Scilight

Journal of Emerging Technologies with Industrial Applications
https://www.sciltp.com/journals/jetia



Review

Check for updates

Industrial Information Integration-Selected Patents and Innovations

Jinzhi Li [1,*,†] and Liuliu Fu [2,†]

[1] Department of Information Technology & Decision Sciences, Strome College of Business, Old Dominion University, Norfolk, VA 23529, USA
[2] Department of Information Systems, California State University-Los Angeles, Los Angeles, CA 90032, USA
* Correspondence: jli037@odu.edu
† These authors contributed equally to this work.

How To Cite: Li, J.; Fu, L. Industrial Information Integration-Selected Patents and Innovations. *Journal of Emerging Technologies with Industrial Applications* **2026**, *1*(2), 11. https://doi.org/10.53941/jetia.2026.100011

Received: 4 August 2026
Revised: 10 October 2026
Accepted: 24 November 2026
Published: 26 November 2026

Abstract: Industrial information integration has been mapped repeatedly on its academic side, through a series of literature reviews spanning two decades of research. Its patent record, what firms build, protect, and commercialize, has not been mapped at all. This paper provides the first systematic survey of that record. From the Lens.org database we retrieve 6627 unique patent families that use the phrase "information integration" and were published between January 2006 and July 2026 across the United States, European, German, and Chinese patent offices, screen them to an in-scope pool of 3312 families, and classify the pool along two separately coded dimensions: the integration mechanism, in seven classes (five mapped to layers of IEC 62264 and RAMI 4.0 and two cross-cutting capability classes), and the application domain, in nine classes aligned with the research-category framework of the parent review series. From this pool, 150 families are selected by a composite representativeness score and manually coded under a written codebook, yielding a descriptive corpus of 116 representative patent families that are introduced individually by domain along the industrial value chain. Annual patenting volume grows roughly threefold across the window when families are dated by their first publication, accelerating after 2015. The two dominant patenting systems differ structurally rather than only in volume: in the retrieved corpus, families published through the US office are concentrated among large vendors, with the ten largest applicant groups holding 26.5% of families, while families published through the Chinese office are more fragmented, with the ten largest holding 10.3% and led by a utility rather than a technology vendor. Patenting concentrates in the data and model classes rather than at the human–machine interface, and about one third of in-scope families name no application domain in their title or abstract.

Keywords: industrial information integration; patent analysis; technology landscape; classification framework; industrial information integration engineering; enterprise integration

1. Introduction

The digital transformation of industry rests on a deceptively simple capability: making information flow across the boundaries of systems that were never designed to share it. Industrial information integration, the discipline concerned with linking the information systems of enterprises, plants, supply chains, and infrastructures into coherent wholes, has emerged as the connective layer of this transformation. Xu's systems-perspective

Copyright: © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).
Publisher's Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations

treatment established the field's intellectual foundation, framing enterprise integration and information architecture as a unified engineering problem spanning data, process, and application layers [1], and the field has since consolidated as industrial information integration engineering (IIIE), an emerging subject at the junction of industrialization and informatization [2], with a dedicated journal, a growing methods repertoire, and a widening application footprint. The programs that define contemporary industrial policy, Industrie 4.0 and its successors [3], industrial Internet initiatives, and national smart-manufacturing and grid-modernization strategies, are, at their technical core, information-integration programs: the cyber–physical systems, digital twins, and industrial IoT platforms that characterize Industry 4.0 [4,5] all presuppose that heterogeneous information can be modeled, moved, mediated, secured, and presented across organizational and technical boundaries.

The field's recent development illustrates both the depth of its technical strands and the systems-science orientation it inherited from its foundation [6]. On the enabling-technology side, research has advanced the full stack through which industrial information is integrated: Internet-of-Things infrastructures and their resource allocation [7,8]; edge- and cloud-based application deployment and service computing [9,10], including cooperative task allocation over crowds and networks of services [11–13] and the embedding and evolution of the networks themselves [14,15]; blockchain as a trust layer for industrial data exchange, surveyed generally [16] and embedded into IoT security [17], service-based applications and validator selection [18,19], business processes [20], maritime and shipping [21], and institutional information management [22]; and artificial intelligence as an integrative capability, from system-level frameworks of AI models [23] and evolutionary computation at scale [24] to multi-agent logistics scheduling [25], federated learning in mobility systems [26], and financial applications [27,28]. Security runs through this stack as a persistent concern, from computational-intelligence-enabled IoT cybersecurity [29] to the behavioral antecedents of cybersecurity protection [30].

The application footprint of this research is equally broad. Within manufacturing, recent work spans the smart factory [31], big-data management maturity in intelligent manufacturing enterprises [32], the transformation of national manufacturing bases [33], and safety management under Industry 4.0 [34]. Beyond the factory, industrial information integration reaches healthcare information systems and consumer wearables [35–38], ultralow-latency financial data infrastructure [39], telecommunications [40], aerospace systems [41], and requirements engineering for complex systems [42]; and the field's frontier is turning human-centric, with Industry 5.0 research agendas [43] and workforce reskilling for Industry 4.0 and beyond [44,45] placing people, rather than only systems, at the center of integration. This breadth is the field's strength, and it is precisely what makes systematic mapping of the field both necessary and difficult.

The academic side of this field has indeed been mapped with unusual thoroughness, mapping itself has become one of the field's scholarly traditions. Chen reviewed the industrial information integration literature of 2006–2015 [46]; Lu reviewed the adjacent industrial-integration literature of 2006–2016 [47]; Chen extended the mapping through 2016–2019 in a survey of monographic scale [48]; and Li has recently carried the series into the present decade in a two-part review organizing 874 papers of 2020–2025 into thirty-four research categories [49,50]. Taken together, these reviews give the field a longitudinal self-portrait few engineering disciplines possess: its themes, methods, sectors, and trajectories are documented at regular intervals across twenty years.

Yet this self-portrait has a systematic blind spot. Literature reviews map what researchers' study; they do not map what firms build, protect, and commercialize. The patent record is the natural complement, patents are practice-facing documents, substantively examined, tied to appropriable technology, and filed by exactly the actors (vendors, utilities, manufacturers, integrators) whose adoption decisions determine whether integration research matters in practice. There is no reason to assume the two records coincide: themes prominent in the literature may be thin in the patent record where the underlying capability resists appropriation or is commercialized as service rather than invention, and dense patent streams may form around deployment programs that generate little academic writing. Whether industrial information integration's patented technology mirrors, lags, or diverges from its academic map is, at present, unknown, because no systematic survey of the field's patent record exists. This paper provides one. We ask: what does the patent record of industrial information integration look like, how has it grown, who patents where, through which integration mechanisms, and in which application domains?

To answer this question, we retrieve 9956 patent records (6627 unique simple families) published between January 2006 and July 2026 across the United States, European, German, and Chinese patent offices, screen them to an in-scope pool of 3312 families, and classify the pool along two separately coded dimensions, integration mechanism (seven classes related to the ISA-95/RAMI 4.0 architecture layers) and application domain (nine classes aligned with the in-scope portion of the review series' category framework [46,49,50]). From the pool we select 150 families by a composite representativeness score, verify their classification by manual coding and adjudicated review under a written codebook, and descriptively introduce the resulting 116-family representative corpus. The record shows a field patenting at scale and accelerating annual output roughly triples across the

window, with distinct growth after 2015. It shows a sharply asymmetric geography: United States patenting is vendor-concentrated, the ten largest applicants applicant groups hold 26.5% of US-office families, and productizes integration as reusable tooling, while Chinese patenting is fragmented, the ten largest applicant groups hold about 10%, and is driven by utilities, universities, and small enterprises patenting within infrastructure programs. It shows patenting concentrated at the data and model layers rather than the presentation layer, a substantial domain-silent stratum (one third of the in-scope pool names no application domain in title or abstract), and, within the selected set, sparsely populated mechanism–domain combinations that pose questions for pool-level study.

The paper makes three contributions. First, it provides the first systematic patent-side map of industrial information integration, extending the field's review tradition [46–50] from the literature to the patent record and enabling the two records to be compared. Second, it contributes a reusable instrument: a two-dimensional mechanism × domain classification framework with a written codebook, a documented coding and review procedure, and a reproducible retrieval-and-selection pipeline over an openly accessible database, designed, like the review series itself, to serve as a fixed baseline for periodic updates. Third, its empirical findings are agenda-setting for the field: the asymmetry between vendor-concentrated and deployment-driven patenting systems, the below-the-presentation-layer concentration of patented integration technology, and the divergences between academic emphasis and patent density identified in the concluding section supply concrete, testable hypotheses about where industrial information integration's research and its industrial appropriation part ways.

The remainder of the paper is organized as follows. Section 2 describes the data source, search strategy, screening, classification framework, selection procedure, and coding reliability. Section 3 presents the patent landscape: a corpus overview (Section 3.1) followed by descriptive introductions of the 116 representative patent families organized by application domain along the industrial value chain (Sections 3.2–3.10). Section 4 concludes with the survey's findings, limitations, and the research directions they open.

2. Data and Methodology

2.1. Data Source and Search Strategy

Patent data were retrieved from Lens.org, a free, openly accessible patent database aggregating bibliographic records from more than one hundred patent-issuing authorities. Lens.org exports include title, abstract, applicants, CPC/IPC classifications, publication year, jurisdiction, simple patent family size, forward citation counts, legal status, and a persistent record URL, enabling the two-pass, abstract-based screening described below. The use of Lens.org for patent-survey research follows established precedent: the database has been characterized as an open resource for mapping the influence of research on innovation [51], and it has served as the retrieval base for published patent landscape studies in other technical fields [52].

The search was conducted on the exact phrase "information integration" using Lens.org's default query field, which matches the phrase anywhere in the patent document (title, abstract, claims, and description text where full text is indexed), restricted to patents published from 1 January 2006 to the retrieval date in July 2026 (the latest retrieved publication is dated 9 July 2026) and to the jurisdictions of the United States (US), the European Patent Office (EP), Germany (DE), and China (CN). The window opens in 2006 to coincide with the period covered by the literature-review series this survey extends [46,49,50], and the jurisdictions were chosen to capture the two largest industrial-patenting systems (US, CN) together with the European route most relevant to industrial-automation patenting (EP, DE). The retrieval (July 2026) returned 9956 publication-level records. The export was not collapsed to one document per simple family, so several publications of one invention (an application and its later grant, or counterparts published by more than one office) appear as separate records. Records were therefore grouped by the simple-family membership field of the Lens.org export, giving 6627 unique simple families, of which 4998 are represented by a single retrieved record and 1629 by two or more. One representative publication was retained per family by a fixed rule: a US publication if the family has one among the retrieved records, otherwise an EP, then a DE, then a CN publication, taking the most recent publication within the chosen office. The rule favours an English-language document where one exists. The representative publication supplies the title, abstract, jurisdiction, and publication year used in all subsequent steps; family size and forward-citation counts are family-level fields. A family filed in more than one of the four offices is thus counted once, under the office of its representative.

The single-phrase query favours precision and reproducibility over recall. "Information integration" is the field's self-designation: it names the discipline [2] and the parent review series [46,48–50], and it is the standard English rendering of the Chinese term 信息集成 in translated abstracts, so that one phrase addresses the two largest patenting systems on the same footing. Neighbouring terms such as "data integration" or "interoperability" are general information-technology vocabulary and would shift the burden of delimiting the field onto the relevance screen. The consequence is that the survey maps patents that describe themselves in the vocabulary of information integration.

