publication bias toward successful case studies or specific domains (e.g., energy analysis) where BIM-AI research is most mature. Furthermore, the reliance on publicly available articles may exclude industry-specific, proprietary research where AI integration with BIM/IFC data may be more advanced but is not disseminated in academic forums.

Based on the synthesis, several crucial directions for future research are identified:

1. Development of AI-Ready Validation Metrics: Future work should focus on establishing quantitative metrics and automated tools to assess the AI-Readiness Score of an IFC model based on the five pillars of the conceptual framework (Structural Consistency, Semantic Completeness, Geometric Fidelity, Temporal Coherence, and Contextual Richness). This would allow model authors to validate their exports before handover to an AI service.

2. Automated Semantic Enrichment and Repair: Research into machine learning and natural language processing (NLP) models, specifically using large language models (LLMs) to automatically infer and enrich missing or ambiguous semantic properties in IFC

files, represents a high-impact area. Furthermore, the development of robust geometric repair algorithms that can automatically correct non-manifold geometry and small overlaps is essential for production-level AI deployment.

3. Standardization of Ontologies: The focus should shift from standardizing the file format (IFC) to standardizing the ontology used for computational interpretation. This involves creating and maintaining open-source, domain-specific AEC knowledge graphs that can universally interpret MAIDR Psets, serving as the essential semantic middleware for truly scalable AI-BIM platforms.

5. CONCLUSION

This systematic review has critically assessed the state of BIM and IFC data readiness for integration with Artificial Intelligence in the AEC industry. The findings confirm the immense potential of AI applications across the design, construction management, and operations life cycles, yet simultaneously reveal a persistent and fundamental challenge: the semantic and topological gap between the IFC data standard and the strict requirements of advanced computational models.

IFC, as a data exchange format, is necessary but insufficient for fully automated AI tasks. Current AI success stories are often predicated on labor-intensive data pre-processing, conversion to graph-based structures, and the utilization of custom, non-standard property sets, which collectively undermine the principle of universal interoperability. Specifically, the findings highlight critical limitations in the reliable representation of geometric fidelity, the consistent definition of spatial boundaries for analysis, and the inclusion of explicit semantic attributes necessary for machine inference.

The path toward scalable, industry-wide AI deployment in AEC necessitates a two-pronged strategic evolution: first, a shift in industry practice towards mandatory adherence to standardized Minimum AI Data Requirements (MAIDR) during the BIM authoring phase; and second, a focus in research on developing semantic middleware, such as automated Knowledge Graph generators, and robust geometric validation and repair tools to ensure the five pillars of AI-Ready data—Structural Consistency, Semantic Completeness, Geometric Fidelity, Temporal Coherence, and Contextual Richness—are met before data ingestion. Addressing these fundamental data quality and semantic challenges is the crucial prerequisite for the successful industrialization of AI in the built environment.

REFERENCES

1. De Souza, M.P.; Fialho, B.C.; Ferreira, R.C.; Fabricio, M.M.; Codinhoto, R. Modelling and Coordination of Building Design: An Experience of BIM Learning/Upskilling. Archit. Eng. Des. Manag. 2023, 19, 74–91.

2. NAM AI in Manufacturing. Available online: https://nam.org/issues/research-innovation-and-technology/ai/ (accessed on 5 December 2023).

3. Pan, Y.; Zhang, L. Integrating BIM and AI for Smart Construction Management: Current Status and Future Directions. Arch. Comput. Methods Eng. 2023, 30, 1081–1110.

4. Allen, G. Understanding AI Technology; Joint Artificial Intelligence Center (JAIC) The Pentagon United States: Arlington, VA, USA,2020.

5. Tomczak, A.; Berlo, L.V.; Krijnen, T.; Borrmann, A.; Bolpagni, M. A Review of Methods to Specify Information Requirements in Digital Construction Projects. IOP Conf. Ser. Earth Environ. Sci. 2021, 1101, 092024.

6. Bloch, T. Connecting Research on Semantic Enrichment of BIM-Review of Approaches, Methods and Possible Applications. J. Inf.Technol. Constr. 2022, 27, 416–440.

7. Noardo, F.; Harrie, L.; Arroyo Ohori, K.; Biljecki, F.; Ellul, C.; Krijnen, T.; Eriksson, H.; Guler, D.; Hintz, D.; Jadidi, M.A.; et al.Tools for BIM-GIS Integration (IFC Georeferencing and Conversions): Results from the GeoBIM Benchmark 2019. ISPRS Int. J.Geo-Inf. 2020, 9, 502.

8. McCarthy, J.; Minsky, M.L.; Rochester, N.; Shannon, C.E. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. AI Mag. 2006, 27, 12.

9. Turing, A.M. Computing Machinery and Intelligence; Springer: Berlin/Heidelberg, Germany, 2009.

10. McCulloch, W.S.; Pitts, W. A Logical Calculus of the Ideas Immanent in Nervous Activity. Bull. Math. Biophys. 1943, 5, 115–133.

11. Jackson, P. Introduction to Expert Systems; Addison-Wesley Longman Publishing Co., Inc.: Petaluma, CA, USA, 1990.

12. Waterman, D.A. A Guide to Expert Systems; Addison-Wesley Longman Publishing Co., Inc.:

Petaluma, CA, USA, 1985.

13. Yoon, Y.; Guimaraes, T.; O'Neal, Q. Exploring the Factors Associated with Expert Systems Success. MIS Q. 1995, 19, 83–106.

14. Sagiroglu, S.; Sinanc, D. Big Data: A Review. In Proceedings of the 2013 International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, USA, 20–24 May 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 42–47.