Integration inventions described only in other terms (for example, "OT–IT integration" or "data interoperability") are not retrieved, and no recall estimate against a broader query was made. This is a deliberate boundary of the survey and not an estimate of all integration-related patenting: a multi-term query would define a different and broader object, whose limits would rest on the relevance screen instead of on the field's own terminology.

2.2. Relevance Screening

Because the phrase "information integration" also occurs in domains outside the scope of industrial information integration, the 6627 families were screened for relevance using a keyword-based classifier applied to titles and abstracts. The screen combines term lists matched in the lower-cased title and abstract with CPC and IPC class prefixes, applied in a fixed order; the complete rules are given in Supplementary Material B, and the screen and classifier were re-run from those rules on the archived export to confirm every count reported here. Mechanism and domain classification used word-boundary regular expressions; the relevance screen used substring matching. An audit during development showed why this matters: short tokens in the keyword list of the infrastructure and transportation domain ("port", "city", "ship") matched as substrings inside common words such as "support", "capacity", and "relationship", inflating that category to 1296 candidate records; word-boundary matching corrected the count to 728.

Records were excluded when their claimed subject matter fell in verticals outside the survey's scope, medical and health, media and entertainment, consumer applications, agriculture, finance, military, and environmental or geological monitoring, or when the record claimed only standalone single-platform sensor fusion (for example, on-vehicle perception for autonomous driving) rather than integration of information across systems. These boundaries were drawn on three grounds. Substantively, the survey covers industrial information integration, the manufacturing, enterprise, energy, and infrastructure core of the field, mirroring the in-scope portion of the category framework used by the academic review series, whose sector verticals outside that core are likewise not carried into this survey [49,50]. Semantically, in the excluded fields the phrase "information integration" predominantly denotes technically distinct problems with survey traditions of their own, image and signal fusion in medical imaging, multi-sensor target tracking in defense systems, media-asset workflow orchestration, market-data aggregation in finance, so their inclusion would dilute the catalog with material that is not industrial systems integration. Practically, defense-related patenting is additionally subject to secrecy-driven incompleteness that would undermine representative coverage of that vertical. Screening produced an in-scope pool of 3312 families. In detail, 1634 of the 6627 families were excluded as out-of-scope verticals or standalone fusion (medical and health 855; media, education and entertainment 326; image or sensor fusion 295; finance 140; military 18), 1594 showed no affirmative industrial or integration signal and were not retained, 69 energy, transport, and construction families were dropped for consumer or agricultural terms, and 18 were removed as consumer-device noise. Environmental and geological monitoring was excluded at manual coding, not by the automated screen.

2.3. Classification Framework

Each patent family is classified along two separately coded dimensions (Table 1). The first dimension, the integration mechanism, captures the technical means by which information is integrated, in seven classes defined with reference to the enterprise–control system integration standard IEC 62264 (ISA-95) [53] and the Reference Architecture Model Industrie 4.0 (RAMI 4.0) [54]: five classes correspond to layers or elements of these models, and two are cross-cutting capability classes, as explained below. The second dimension, the application domain, comprises nine classes covering the industrial core of the field. Table 1 defines both dimensions.

The seven mechanism classes are not seven layers. Five correspond to layers or elements of the two reference models: connectivity to the asset, integration, and communication layers of RAMI 4.0 and to acquisition at Levels 0–2 of IEC 62264; data models and databases to the information layer and to the object models and Level 3–4 information exchanges of IEC 62264; middleware and services to the communication and functional layers and to the transaction and messaging parts of IEC 62264; and human–machine interfaces to the human-facing end of the functional and business layers. AI and analytics and cybersecurity are cross-cutting capabilities with no dedicated layer in either model: RAMI 4.0 treats security as a property spanning its layers, and industrial security is standardized separately in IEC 62443 [55]. They are retained as classes because a substantial body of integration patents state their contribution in these terms and because the same themes recur as sub-groups throughout the parent reviews. Since each family receives exactly one mechanism code, the codebook assigns a cross-cutting class only when the learned model or the security purpose is the claimed contribution (rules R-C3 and R-C5 in Supplementary Material A); counts for these two classes are therefore lower bounds on how often the capabilities appear. The full correspondence is tabulated in Supplementary Material A.

Table 1. The two-dimensional classification framework: integration mechanisms and application domains.

Category	Scope
Integration mechanism (seven classes)	
Data models, semantic interoperability & digital twins	Declared representations that serve as the integrating structure: ontologies, schemas, metadata standards, exchange formats, information models, digital twins and model-based engineering artifacts.
Databases, fusion & synchronization	Movement and storage of data: ETL, data warehousing and lakes, multi-source fusion, synchronization and replication, federated query over heterogeneous stores.
Middleware, APIs & services	Interface and brokering layers: middleware, service buses, API gateways, web services and microservices, protocol adapters, integration platforms whose novelty lies in mediation.
Industrial IoT, edge, cloud & connectivity	Acquisition and transport of data: sensor networks, RFID, industrial wireless and 5G, edge and cloud infrastructures, remote transmission and monitoring architectures.
AI & analytics	Learned or inferential models applied to integrated data: machine learning, data mining, predictive analytics, learned knowledge graphs, intelligent decision support. Cross-cutting capability class.
Cybersecurity & trusted exchange	Mechanisms defined by security purpose: encryption, authentication, access control, blockchain-based trust, privacy preservation, tamper-proofing of industrial data flows. Cross-cutting capability class.
Human–machine interfaces	Presentation layers where integration serves human consumption: visualization, dashboards, AR/VR interfaces, interactive displays.
Application domain (nine classes)	
Domain-general/cross-cutting	Generic integration methods claimed without commitment to an industry or business function.
Product design & engineering	Claims a design or engineering artifact: CAD models, PLM/PDM records, engineering BOM, design collaboration, requirements traceability, digital thread.
Production & shop floor	Manufacturing execution, workshop operations, assembly, machining, process control in a production context.
Quality & inspection	Quality management, inspection, defect detection, metrology, traceability for quality purposes.
Maintenance & asset management	Condition monitoring, fault diagnosis, predictive maintenance, equipment and asset health management.
Logistics, supply chain & warehousing	Supply chain coordination, logistics, warehousing, inventory, fleet and freight information systems.
Enterprise & cross-functional	ERP, office automation, business process management, workflow, CRM and management information systems.
Energy & utilities	Power systems, smart grid, utility operations, energy management information integration.
Infrastructure & transportation	Construction and project/asset-level BIM, smart-city infrastructure, transport information systems, mining operations.

The domain dimension is deliberately hybrid. Six classes are value-chain functions (design, production, quality, maintenance, logistics, enterprise) and identify the business function whose information is integrated, in whatever industry. Two are sector classes (energy and utilities; infrastructure and transportation), needed because in these sectors the object of integration is the sector's own operating system (the grid and its dispatch, metering, and distribution operations; the built asset, the construction project, the traffic or rail network), which has no home among the factory functions. The domain-general class receives families that claim neither. Consistency is maintained by a precedence rule (R-D5 in Supplementary Material A): a claimed value-chain function is coded as that function even in an energy or infrastructure setting, and a sector class is assigned only when the claimed contribution is the operation of the sector's own system.

The framework is designed for continuity with the classification lineage of the academic review series. The application-domain categories align nominally with the in-scope portion of the categories used by Chen and by Li [48–50], restricted to the enterprise, manufacturing, and infrastructure core and excluding the sector verticals outside this survey's scope. The domain-general class follows the precedent of the review series itself, whose framework relies on three domain-general categories (Algorithms, Mathematical Modeling, and ICT) jointly holding roughly one quarter of the reviewed literature; a legitimate destination for domain-silent patents proved equally necessary here. The mechanism dimension generalizes a pattern internal to the same series: near-identical mechanism sub-groups (ontology and semantic integration; architecture, frameworks and middleware; data integration; IIoT and edge; AI and analytics; security) recur inside almost every one of the series' domain categories, which motivates factoring mechanism out as an orthogonal dimension rather than re-deriving it within each domain.

For pool-level statistics, mechanism and domain labels were assigned by the word-boundary keyword classifier. Each family receives the first class whose term list matches, with classes tested in a fixed order

(Supplementary Material B). Of the 3312 in-scope families, 1294 matched both a mechanism and a domain term; 650 matched a mechanism term only, 871 a domain term only, and 497 neither. The 2018 families without two labels remain in all pool-level statistics that do not depend on classification (annual volume, jurisdiction, and applicants; Section 3.1), but they were not eligible for selection and were not manually coded. A missing label reflects the term lists and the wording of the abstract, not the absence of a mechanism or a domain; in particular, the 1147 families with no domain term are domain-silent in title and abstract, which is an upper bound on genuinely domain-general patenting. As an unsupervised cross-check, latent Dirichlet allocation (LDA) topic models (k = 8 and k = 12) [56] were fitted to the 1294 eligible families, following the established use of topic models for validating and complementing patent-text classification [57,58] and, specifically, for inducing technology topics from patent abstracts in patent-portfolio surveys [59]. The induced topics corroborate the scheme's major themes but conflate its two dimensions: some topics track integration mechanisms (services and platforms; databases and queries; prediction and analytics), while others track application domains (power systems; building and construction) or patent drafting conventions (Figure 1). A single induced classification would therefore mix mechanism and domain in one axis, which supports the use of the expert-defined two-dimensional scheme over an induced one.

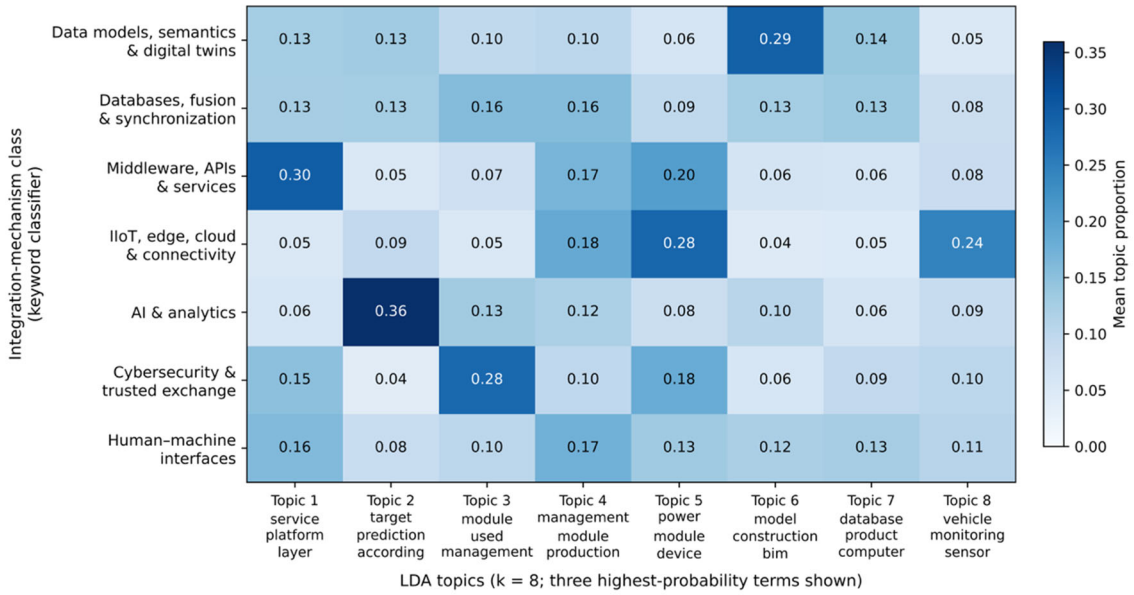


Figure 1. Unsupervised cross-check of the classification scheme: mean LDA topic proportions (k = 8) by keyword-assigned integration-mechanism class. Topics align partly with mechanisms and partly with application domains, supporting the two-dimensional design.

2.4. Selection of Representative Patents

Following the descriptive-survey genre, in which a curated body of recent patents is introduced and thematically organized rather than statistically modelled [59–61], the paper introduces a curated set of representative patents rather than a statistical sample. Eligibility for selection required classifiability on both dimensions, satisfied by 1294 families. Each eligible family received a composite score, the weighted sum of four indicators that each lie between 0 and 1 and are computed over the eligible pool. (i) Forward citations (weight 0.40): the percentile rank of the family's forward-citation count among families of the same jurisdiction and three-year publication cohort, with tied counts given their average rank; a missing count is treated as zero. The within-cohort percentile removes the age and office effects that make raw counts incomparable. (ii) Family size (0.20): the natural logarithm of one plus the simple family size, rescaled min–max over the eligible pool; a missing size is treated as one. Simple family size counts the publications in the family, so it reflects both later publications of the same application (for example, a grant following an application) and filings in other offices. (iii) Legal status (0.20): a graded score from the Lens.org legal-status field of the representative publication: active 1.0, expired 0.8, pending 0.5, inactive 0.3, discontinued 0. The term rewards rights that are or were kept in force. It is not a grant indicator, because Chinese family representatives are application publications (kind code A) from which grant cannot be read. (iv) Typicality (0.20): the cosine similarity between the family's TF-IDF vector and the centroid of its keyword-assigned mechanism class. For this indicator, titles and abstracts were concatenated and lower-

cased; terms were TF-IDF weighted after removing English stop words and terms occurring in fewer than five or in more than half of the eligible families, retaining the 6000 most frequent; vectors were scaled to unit length; and each centroid is the mean vector of the eligible families in that mechanism class, rescaled to unit length. The weights are judgement-based rather than estimated: citations receive the largest weight as the most established single indicator of technical importance, and the remaining three share the balance equally as proxies for applicant-assigned value, substantive examination, and closeness to the centre of the class. Effective influence differs from nominal weight: the status term has the widest spread across the pool, so it and the citation term drive the ranking, while typicality, which ranges only from 0.05 to 0.55, has little influence. The 150 highest-scoring families were carried into manual coding. The selection is stable when the four terms are reweighted: 136 of the 150 families remain selected under equal weights, 139 with the citation weight raised to 0.50, and 145 with typicality removed. It is not stable to removing the status term, under which only 82 remain, which shows how strongly the rule favours rights in force. Two properties of this rule bear on everything that follows. It favours cited, in-force families with several publications, and it draws only from families with both keyword labels, so families that are domain-silent in title and abstract could not be selected. The selected set therefore describes notable, well-characterized families; its composition is not an estimate of the composition of the pool.

2.5. Manual Coding, Reliability, and the Final Corpus

The machine-assigned labels of the top 150 set were verified by manual coding in two rounds. In a first, blind round conducted under the initial category definitions and covering the 113 records then shortlisted, the first author and an independent model-assisted pass each assigned mechanism and domain labels from titles and abstracts; agreement was 54% for mechanism (Cohen's $\kappa = 0.43$) and 50% for domain ($\kappa = 0.41$) [62], "moderate" agreement against conventional benchmarks [63]. Consistent with the methodological literature on structured content analysis, agreement at this level indicates underspecified category definitions rather than coder error [64,65], and the disagreement pattern (concentrated in one domain category functioning as a residual bucket, and on the boundary between the data-model and database classes, representation versus movement of data) was used to develop a written codebook. The codebook gives each category a positive definition with inclusion criteria, five explicit boundary decision rules for the mechanism dimension, five routing rules for the domain dimension including the rule routing domain-silent records to the domain-general class, and four relevance-exclusion rules; it is reproduced in full as Supplementary Material A.

In the second round, all 150 records were first coded against the codebook from titles and abstracts, blind to all first-round labels. This first pass was produced with a large language model (Claude, Anthropic) applying the written codebook; recent methodological evidence supports the use of large language models as text annotators when they operate under explicit coding instructions and human review [66,67]. The second author then reviewed every record with the first-pass codes visible, checking each against the codebook and, where necessary, the claims. The authors bear full responsibility for the final classification. The review left the include/exclude verdict unchanged on all 150 records, changed the mechanism code on 4 records (146 of 150 retained, 97.3%), and changed the domain code on 1 record (149 of 150, 99.3%); all five changes fell within boundary zones the codebook had flagged. These figures are retention rates under review, not inter-rater reliability: a reviewer who sees a proposed code is more likely to accept it than to reach it unaided, so the figures overstate the agreement two independent coders would achieve. They are not comparable with the first-round statistics, which were moreover obtained under a different category scheme. Codebook development from first-round disagreements, followed by recoding and adjudicated review, follows established practice in content analysis [64,65]. An independent inter-rater statistic for the final codebook is therefore not available; obtaining one is left to future applications of the instrument.

Manual coding excluded 33 of the 150 records under the codebook's relevance rules (19 single-platform autonomy or perception records; 8 out-of-scope verticals; 6 records claiming no integration mechanism), leaving 117 families. One further family, CN 118552523 A, was removed at revision. It claims an improved defect-detection network and uses the phrase "information integration" only for a component of that network, so it integrates no information across sources. It had been admitted under a rule that accepted factory machine vision by its industrial setting alone; that rule (R-X2 in Supplementary Material A) now requires integration of information from more than one source or system, and the other 22 vision, inspection, and robotics records among the 117 were re-examined against it and retained. The energy and infrastructure entries were also re-checked at revision against the function-over-sector rule (R-D5), and two families whose claimed contribution is the management or condition monitoring of equipment and structures (CN 104167817 A [68] and CN 120948614 A [69]) were recoded to maintenance and asset management. The final corpus is 116 patent families, whose distribution across the mechanism × domain matrix is shown in Table 2. Table 2 describes the selected families only; because of the

selection rule (Section 2.4), neither its row and column shares nor its empty cells estimate the distribution of the pool. Manual coding also measures the keyword classifier. For the 116 retained families, the keyword label matched the final code on mechanism for 61 (53%), on domain for 57 (49%), and on both for 31 (27%); the keyword scheme had no domain-general class, which accounts for 18 of the domain mismatches. Together with the 34 exclusions (23% of the selected records that the relevance screen had admitted), this indicates that pool-level counts derived from keyword labels are indicative only. These 116 families are the subject of the descriptive catalog in Section 3.

Table 2. Distribution of the 116 representative patent families across integration mechanisms (rows) and application domains (columns). Column headers abbreviate the domain names defined in Table 1. The table describes the selected set and is not an estimate of the pool distribution.

Integration mechanism\Application domain	General	Design	Production	Quality	Maintenance	Logistics	Enterprise	Energy	Infrastructure	Total
Data models, semantics & twins	–	4	4	–	2	–	2	3	10	25
Databases, fusion & sync.	7	2	2	1	3	–	4	6	8	33
Middleware, APIs & services	1	–	2	–	1	1	1	3	–	9
IIoT, edge, cloud & connectivity	3	–	4	1	2	1	–	2	5	18
AI & analytics	5	1	1	1	2	1	4	4	2	21
Cybersecurity & trusted exchange	2	–	2	–	–	1	2	–	–	7
Human–machine interfaces	–	–	1	–	1	–	–	–	1	3
Total	18	7	16	3	11	4	13	18	26	116

2.6. Methodological Limitations

Five limitations bound the method. First, pool-level statistics (Section 3.1) rest on the keyword classifier, which carries residual misassignment that manual coding corrected only within the top 150 set; pool-level figures should be read as approximate (keyword labels matched final manual codes for about half of the selected families; Section 2.5). Second, manual coding was performed on titles and abstracts, consulting claims only for adjudication; Chinese-language abstracts were machine-translated, and translation ambiguity (for example, whether a "vehicle" system is roadside or on-board) is the most likely locus of residual coding error. Third, retrieval and coverage are uneven across offices: Lens.org's full-text indexing is deeper for US documents than for other jurisdictions, so description-text matches contribute proportionally more US retrievals; EP-designated families are nearly absent from the corpus relative to US and CN families; and legal-status labels for EP and CN records are less reliable than for US records, so grant-status-dependent findings rest primarily on the US subset; applicant names are used as standardized by Lens.org, and uneven standardization across offices may affect measured applicant concentration. Fourth, the selection rule's legal-status and citation terms deliberately over-represent in-force, cited families relative to the pool: 108 of the 116 selected families have active status, against 30% of the eligible pool. The 83 Chinese representatives among the 116 are all application publications (kind code A), for which grant cannot be read from the representative document. The corpus is a curated set of notable families, not a set of granted patents and not a probability sample, and no inferential statistics are computed from it. Its distribution across Table 2 describes the selection only. Fifth, retrieval rests on a single exact phrase: the survey covers patents that describe themselves in the vocabulary of information integration, and integration inventions described only in other terms are unobserved.

3. The Patent Landscape of Industrial Information Integration

3.1. Corpus Overview

The retrieval corpus comprises 6627 unique simple patent families published between January 2006 and July 2026. Annual output grows steadily across the window (Figure 2): dated by the earliest publication in each family, the corpus averages 175 families per year in 2006–2008 and 550 in 2023–2025, a roughly threefold increase. Families are dated by first publication because the representative publication of a US family is usually a later document, typically the grant: for 1368 US-office families it postdates the first publication, by three years on average, so dating by the representative would move these families toward the end of the window and overstate growth. Dating by earliest application year or earliest priority year gives the same result (factors of 3.1 and 2.9). Growth is steeper within the in-scope pool, from 29 families per year to 210. In the retrieval corpus the rise is steady to 2015, with growth accelerating after 2015, the period in which Industry 4.0 programs, industrial IoT platforms, and digital-twin initiatives entered mainstream industrial practice. Jurisdictionally, China accounts for 3603 family representatives (54%), the United States for 2878 (43%), and the European route for a small remainder

(EP 140, DE 6); the Chinese share is concentrated in the second half of the window, mirroring the broader expansion of Chinese domestic patenting.