15. Parate, H., Kishore Bandela, & Paniteja Madala. (2025). Quantity Take-Off Strategies: Reducing Errors in Roadway Construction Estimation. Journal of Mechanical, Civil and Industrial Engineering, 6(3), 01-09. https://doi.org/10.32996/jmcie.2025.6.3.1

16. Jordan, M.I.; Mitchell, T.M. Machine Learning: Trends, Perspectives, and Prospects. Science 2015, 349, 255–260.

17. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. Nature 2015, 521, 436–444.

18. Chang, H.S.; Fu, M.C.; Hu, J.; Marcus, S.I. Google DeepMind's AlphaGo: Operations Research's Unheralded Role in the Path-Breaking Achievement. OrMs Today 2016, 43, 24–30.

19. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Li, F. Imagenet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; IEEE: Piscataway, NJ,USA, 2009; pp. 248–255.

20. Vaswani, A. Attention Is All You Need. In Advances in Neural Information Processing Systems; MIT Press: Cambridge, MA, USA,2017.

21. Radford, A. Improving Language Understanding by Generative Pre-Training. 2018. Available online: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf (accessed on 5 December 2023).

22. Brown, T.B. Language Models Are Few-Shot Learners. arXiv 2020, arXiv:200514165.

23. Bubeck, S.; Chandrasekaran, V.; Eldan, R.; Gehrke, J.; Horvitz, E.; Kamar, E.; Lee, P.; Lee, Y.T.; Li, Y.; Lundberg, S.; et al. Sparks of Artificial General Intelligence: Early Experiments with Gpt-4. arXiv 2023, arXiv:230312712.

24. International Organization for Standardization ISO/IEC 22989:2022 Information Technology—Artificial Intelligence—Artificial Intelligence Concepts and Terminology 2022. Available online: https://www.iso.org/standard/74296.html (accessed on 5 December 2023).

25. Standards Australia Standards Australia. Available online: https://www.standards.org.au/standards-catalogue/standarddetails?designation=as-iso-iec-22989-2023 (accessed on 5 December 2023).

26. El Naqa, I.; Murphy, M.J. What Is Machine Learning? Springer: Berlin/Heidelberg, Germany, 2015.

27. Gurney, K. An Introduction to Neural Networks; CRC Press: Boca Raton, FL, USA, 2018.

28. Calder, B. Architecture: From Prehistory to Climate Emergency; Penguin: London, UK, 2021.

29. García, M.Á. Challenges of the Construction Sector in the Global Economy and the Knowledge Society. Int. J. Strateg. Prop. Manag.2005, 9, 65–77.

30. Young, D.; Panthi, K.; Noor, O. Challenges Involved in Adopting BIM on the Construction Jobsite. EPiC Ser. Built Environ. 2021, 2,302–310.

31. Buildings 2024, 14, 3305 50 of 54 Mahmoodzadeh, A.; Nejati, H.R.; Mohammadi, M. Optimized Machine Learning Modelling for Predicting the Construction Costand Duration of Tunnelling Projects. Autom. Constr. 2022, 139, 104305.

32. Hinze, J.; Appelgate, L.L. Costs of Construction Injuries. J. Constr. Eng. Manag. 1991, 117, 537–550.

33. You, H.; Ye, Y.; Zhou, T.; Zhu, Q.; Du, J. Robot-Enabled Construction Assembly with Automated Sequence Planning Based on ChatGPT: RoboGPT. Buildings 2023, 13, 1772.

34. Choi, W.; Na, S.; Heo, S. Integrating Drone Imagery and AI for Improved Construction Site Management through Building Information Modeling. Buildings 2024, 14, 1106.

35. Based on Deep Learning Network with Improved Particle Swarm Optimization. Artif. Intell. Evol. 2023, 4, 216–225.

36. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.;Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. BMJ 2021, 372, 71.

37. Han, J.; Zhou, X.; Zhang, W.; Guo, Q.; Wang, J.; Lu, Y. Directed Representative Graph Modeling of MEP Systems Using BIM Data.Buildings 2022, 12, 834.

38. Zhao, Y.; Deng, X.; Lai, H. Reconstructing BIM from 2D Structural Drawings for Existing Buildings. Autom. Constr. 2021, 128,103750.

39. Vinod Kumar Enugala. (2025). "BIM-to-Field" Inspection Workflows for Zero Paper Sites. Utilitas Mathematica, 122(2), 372–404. Retrieved from https://utilitasmathematica.com/index.php/Index/article/view/2711

40. Rausch, C.; Talebi, S.; Poshdar, M.; Li, B.; Schultz, C. Tolerance Management Domain Model for Semantic Enrichment of BIMs.

41. Jia, J.; Gao, J.; Wang, W.; Ma, L.; Li, J.; Zhang, Z. An Automatic Generation Method of Finite Element Model Based on BIM and Ontology. Buildings 2022, 12, 1949.

42. Chen, Z.; Chen, K.; Song, C.; Zhang, X.; Cheng, J.C.P.; Li, D. Global Path Planning Based on BIM and Physics Engine for UGVs in Indoor Environments. Autom. Constr. 2022, 139, 104263.

43. Hamieh, A.; Ben Makhlouf, A.; Louhichi, B.; Deneux, D. A BIM-Based Method to Plan Indoor Paths. Autom. Constr. 2020, 113,103120.

44. Feng, Y.; Wang, J.; Fan, H.; Hu, Y. A BIM-Based Coordination Support System for Emergency Response. IEEE Access 2021, 9,68814–68825.