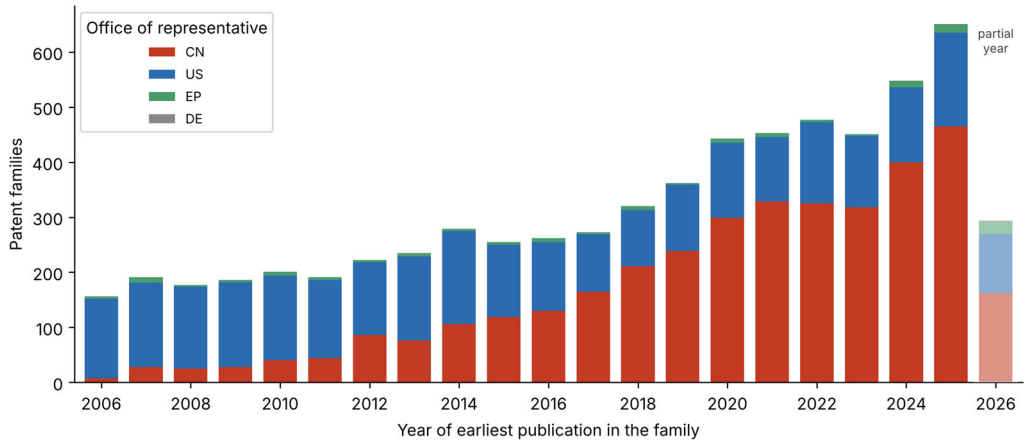


Figure 2. Annual publication volume of the 6627 retrieved patent families by jurisdiction of the representative publication, dated by the earliest publication in each family, January 2006—July 2026 (final year partial).

The US-office and Chinese-office families differ in applicant structure (Figure 3). The two groups are defined by the office of the family's representative publication, not by applicant nationality: several of the largest filers in the US group are Japanese, Korean, or German companies. The US side is concentrated in large information-technology and automation vendors: IBM alone accounts for 279 retrieved families, nearly one in ten US families, followed by Microsoft, NEC, Hitachi, SAP, and Sony, and the ten largest applicant groups in the US group together hold 26.5% of its families (22.6% with names exactly as recorded). Within the classified eligible pool, IBM, Oracle, Boeing, Fisher-Rosemount/Emerson, and Ingram Micro lead. The Chinese side is more fragmented: the ten largest applicant groups hold 10.3% of Chinese-office families (5.7% with names as recorded). The largest, State Grid Corporation of China (73 retrieved families under the parent's own name, and 152 when its provincial companies and research institutes are grouped), is a utility rather than a technology vendor, and it is followed by a long tail of universities, Shanghai Jiao Tong, Zhejiang, Chongqing, Southeast, Beihang, with 15–20 families each and thousands of small and medium-sized enterprises with one or two. Applicant names are those standardized by Lens.org, with each co-applicant credited with the family. Shares are reported after grouping legal entities under their corporate parent by the same name-based concordance on both sides (Supplementary Material B) and, in parentheses, with names exactly as recorded. Grouping matters more on the Chinese side, where State Grid appears under 124 applicant-name variants and China Southern Power Grid under 44. Because the concordance relies on the parent's name appearing in the applicant string, it is a lower bound on consolidation. These statistics describe the retrieval corpus, which includes families later screened out as out of scope. The same pattern holds within the in-scope pool of 3312 families (2059 Chinese-office and 1184 US-office), where the ten largest groups hold 29.6% of US-office and 13.0% of Chinese-office families. The comparison is also conditioned by database coverage: Lens.org indexes US full text more deeply than Chinese full text, and applicant-name standardization may be less complete for Chinese entities, which would itself raise apparent fragmentation. The contrast (26.5% against 10.3%) is therefore a descriptive feature of the retrieved corpus under these rules, not an estimate of industry structure in the two countries. This contrast, vendor-concentrated patenting in the US versus utility- and university-driven patenting in China, recurs throughout the catalog below. A further corpus-level observation is that roughly one third of in-scope records name no application domain in title or abstract, which motivated the domain-general class of the framework and is taken up in Section 3.10.

Sections 3.2–3.10 introduce the 116 representative patent families by application domain, ordered along the industrial value chain, from product design through production, quality, maintenance, and logistics to enterprise systems, followed by the two infrastructure-facing domains (energy; infrastructure and transportation) and closing with domain-general technologies. Within each section, patents are grouped by integration mechanism in the layer order of Table 1.

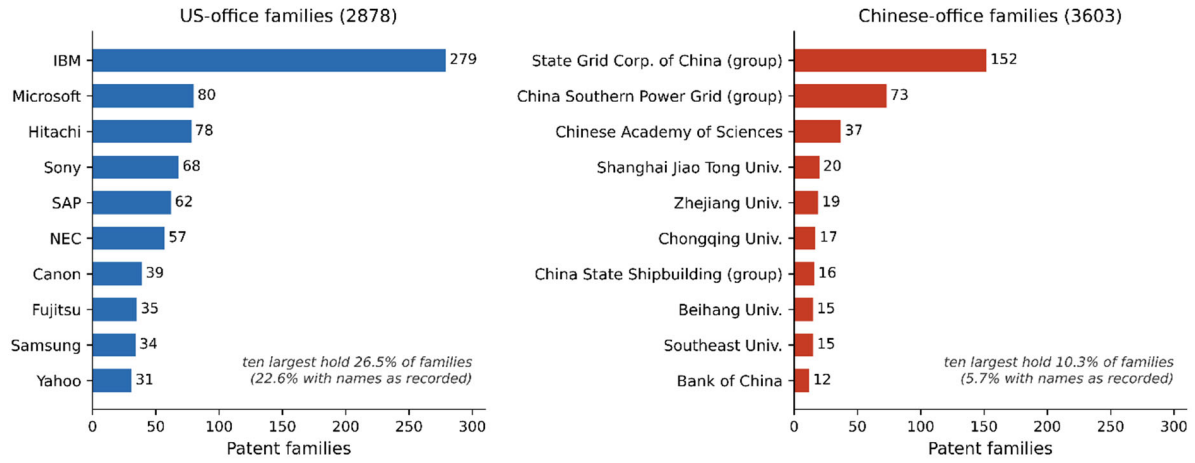


Figure 3. Applicant structure of the retrieval corpus: ten largest applicants by patent families, United States (**left**) and China (**right**). Legal entities are grouped under their corporate parent on both sides (Supplementary Material B). The top 10 concentration shares (26.5% vs. 10.3% after grouping; 22.6% vs. 5.7% with names as recorded) quantify the contrast between vendor-concentrated US patenting and fragmented, utility- and university-driven Chinese patenting.

3.2. Product Design and Engineering (7 Families)

Design-domain patents concentrate on model-centric integration: the engineering model itself, CAD geometry, BIM component libraries, or a design-analysis loop, is the artifact through which information is unified. US 10303159 B2 [70] (Divergent Technologies, 2019) integrates additive manufacturing of vehicle structures around a shared design model: multiple analysis components evaluate the model under different engineering factors, and an integrator iterates it until printing criteria are met, closing a design–analysis–fabrication loop. US 12314639 B1 [71] (Nanjing University of Technology, 2025) builds a BIM-based component parameter library for prefabricated buildings, giving architects and constructors a shared, parameterized vocabulary of components that couples design decisions to construction execution. CN 109472074 A [72] (2019) claims a forward digital design and virtual construction method that maintains a three-dimensional design-and-construction model across the project life cycle, while CN 109145520 A [73] (Chongqing University, 2019) fuses digital imagery of geological sections with support-type and measurement data into a three-dimensional information model for earth–rock tunnel design.

Two families integrate engineering data by movement rather than by model. CN 112699496 A [74] (Beijing Institute of Spacecraft System Engineering, 2021) provides a client–server system that unifies spacecraft maintainability design information, requirements, allocation, modeling, and prediction, in one managed flow, and CN 119357632 A [75] (2025) generates engineering drawings automatically by integrating project IO tables with PLC module information and formatting the recognized drawing logic to user preference. Finally, CN 121479858 A [76] (Xiangtan University, 2026) represents the AI-generative turn in design integration: parameter-driven CAD models are constructed and reverse-reconstructed through topology analysis, feature learning, and generative optimization, folding machine learning into the modeling loop itself.

3.3. Production and Shop Floor (16 Families)

The production domain shows the fullest mechanism spectrum in the corpus, all seven mechanism classes appear, and a clear historical arc from transactional OT–IT bridges to software-defined, secured control architectures. On the model side, CN 112507557 A [77] (China Iron & Steel Research Institute, 2021) builds a multi-agent simulation of the iron–steel interface in which agent models of production processes and transport equipment coordinate through a defined communication and information-integration mechanism; CN 118331175 A [78] (Zhejiang University, 2024) constructs a digital-twin eigenmodel of a CNC machine tool that synchronously evolves with the physical machine's dynamic and static errors; CN 117574691 A [79] (Xiangjiang Laboratory, 2024) encodes a workshop as matrices of entity attributes, spatial layout, environment state, and production constraints, a declarative virtual-entity data system; and US 12524389 B2 [80] (Fisher-Rosemount, 2026) introduces an enterprise engineering and configuration framework whose common configuration schema spans control hardware and software across multiple plant sites, even where the sites run different control systems.

Platform- and data-layer integration anchors the middle of the arc. CN 105467946 A [81] (2016) is an aluminum-electrolysis MES that fuses online tank detection, operator positioning, and production billboards into an information-integration platform for perception-to-decision management, and CN 107168258 A [82] (2017) claims a digitalized flexible manufacturing system for automobile production that couples plan, process, equipment, material, and quality management units around one integrating unit with plant-wide visualization. The earliest US family in this section, US 7162534 B2 [83] (Fisher-Rosemount, 2007), is a transactional data-communications bridge that formats process-control information against an XML schema and exposes it to enterprise IT through web services, a canonical OT–IT mediation patent, while CN 115065697 A [84] (Shanghai Jiao Tong University, 2022) layers field devices, an information-integration layer, and a control layer over OPC UA, the contemporary standard for the same boundary.

Connectivity-layer families supply the acquisition fabric. US 11531325 B2 [85] (Fisher-Rosemount, 2022) manages nodal communication networks of highly versatile field devices, mapping device tag identifications to network endpoints through DHCP/DNS-style services so that field instruments become addressable information nodes; CN 106533886 A [86] (2017) configures a fully interconnected manufacturing network over IPv6, bridging legacy fieldbus segments; CN 105867337 A [87] (2016) wires a backlight-module production line's PLC, data server, and visual management into one line-management system; and CN 104852992 A [88] (2015) spans multiple communication networks and sensing devices into an online safety-supervision platform with a production-safety cloud data center. The remaining families extend production integration upward: CN 114647741 A [89] (2022) reasons over a process knowledge graph built from a three-level model of part, process, and equipment information to automate process decisions; CN 114297020 A [90] (2022) and US 12449789 B2 [91] (Fisher-Rosemount, 2025) secure the integrated shop floor, the former as an industrial-control "safety brain" aggregating structured logs for real-time analysis, the latter embedding container-based security services in a software-defined control system; and US 7380213 B2 [92] (Kimberly-Clark, 2008) contributes the presentation layer, a user interface reporting event-based production information with timestamps, event codes, and cost measures.

3.4. Quality and Inspection (3 Families)

Quality-domain patents divide between event-based quality data management and learned inspection. US 7882438 B2 [93] (2011) applies labels and smart tags in event-based manufacturing to capture material history, machine productivity, waste, and delay on an event basis, quality management grounded in tagged acquisition, and its sibling US 8799113 B2 [94] (2014) validates bills of materials during manufacture against a stored product-specification data structure, so that material events are checked at the point of use. On the learned side, CN 119335261 A [95] (2025) integrates fabric production-processing data with electromagnetic detection signatures to fingerprint chemical residues in textiles, quality inspection that explicitly fuses production information with sensing.

3.5. Maintenance and Asset Management (11 Families)

Maintenance patents integrate information around the asset. Model-centric families include CN 102411764 A [96] (2012), which manages electrical-network equipment data through a three-dimensional grid design model linked to operating-environment and spatial-relation modules, and CN 115034578 A [97] (PowerChina Huadong, 2022), which constructs digital twins of hydraulic metal structures whose physical and information spaces interact through historical, online, and predictive twin data. Data-fusion families extend the reach of inspection: CN 115113650 A [98] (2022) fuses ground information databases with real-time UAV acquisition for autonomous inspection operations, and CN 119989235 A [99] (2025) predicts pipeline state by weighting multi-source monitoring inputs by source credibility before fusion. CN 104167817 A [68] (2014) keeps a power utility's equipment-data platform current by mirroring the real-time databases of its production-side subsystems and online-monitoring systems whenever their data change.

The remaining families cover the mediation, acquisition, learned, and presentation layers. US 10248914 B2 [100] (2019) sustains a fleet of configuration-controlled assets by integrating previously separate systems through gateways and a common message format, packaging and distributing operational data to fleet applications under configuration control. US 9701325 B2 [101] (Korea Railroad Research Institute, 2017) monitors railroad-vehicle equipment through energy-harvesting wireless sensor devices that power themselves from vehicle motion, acquisition-layer integration for rolling-stock health. CN 120948614 A [69] (2025) monitors the beam-column connections of fabricated steel structures ultrasonically, guided by node parameters extracted from a BIM restoration of the design. Two Chinese families of 2025 bring learned health assessment to process assets: CN 120471323 A [102] manages equipment-state data on an automated herbal-medicine production line for precise

health prediction and active maintenance, and CN 119722027 A [103] constructs power-plant equipment health-assessment models over station-information data channels. CN 111176642 A [104] (2020) closes the section at the presentation layer, converting plant equipment diagrams into interactive web-based visual planar graphs whose programmed primitives mirror live equipment state for asset-intensive enterprises.

3.6. Logistics, Supply Chain and Warehousing (4 Families)

The four logistics families each occupy a different mechanism layer, sketching the stack of supply-chain integration in miniature. CN 108389018 A [105] (2018) orchestrates bulk-commodity logistics through a unified service-layer platform whose BPM rule engine calls the service systems relevant to each logistics transaction; CN 114331284 A [106] (2022) manages intelligent warehousing over cloud computing, integrating image-sensed warehouse-in/out events into real-time multi-warehouse inventory; CN 106997509 A [107] (Nanjing University of Aeronautics and Astronautics, 2017) predicts emergency material-distribution demand by integrating indeterminate information from historical disaster cases through rough-set attribute weighting; and CN 111737352 A [108] (Sichuan Changhong, 2020) chains supply-chain participants into private and interconnected blockchains with hybrid consensus, establishing full-link mutual trust as the basis for collaborative information management.

3.7. Enterprise and Cross-Functional Integration (13 Families)

The enterprise domain is the most US-weighted section of the catalog (eleven of thirteen families) and traces enterprise information integration from ontology and warehouse infrastructure to AI-driven business communication. On the representation side, US 8473521 B2 [109] (2013) discovers relationships among reference data in multiple databases by mapping both against a shared enterprise ontology, master-data integration through a declared semantic layer, and US 10521442 B1 [110] (EMC, 2019) governs enterprise data assets through a hierarchical architecture that couples technical and business metadata hubs into a value-annotated data-asset catalog.

Data- and process-layer families span two decades of enterprise integration practice. US 7139999 B2 [111] (Accenture, 2006), the oldest family in the catalog, manages the information environment of software-delivery projects, security requirements, quality metrics, plans, and tracking, in one development-architecture framework. US 7357298 B2 [112] (Kimberly-Clark, 2008) integrates event-based production information with financial and purchasing systems so that manufacturing events populate financial reports directly, a cross-functional bridge between the shop floor and the ledger. US 7890517 B2 [113] (Metatomix, 2011) provides an enterprise-information-integration appliance that unifies disparate agency data sources into a single real-time organizational view, and US 11790155 B2 [114] (DocuSign, 2023) integrates document-workflow state into e-signature services by generating action-responsive document copies while the underlying content remains non-editable. At the mediation layer, EP 4607427 A1 [115] (Ingram Micro, 2025) automates configure-to-order and quote-to-order processes across vendor systems, retrieving bill-of-materials data and market pricing through a real-time data mesh and transferring validated quotes into orders.

The AI-and-analytics cluster concentrates in customer-facing enterprise functions. US 10607165 B2 [116] (Salesforce, 2020) generates suggested next actions inside relationship-management processes by analyzing communication content; US 10623572 B1 [117] and US 10951763 B2 [118] (N3, 2020–2021) build semantic CRM, structured transcripts of communication sessions and a semantic AI agent that administers customer contacts on behalf of tele-agents; and US 11379529 B2 [119] (Microsoft, 2022) enriches message composition by predicting user intents and executing task queries over user-knowledge sources. Two Chinese security families complete the section: CN 108241799 A [120] (2018) provides cross-system access by credentialed simulated login so that a user's authorized identity spans multiple business systems, and CN 114584404 A [121] (2022) protects enterprise project-workflow data in cloud storage by marking data interactions and integrating permission-account information across business nodes.

3.8. Energy and Utilities (18 Families)

Energy is the largest single-utility domain in the catalog and the clearest demonstration of the applicant contrast of Section 3.1: sixteen of eighteen families are Chinese, and State Grid Corporation entities or provincial subsidiaries appear as applicants throughout, patenting the information infrastructure of grid modernization layer by layer. On the semantic layer, CN 102542513 A [122] (Shanghai Jiao Tong University, 2012) verifies power-grid CIM/XML models ontologically, using OWL to check semantic consistency against the standard information model; CN 103490958 A [123] (State Grid, 2014) tests the interchangeability of metering-terminal communication units, operationalizing interoperability as conformance testing; and CN 111160795 A [124] (2020) manages distribution-network graph-model transactions so that the network model itself is versioned as dispatchers issue changes.

The data-integration cluster is the utility workhorse. CN 102646228 A [125] (2012) routes multi-service real-time data of the smart grid through an information-integration platform into a visualized platform database; CN 103715767 A [126] (2014) integrates and displays grid information platform-wide; CN 104213534 A [127] (China Three Gorges, 2014) dispatches cascade reservoirs by integrating external and internal multi-source information under self-adaptive optimization; CN 111967658 A [128] and CN 111985794 A [129] (2020) analyze power-failure plans and line losses on a marketing-and-distribution information-integration platform that pulls from production, marketing, metering, and GIS systems; and CN 118980884 A [130] (State Grid Henan, 2024) positions distribution-network faults rapidly by fusing environment sensing with spatial information integration.

Service-layer families industrialize the platform itself: CN 102710435 A [131] (China EPRI, 2012) adapts system interfaces through a CIM-based adapter stack; CN 104124756 A [132] (State Grid, 2014) monitors a province-level distribution network through a three-layer architecture whose integration layer feeds an information interaction bus; and CN 106230987 A [133] (China Southern Power Grid, 2016) rebuilds utility applications as microservices on an electric-power PaaS cloud with dedicated application- and data-integration modules. On the connectivity layer, CN 110190605 A [134] (2019) regulates distribution-network voltage through edge computing coordinated with air-conditioner demand response, and US 11662760 B2 [135] (General Electric, 2023), one of the section's two US families, secures wireless communication between intelligent electronic devices, power equipment, and management systems. The learned layer closes the arc: US 10418811 B2 [136] (SAS Institute, 2019) predicts grid supply and load from cleansed meter time series; CN 112149877 A [137] (State Grid Jiangsu, 2020) drives urban-grid fault prediction from multi-source data by stepwise regression; CN 118940150 A [138] (2024) positions faults intelligently through a Bayesian network trained on integrated line, load, and equipment-state information; and CN 119864807 A [139] (State Grid Shanghai, 2025) forecasts photovoltaic output with an improved graph neural network over integrated feature data.

3.9. Infrastructure and Transportation (26 Families)

The largest domain section divides into two coherent halves: BIM-centered construction integration, almost entirely Chinese, and transport-system information integration spanning road, rail, maritime, and mining. The BIM cluster is the corpus's densest single technology stream, ten families across a decade. CN 105912651 A [140] (2016) converts heterogeneous engineering data into IFC format for a Web-BIM management system; CN 107844635 A [141] (2018) couples BIM models with microscopic traffic simulation bidirectionally; CN 109214068 A [142] (2019) extracts component and management information from a global BIM model of assembled buildings; CN 112651126 A [143] (2021) optimizes construction through a BIM-modeled feedback loop from site acquisition to human–machine interaction; CN 115034911 A [144] (2022) runs whole-process cost consultation on an evolving project BIM model from investment decision through bidding; CN 115271269 A [145] (2022) plans and safety-controls large prefabricated-part hoisting inside a BIM spatial model; US 2023/0267240 A1 [146] (2023) compares handheld-LiDAR point clouds against the cloud-hosted BIM to manage construction progress in real time; CN 116611155 A [147] (2023) fuses per-mode network topology models into a computable multi-mode transport atlas on a BIM basis; CN 117807691 A [148] (2024) codes BIM components under a standard system with NLP-assisted extraction and blockchain-backed distributed sharing; and CN 119248777 A [149] (2025) manages engineering data through coded virtual-building interaction. The cluster's one non-model entry, US 9070216 B2 [150] (University of Illinois, 2015), overlays photogrammetric as-built point clouds on the as-planned BIM in four-dimensional augmented reality, construction-progress integration delivered at the presentation layer.

The transport half integrates information across vehicles and infrastructure rather than within any one platform. Road-traffic families run from CN 102054025 A [151] (2011), an early traffic-information integration pipeline from acquisition through an integration library to displays, through CN 107729413 A [152] (2018), which re-architects regional traffic management on distributed databases, message queues, and stream computing, to roadside fusion systems: CN 107888877 A [153] (2018) tracks vehicles by storing wireless-signal and video-surveillance data in one database, CN 114120625 A [154] (2022) fuses roadside multi-target radar with cameras to compute per-vehicle speed and trajectory, CN 109257422 A [155] (2019) reconstructs Internet-of-Vehicles perception-network information under unified data-interface rules, and CN 111737524 A [156] (Shanghai University, 2020) integrates UAV multispectral imagery with road-surface sensing in a cloud system for road-anomaly monitoring. CN 110232235 A [157] (2019) standardizes checked building-floor models into cloud databases, model data logistics for the built environment, and CN 107239575 A [158] (2017) serves urban rainstorm-waterlogging risk analysis through a layered IoT service system.

Rail, maritime, and mining complete the domain. CN 112254660 A [159] (China Railway Seventh Group, 2021) integrates tunnel portal and body monitoring assemblies into an early-warning master system; CN 115883789 A [160] (China Academy of Railway Sciences, 2023) monitors railway infrastructure over 5G with an integral information-security protection layer; CN 103353756 A [161] (Wuhan University of Technology, 2013) fuses AIS transponder and VTS radar information through Dempster–Shafer evidence theory to monitor ships underway; CN 111452669 A [162] (2020) guides intelligent bus-fleet charging through an enterprise private cloud and dynamic queuing; CN 112746864 A [163] (2021) coordinates a robot group for fully mechanized mining advance support through an upper-computer information-integration console; CN 120373537 A [164] (2025) predicts periodic roof weighting on steep coal faces by coupling a digital-twin model with clustered machine-learning prediction; and CN 119938916 A [165] (COSCO Shipping Technology, 2025) constructs maritime-document knowledge graphs with large-language-model extraction, integration, and summarization, the catalog's first LLM-based integration family.

3.10. Domain-General Integration Technologies (18 Families)

Nearly one record in six of the final corpus claims integration technology without naming an industry, the patent-side counterpart of the domain-general categories that hold a quarter of the academic review series' literature. Because domain-silent families were not eligible for selection (Section 2.4), these 18 families entered through manual recoding, and their share understates the weight of domain-general patenting in the pool. The section's core is the commercial data-integration tooling of the US software industry, led by Oracle's ETL-design series: US 9507838 B2 [166] (2016) and US 10216814 B2 [167] (2019) simplify ETL map design through component types that propagate attributes and combine flow-based with entity-relationship mapping; US 10073867 B2 [168] (2018) generates executable integration code from platform-independent logical designs via knowledge modules; and US 11847131 B2 [169] (2023) optimizes incremental warehouse loading through partitioning plans and selective index management. They are flanked by US 7421458 B1 [170] (Informatica, 2008), which versions and deploys interrelated database objects across enterprise warehouses, US 8396884 B2 [171] (2013), which indexes and mines graph databases through closure trees and random-walk clustering, and CN 118643012 A [172] (2024), a big-data platform storage-management method for semi- and unstructured data at scale.

The remaining families generalize the connectivity, learned, and security layers. CN 108924198 A [173] (China Unicom, 2018) schedules smart-industrial-internet data across edge-computing resources; CN 113518127 A [174] (2021) packages an industrial-IoT information-integration platform, acquisition gateway, integration and distribution servers, display control, for arbitrary target objects; CN 114867128 A [175] (2022) adapts random access for satellite IoT under interference; and CN 113094601 A [176] (2021) manages multi-vendor GIS map services in the cloud behind standard web-service requests. Learned-layer families include CN 110706095 A [177] (2020), which fills missing key information of a target node from its association network; US 11399060 B2 [178] (2022), which manages continuous AI pipelines with deductive attribution and auditability over ontologies; CN 119647696 A [179] (2025), which predicts carbon emissions from integrated streaming sensor data with transfer learning; CN 119884215 A [180] (2025), which mines conditional, regional, and social geographic information in breadth; and CN 121301554 A [181] (Jiangnan University, 2026), a verifiable retrieval-augmented generation system that grounds large-model answers in evidence chains, domain-general information integration in its most current form. Two security families close the catalog: CN 109587125 A [182] (2019) integrates heterogeneous internal and external sources into streaming security analytics under deep-learning models, and CN 114817340 A [183] (2022) traces data across application systems by link-tracking identities for audit.

4. Conclusions

This survey provides the first systematic mapping of the patent record of industrial information integration, complementing the literature-review series that has periodically mapped the field's academic output [48–50]. From 6627 patent families retrieved across the US, European, German, and Chinese offices for 2006–2026, a screened pool of 3312 in-scope families was classified along two dimensions, integration mechanism and application domain, and 116 representative families were selected, manually coded and reviewed under a written codebook, and introduced descriptively. Four findings organize what the patent record shows.

First, industrial information integration is a growing patenting field with a distinctive geography. Annual output roughly tripled across the window, accelerating after 2015 with the mainstreaming of Industry 4.0, industrial IoT, and digital-twin programs. The two dominant patenting systems differ structurally rather than merely in volume: US patenting is vendor-concentrated, the ten largest applicant groups hold 26.5% of US-office families, led by IBM, and productizes integration as reusable tooling, from ETL design suites and control-system platforms

to CRM intelligence. Chinese patenting is fragmented, the ten largest applicant groups hold about 10% of families, and deployment-driven, led by a utility rather than a vendor and populated by universities and thousands of small enterprises patenting integration in the context of specific infrastructure programs: grid modernization, BIM-based construction, rail and mining systems. This contrast is conditional on the applicant-name treatment and the database coverage discussed in Section 3.1.

Second, patenting concentrates in the data and model classes rather than at the human–machine interface. Two imperfect sources agree on this ordering. Among the 116 manually coded families, 58 fall in the two data classes (data models, semantics and digital twins; databases, fusion and synchronization) and three in human–machine interfaces. Among the 1944 pool families to which the keyword classifier assigned a mechanism, the corresponding figures are 1080 (56%) and 45 (2%). The first is a selected set and the second rests on labels that matched manual codes about half the time, so the shares are indicative, but the direction is the same in both. Industrial information integration is patented, in other words, chiefly as data- and model-layer infrastructure rather than as user-facing technology, a pattern worth comparing against the academic literature, where human-centred and interface themes are more prominent.

Third, the mechanism × domain matrix (Table 2) shows where the selected families cluster, and it should be read as a description of that selection only. Three streams are dense among the selected families: model-centric BIM integration in construction, platform-by-platform grid informatization in energy, and domain-general ETL tooling. Some cells are empty: no selected family combines middleware and services with infrastructure and transportation, or cybersecurity with energy, and logistics and supply chain, among the most prominent domains of the academic series, contributes only four families. Because the selection favours cited, granted patents and excludes domain-silent families, an empty cell does not show that such patents are absent from the pool. These cells are better treated as questions for a pool-level study (for example, whether the BIM model absorbs the mediation role in construction, or whether utility security patenting is expressed in other vocabulary) than as findings.

Fourth, a substantial share of integration patenting appears domain-general: about one third of the in-scope pool names no application domain in title or abstract, an upper bound on domain-general patenting, and 18 of the 116 selected families were coded domain-general on manual reading. This mirrors the domain-general categories that hold a quarter of the academic series' literature and supports treating industrial information integration as a horizontal technology field, one whose core artifacts (schemas, pipelines, platforms, learned models) migrate across verticals. The window closes with that horizontal layer visibly turning toward AI: learned mechanisms are prominent among the late-window families of the selected set, and the corpus's most recent entries, large-language-model knowledge-graph construction and verifiable retrieval-augmented generation, indicate that the field's next patenting wave is already legible in 2025–2026 filings.

The method's limitations are stated in Section 2.6; three bound the conclusions most directly. The retrieval rests on a single exact phrase, so integration inventions described in other vocabulary are unobserved; coverage and legal-status reliability are uneven across offices, with Japanese and Korean domestic filings outside the searched jurisdictions altogether even though Japanese and Korean firms appear among US applicants; and the descriptive corpus is a curated, grant-weighted sample rather than a probability sample. In addition, the second-round coding figures are retention rates under review rather than independent inter-rater reliability. These limitations chart the extensions: broadening the query vocabulary and jurisdictions (JP, KR, WO); citation-network analysis over the retrieval corpus to trace how integration technology flows between the vendor-concentrated and deployment-driven systems; and, following the precedent of the review series this survey extends, periodic updates, for which the written codebook, the archived dataset, and the reproducible pipeline reported here are designed to serve as the fixed baseline. As industrial information integration's academic literature continues to be mapped at intervals, its patent record now has a first map to be measured against.

Supplementary Materials

The additional data and information can be downloaded at: https://media.sciltp.com/articles/others/2611261247196481/JETIA-26080021-Supplementary-Materials.zip.

Author Contributions

J.L. developed the conceptual framework of the survey, designed the methodology, and wrote the main body of the manuscript. L.F. assisted with the coding of the patent families and with reviewing and editing the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Data will be made available on request.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work the authors used Claude (Anthropic) in order to improve the grammar and language of the manuscript. The use of the same tool as a coder in the patent classification is described in Section 2. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1. Xu, L.D. *Enterprise Integration and Information Architecture: A Systems Perspective on Industrial Information Integration*; CRC Press: Boca Raton, FL, USA, 2015.
2. Xu, L.D. Industrial information integration—An emerging subject in industrialization and informatization process. *J. Ind. Inf. Integr.* **2020**, *17*, 100128. https://doi.org/10.1016/j.jii.2020.100128.
3. Kagermann, H.; Wahlster, W.; Helbig, J. *Recommendations for Implementing the Strategic Initiative Industrie 4.0: Final Report of the Industrie 4.0 Working Group*; acatech, National Academy of Science and Engineering: Frankfurt am Main, Germany, 2013.
4. Xu, L.D.; Xu, E.L.; Li, L. Industry 4.0: State of the art and future trends. *Int. J. Prod. Res.* **2018**, *56*, 2941–2962. https://doi.org/10.1080/00207543.2018.1444806.
5. Lu, Y. Industry 4.0: A survey on technologies, applications and open research issues. *J. Ind. Inf. Integr.* **2017**, *6*, 1–10. https://doi.org/10.1016/j.jii.2017.04.005.
6. Xu, L.D. The contribution of systems science to Industry 4.0. *Syst. Res. Behav. Sci.* **2020**, *37*, 618–631. https://doi.org/10.1002/sres.2705.
7. Sigov, A.; Ratkin, L.; Ivanov, L.A.; et al. Emerging Enabling Technologies for Industry 4.0 and Beyond. *Inf. Syst. Front.* **2024**, *26*, 1585–1595. https://doi.org/10.1007/s10796-021-10213-w.
8. Li, X.; Xu, L.D. A Review of Internet of Things—Resource Allocation. *IEEE Internet Things J.* **2021**, *8*, 8657–8666. https://doi.org/10.1109/jiot.2020.3035542.
9. Zhao, L.; Li, B.; Tan, W.; et al. Joint Coverage-Reliability for Budgeted Edge Application Deployment in Mobile Edge Computing Environment. *IEEE Trans. Parallel Distrib. Syst.* **2022**, *33*, 3760–3771. https://doi.org/10.1109/tpds.2022.3166163.
10. Zhao, L.; Tan, W.; Li, B.; et al. Joint Shareability and Interference for Multiple Edge Application Deployment in Mobile-Edge Computing Environment. *IEEE Internet Things J.* **2022**, *9*, 1762–1774. https://doi.org/10.1109/jiot.2021.3088493.
11. Zhao, L.; Tan, W.; Xu, L.; et al. Crowd-Based Cooperative Task Allocation via Multicriteria Optimization and Decision-Making. *IEEE Syst. J.* **2020**, *14*, 3904–3915. https://doi.org/10.1109/jsyst.2020.2966646.
12. Tan, W.; Zhao, L.; Li, B.; et al. Multiple Cooperative Task Allocation in Group-Oriented Social Mobile Crowdsensing. *IEEE Trans. Serv. Comput.* **2022**, *15*, 3387–3401. https://doi.org/10.1109/tsc.2021.3086097.
13. Zhao, L.; Tan, W.; Li, B.; et al. Multiple cooperative task assignment on reliability-oriented social crowdsourcing. *IEEE Trans. Serv. Comput.* **2022**, *15*, 3402–3416. https://doi.org/10.1109/tsc.2021.3103636.
14. Niu, W.; Tan, W.; Jia, W.; et al. Method Toward Network Embedding Within Homogeneous Attributed Network Using Influential Node Diffusion-Aware. *IEEE Trans. Comput. Soc. Syst.* **2024**, *11*, 2620–2631. https://doi.org/10.1109/tcss.2023.3276159.

15. Tan, W.; Niu, W.; Jia, W.; et al. TNSAR: Temporal Evolution Network Embedding Based on Structural and Attribute Retention. *IEEE Trans. Serv. Comput.* **2023**, *16*, 4418–4431. https://doi.org/10.1109/tsc.2023.3322588.
16. Gorkhali, A.; Li, L.; Shrestha, A. Blockchain: A literature review. *J. Manag. Anal.* **2020**, *7*, 321–343. https://doi.org/10.1080/23270012.2020.1801529.
17. Xu, L.D.; Lu, Y.; Li, L. Embedding Blockchain Technology Into IoT for Security: A Survey. *IEEE Internet Things J.* **2021**, *8*, 10452–10473. https://doi.org/10.1109/jiot.2021.3060508.
18. Viriyasitavat, W.; Xu, L.D.; Bi, Z. Specification Patterns of Service-Based Applications Using Blockchain Technology. *IEEE Trans. Comput. Soc. Syst.* **2020**, *7*, 886–896. https://doi.org/10.1109/tcss.2020.2999574.
19. Viriyasitavat, W.; Xu, L.; Bi, Z.; et al. User-Oriented Selections of Validators for Trust of Internet-of-Thing Services. *IEEE Trans. Ind. Informatics* **2022**, *18*, 4859–4867. https://doi.org/10.1109/tii.2021.3136928.
20. Viriyasitavat, W.; Xu, L.D.; Niyato, D.; et al. Applications of Blockchain in Business Processes: A Comprehensive Review. *IEEE Access* **2022**, *10*, 118900–118925. https://doi.org/10.1109/access.2022.3217794.
21. Li, L.; Zhou, H. A survey of blockchain with applications in maritime and shipping industry. *Inf. Syst. e-Bus. Manag.* **2021**, *19*, 789–807. https://doi.org/10.1007/s10257-020-00480-6.
22. Chen, W.; Bohloul, S.M.; Ma, Y.; et al. A blockchain-based information management system for academic institutions: A case study of international students' workflow. *Inf. Discov. Deliv.* **2022**, *50*, 343–352. https://doi.org/10.1108/idd-01-2021-0010.
23. Sun, Y.; Li, L.; Yu, Z.; et al. Exploring AI models and applications within a system framework. *Syst. Res. Behav. Sci.* **2025**, *42*, 1163–1180. https://doi.org/10.1002/sres.3036.
24. Chen, Y.; Zhou, Z.; Liu, Y.; et al. Dynamic Grouping With a Self-Aware Computational Resource Allocation for Large-Scale Multi-Objective Optimization. *IEEE Trans. Evol. Comput.* **2026**, *30*, 795–809. https://doi.org/10.1109/tevc.2025.3564335.
25. Guan, Y.; Chen, Y.; Liu, Y.; et al. Real-Time Scheduling Framework for Multiagent Cooperative Logistics With Dynamic Supply Demands. *IEEE Trans. Ind. Informatics* **2025**, *21*, 3007–3016. https://doi.org/10.1109/tii.2024.3516131.
26. Sun, Y.; Long, Y.; Xu, L.; et al. Long-Term Matching Optimization With Federated Neural Temporal Difference Learning in Mobility-on-Demand Systems. *IEEE Internet Things J.* **2023**, *10*, 1426–1445. https://doi.org/10.1109/jiot.2022.3206491.
27. Li, X.; Sigov, A.; Ratkin, L.; et al. Artificial intelligence applications in finance: A survey. *J. Manag. Anal.* **2023**, *10*, 676–692. https://doi.org/10.1080/23270012.2023.2244503.
28. Zou, B.; Wang, H.; Li, H.; et al. Predicting stock index movement using twin support vector machine as an integral part of enterprise system. *Syst. Res. Behav. Sci.* **2022**, *39*, 428–439. https://doi.org/10.1002/sres.2862.
29. Zhao, S.; Li, S.; Qi, L.; et al. Computational Intelligence Enabled Cybersecurity for the Internet of Things. *IEEE Trans. Emerg. Top. Comput. Intell.* **2020**, *4*, 666–674. https://doi.org/10.1109/tetci.2019.2941757.
30. Li, L.; Xu, L.; He, W. The effects of antecedents and mediating factors on cybersecurity protection behavior. *Comput. Hum. Behav. Rep.* **2022**, *5*, 100165. https://doi.org/10.1016/j.chbr.2021.100165.
31. Shi, Z.; Xie, Y.; Xue, W.; et al. Smart factory in Industry 4.0. *Syst. Res. Behav. Sci.* **2020**, *37*, 607–617. https://doi.org/10.1002/sres.2704.
32. Ge, J.; Wang, F.; Sun, H.; et al. Research on the maturity of big data management capability of intelligent manufacturing enterprise. *Syst. Res. Behav. Sci.* **2020**, *37*, 646–662. https://doi.org/10.1002/sres.2707.
33. Sun, Y.; Li, L.; Shi, H.; et al. The transformation and upgrade of China's manufacturing industry in Industry 4.0 era. *Syst. Res. Behav. Sci.* **2020**, *37*, 734–740. https://doi.org/10.1002/sres.2714.
34. Liu, Z.; Xie, K.; Li, L.; et al. A paradigm of safety management in Industry 4.0. *Syst. Res. Behav. Sci.* **2020**, *37*, 632–645. https://doi.org/10.1002/sres.2706.
35. Fu, L.; Li, L.; Li, L.; et al. Impact of hospital size on healthcare information system effectiveness: Evidence from healthcare data analytics. *J. Manag. Anal.* **2022**, *9*, 211–231. https://doi.org/10.1080/23270012.2022.2036647.
36. Dobson, R.; Stowell, M.; Warren, J.; et al. Use of Consumer Wearables in Health Research: Issues and Considerations. *J. Med. Internet Res.* **2023**, *25*, e52444. https://doi.org/10.2196/52444.
37. Zhao, Y.; Wang, X.; Peng, X.; et al. Selecting and Evaluating Key MDS-UPDRS Activities Using Wearable Devices for Parkinson's Disease Self-Assessment. *IEEE J. Sel. Areas Sensors* **2024**, *1*, 177–189. https://doi.org/10.1109/jsas.2024.3432714.
38. Gu, Y.; Kalibatseva, Z.; Song, X.; et al. Effective use of online COVID-19 information and eHealth information literacy among US university students. *J. Am. Coll. Health* **2024**, *72*, 1458–1465. https://doi.org/10.1080/07448481.2022.2080505.
39. Jia, H.; Huan, Y.; Ding, C.; et al. A Domain-Specific Accelerator for Ultralow Latency Market Data Distribution System. *IEEE Trans. Ind. Informatics* **2023**, *19*, 5465–5475. https://doi.org/10.1109/tii.2022.3182242.
40. Chen, H.; Li, L.; Chen, Y. Sustainable growth research—A study on the telecom operators in China. *J. Manag. Anal.* **2022**, *9*, 17–31. https://doi.org/10.1080/23270012.2021.1980445.

41. Shen, B.; Xu, L.D.; Cai, H.; et al. Enhancing Context-Aware Reactive Planning for Unexpected Situations of On-Orbit Spacecraft. *IEEE Trans. Aerosp. Electron. Syst.* **2022**, *58*, 4965–4983. https://doi.org/10.1109/taes.2022.3172022.
42. Wang, W.; Niu, N.; Alenazi, M.; et al. Complementarity in Requirements Tracing. *IEEE Trans. Cybern.* **2020**, *50*, 1395–1404. https://doi.org/10.1109/tcyb.2018.2889830.
43. Li, L.; Duan, L. Human centric innovation at the heart of industry 5.0—Exploring research challenges and opportunities. *Int. J. Prod. Res.* **2025**, *64*, 1–33. https://doi.org/10.1080/00207543.2025.2462657.
44. Li, L. Reskilling and Upskilling the Future-ready Workforce for Industry 4.0 and Beyond. *Inf. Syst. Front.* **2024**, *26*, 1697–1712. https://doi.org/10.1007/s10796-022-10308-y.
45. Li, L. Education supply chain in the era of Industry 4.0. *Syst. Res. Behav. Sci.* **2020**, *37*, 579–592. https://doi.org/10.1002/sres.2702.
46. Chen, Y. Industrial information integration—A literature review 2006–2015. *J. Ind. Inf. Integr.* **2016**, *2*, 30–64. https://doi.org/10.1016/j.jii.2016.04.004.
47. Lu, Y. Industrial Integration: A Literature Review. *J. Ind. Integr. Manag.* **2016**, *1*, 1650007. https://doi.org/10.1142/s242486221650007x.
48. Chen, Y. A Survey on Industrial Information Integration 2016–2019. *J. Ind. Integr. Manag.* **2020**, *5*, 33–163. https://doi.org/10.1142/s2424862219500167.
49. Li, J. Part I: Industrial Information Integration Review 2020–2025. *J. Emerg. Technol. Ind. Appl.* **2026**, *1*, 1. https://doi.org/10.53941/jetia.2026.100001.
50. Li, J. Part II: Industrial Information Integration Review 2020–2025. *J. Emerg. Technol. Ind. Appl.* **2026**, *1*, 2. https://doi.org/10.53941/jetia.2026.100002.
51. Jefferson, O.A.; Jaffe, A.; Ashton, D.; et al. Mapping the global influence of published research on industry and innovation. *Nat. Biotechnol.* **2018**, *36*, 31–39. https://doi.org/10.1038/nbt.4049.
52. Litvinova, O.; Hammerle, F.P.; Stoyanov, J.; et al. Patent and Bibliometric Analysis of the Scientific Landscape of the Use of Pulse Oximeters and Their Prospects in the Field of Digital Medicine. *Healthcare* **2023**, *11*, 3003. https://doi.org/10.3390/healthcare11223003.
53. IEC. *Enterprise-Control System Integration, Part 1: Models and Terminology*; International Electrotechnical Commission: Geneva, Switzerland, 2013.
54. DIN. *SPEC 91345:2016-04*; Reference Architecture Model Industrie 4.0 (RAMI 4.0), DIN: Berlin, Germany, 2016. https://doi.org/10.31030/2436156.
55. IEC. *Security for Industrial Automation and Control Systems*; 62443 (Series); International Electrotechnical Commission: Geneva, Switzerland, 2019.
56. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022. https://doi.org/10.5555/944919.944937.
57. Venugopalan, S.; Rai, V. Topic based classification and pattern identification in patents. *Technol. Forecast. Soc. Chang.* **2015**, *94*, 236–250. https://doi.org/10.1016/j.techfore.2014.10.006.
58. Suominen, A.; Toivanen, H.; Seppänen, M. Firms' knowledge profiles: Mapping patent data with unsupervised learning. *Technol. Forecast. Soc. Chang.* **2017**, *115*, 131–142. https://doi.org/10.1016/j.techfore.2016.09.028.
59. Mastilović, J.; Kukolj, D.; Kevrešan, Ž.; et al. Emerging Perspectives of Blockchains in Food Supply Chain Traceability Based on Patent Analysis. *Foods* **2023**, *12*, 1036. https://doi.org/10.3390/foods12051036.
60. Ahmed, T.A.E.; Younes, M.; Wu, L.; et al. A Survey of Recent Patents in Engineering Technology for the Screening, Separation and Processing of Eggshell. *Front. Bioeng. Biotechnol.* **2021**, *9*, 677559. https://doi.org/10.3389/fbioe.2021.677559.
61. Ahmed, T.A.E.; Wu, L.; Younes, M.; et al. Biotechnological Applications of Eggshell: Recent Advances. *Front. Bioeng. Biotechnol.* **2021**, *9*, 675364. https://doi.org/10.3389/fbioe.2021.675364.
62. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. https://doi.org/10.1177/001316446002000104.
63. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159–174. https://doi.org/10.2307/2529310.
64. MacQueen, K.M.; McLellan, E.; Kay, K.; et al. Codebook Development for Team-Based Qualitative Analysis. *CAM J.* **1998**, *10*, 31–36. https://doi.org/10.1177/1525822x980100020301.
65. Campbell, J.L.; Quincy, C.; Osserman, J.; et al. Coding In-depth Semistructured Interviews. *Sociol. Methods Res.* **2013**, *42*, 294–320. https://doi.org/10.1177/0049124113500475.
66. Gilardi, F.; Alizadeh, M.; Kubli, M. ChatGPT outperforms crowd workers for text-annotation tasks. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2305016120. https://doi.org/10.1073/pnas.2305016120.
67. Ziems, C.; Held, W.; Shaikh, O.; et al. Can Large Language Models Transform Computational Social Science? *Comput. Linguist.* **2024**, *50*, 237–291. https://doi.org/10.1162/coli_a_00502.

68. Shenzhen Power Supply Bureau. Power Equipment Real-Time Information Integration System and Method. Patent CN 104167817 A, 2014.
69. Nantong Huazhenglong Steel Mfg Co. Ltd. Fabricated Steel Structure Beam-Column Connection Performance Monitoring Method Based on Ultrasound. Patent CN 120948614 A, 2025.
70. Divergent Tech Inc. Systems and Methods for Additive Manufacturing of Transport Structures. Patent US 10303159 B2, 2019.
71. Nanjing University of Technology. Prefabricated Building Design and Construction Integrated Collaborative Method Based on Component Parameter Library. Patent US 12314639 B1, 2025.
72. Liu, Y. A Digital Design and Virtual Construction Method Capable of Implementing forward Construction Engineering. Patent CN 109472074, 2019.
73. University of Chongqing. Design Method of Earth-Rock Mixture Tunnel Based on Digital Image and Big Data. Patent CN 109145520 A, 2019.
74. Beijing Inst Spacecraft System Engineering. Maintainability Design and Analysis Integrated System for Spacecraft. Patent CN 112699496 A, 2021.
75. Shanghai Jiling Information Tech Co. Ltd. Project Information Multi-Dimensional Integration-Combined Drawing Automatic Generation Method and Device. Patent CN 119357632 A, 2025.
76. University of Xiangtan. CAD Automatic Modeling Method and System Based on Parameter Driving. Patent CN 121479858 A, 2026.
77. China Iron & Steel Res Institute Group. Iron-Steel Interface Simulation System Based on Multiple Agents. Patent CN 112507557 A, 2021.
78. University of Zhejiang. Numerical Control Machine Tool Digital Twinning Synchronous Evolution Control Method Associated with Dynamic and Static Errors. Patent CN 118331175 A, 2024.
79. Xiangjiang Laboratory. Virtual Entity Data System Construction Method and Related Equipment. Patent CN 117574691 A, 2024.
80. Fisher Rosemount Systems Inc. Enterprise Engineering and Configuration Framework for Advanced Process Control and Monitoring Systems. Patent US 12524389 B2, 2026.
81. Guiyang Aluminum Magnesium Design & Res Inst Co. Ltd. Aluminum Electrolytic MES System Based on Accurate Perception and Intelligent Decision. Patent CN 105467946 A, 2016.
82. Datang Radio And Television Tech (Wuhan) Co. Ltd. Digitalized Flexible System and Management Method Used for Automobile Manufacturing. Patent CN 107168258 A, 2017.
83. Fisher Rosemount Systems Inc. Transactional Data Communications for Process Control Systems. Patent US 7162534 B2, 2007.
84. University of Shanghai Jiaotong. Industrial Network Equipment Information Integration System and Method Based on OPC UA. Patent CN 115065697 A, 2022.
85. Fisher Rosemount Systems Inc. Node Management of Nodal Communication Networks for Highly Versatile Field Devices in Control and Automation Systems. Patent US 11531325 B2, 2022.
86. University of Chongqing Posts & Telecom. Configuration and Management Method of Full-Interconnected Manufacturing Network Based on IPv6 Protocols. Patent CN 106533886 A.
87. Heyuan Coe Communication Tech Co. Ltd. Management System of Automatic Production Line of Backlight Modules and Method of Management System. Patent CN 105867337 A, 2016.
88. Hubei Skyshield Electronical Technology Co. Ltd. On-Line Intelligent System for Safety Emergency Based on Multiple Communication Networks and Sensing Equipment. Patent CN 104852992 A, 2015.
89. Guangdong Technology Normal Univ. Process Automatic Decision-Making and Reasoning Method and Device, Computer Equipment and Storage Medium. Patent CN 114647741 A, 2022.
90. Jiangsu Forest And Ocean Energy Co. Ltd. Enterprise Industrial Control Safety Brain Platform System and Operation Method. Patent CN 114297020 A, 2022.
91. Fisher Rosemount Systems Inc. Security Services in a Software Defined Control System. Patent US 12449789 B2, 2025.
92. Kimberly Clark Co. User Interface for Reporting Event-Based Production Information in Product Manufacturing. Patent US 7380213 B2, 2008.
93. Binforma Group Ltd. Liability Company. Quality Management and Intelligent Manufacturing with Labels and Smart Tags in Event-Based Product Manufacturing. Patent US 7882438 B2, 2011.
94. Markham Charles Earl. Quality Management by Validating a Bill of Materials in Event-Based Product Manufacturing. Patent US 8799113 B2, 2014.
95. Hunan Fuli Rich Home Textiles Co. Ltd. Textile Chemical Residue Detection Method and System and Medium. Patent CN 119335261 A, 2025.

96. Guangdong Elec Power Des Inst. Data Management System of Electrical Network Equipment and Management Method Thereof. Patent CN 102411764 A, 2012.
97. Powerchina Huadong Engineering Corp. Ltd. Hydraulic Metal Structure Equipment Intelligent Management Construction Method and System Based on Digital Twinning. Patent CN 115034578 A, 2022.
98. Nanjing Jincheng Data Tech Co. Ltd. Air-Ground Integrated Data Fusion System Based on Autonomous Operation of Unmanned Aerial Vehicle. Patent CN 115113650 A, 2022.
99. Ailf Shandong Instr Co. Ltd. Pipeline State Prediction Method Based on Big Data. Patent CN 119989235 A, 2025.
100. Stoughton Chris. Sustaining a Fleet of Configuration-Controlled Assets. Patent US 10248914 B2, 2019.
101. Korea Railroad Res Inst. Energy Harvester, Wireless Sensor Device Having the Energy Harvester, and System for Monitoring Railroad Vehicle Using the Same. Patent US 9701325 B2, 2017.
102. Jiangsu Jurong Pharmaceutical Group Co. Ltd. Equipment State Data Management Method and System Applied to Automatic Radix Astragali seu Hedysari Refined Production Line. Patent CN 120471323 A, 2025.
103. Nat Energy Nanning Power Generation Co. Ltd. Construction Method and System of Power Plant Equipment Health Assessment Model. Patent CN 119722027 A, 2025.
104. Zhuhai Tohi Tech Co. Ltd. Interactive Visual Processing System Based on Planar Graph and Application Method. Patent CN 111176642 A, 2020.
105. Circle Logistics Co. Ltd. Intelligent Bulk Commodity Logistics Management System. Patent CN 108389018 A, 2018.
106. Shenzhen Kaongyuan Modern Logistics Co. Ltd. Intelligent Warehousing Service Management System Based on Cloud Computing. Patent CN 114331284 A, 2022.
107. University of Nanjing Aeronautics & Astronautics. Indeterminate Information Integration Emergency Material Distribution Demand Prediction Method. Patent CN 106997509 A, 2017.
108. Sichuan Changhong Electric Co. Ltd. Supply Chain Information Collaborative Management Method Based on Block Chain. Patent CN 111737352 A, 2020.
109. Fot Dmitriy, Ontology Guided Reference Data Discovery. Patent US 8473521 B2, 2013.
110. EMC Ip Holding Co. LLC. Hierarchical Value-Based Governance Architecture for Enterprise Data Assets. Patent US 10521442 B1, 2019.
111. Accenture Llp. Development Architecture Framework. Patent US 7139999 B2, 2006.
112. Kimberly Clark Co. Integrating Event-Based Production Information with Financial and Purchasing Systems in Product Manufacturing. Patent US 7357298 B2, 2008.
113. Metatomix Inc. Appliance for Enterprise Information Integration and Enterprise Resource Interoperability Platform and Methods. Patent US 7890517 B2, 2011.
114. Docusign Inc. Electronic Signing Using Action Responsive Document Copy Generation. Patent US 11790155 B2, 2023.
115. Ingram Micro Inc. Systems and Methods for Automated Configuration to Order and Quote to Order. Patent EP 4607427 A1, 2025.
116. Salesforce Com Inc. Systems and Methods for Automatic Suggestions in a Relationship Management System. Patent US 10607165 B2, 2020.
117. N3 LLC. Semantic CRM Transcripts from Mobile Communications Sessions. Patent US 10623572 B1, 2020.
118. N3 LLC. Semantic Artificial Intelligence Agent. Patent US 10951763 B2, 2021.
119. Microsoft Technology Licensing LLC. Composing Rich Content Messages. Patent US 11379529 B2, 2022.
120. Guangzhou Transense Software Tech Co. Ltd. Cross-System Access Method, System and Device and Computer Readable Storage Medium. Patent CN 108241799 A, 2018.
121. Suzhou Wishi Hui Information Tech Limited Company. Data Security Protection System and Method Based on Cloud Storage Technology. Patent CN 114584404 A, 2022.
122. University of Shanghai Jiaotong. Body-Based Verification Tool of Power Grid Public Information Model and Method Thereof. Patent CN 102542513 A, 2012.
123. State Grid Corp. China. Method for Testing Local Communication Unit Interchangeability of Acquisition Terminal. Patent CN 103490958 A, 2014.
124. Wuhu Power Supply Company State Grid Anhui Electric Power Company. Management System and Method for Distribution Network Graph-Model Transaction and Power Failure Lean Assessment. Patent CN 111160795 A, 2020.
125. Tianjin Electric Power Corp. Multi-Service Real-Time Data Integration Processing System and Method of Intelligent Power Grid. Patent CN 102646228 A, 2012.
126. State Grid Corp. China. Smart Power Grid Information Integration and Display Platform. Patent CN 103715767 A, 2014.
127. China Three Gorges Corp. Cascade-Reservoir Self-Adaptive Integrated Dispatching System and Dispatching Method Integrating Multi-Source Information. Patent CN 104213534 A, 2014.
128. Guangdong Topway Network Co. Ltd. Comprehensive Power Failure Analysis Method Based on Marketing and Distribution Information Integration Platform. Patent CN 111967658 A, 2020.

129. Guangdong Topway Network Co. Ltd. Line Loss Management Method Based on Marketing and Distribution Information Integration Platform. Patent CN 111985794 A, 2020.
130. State Grid Henan Electric Power Company Gongyi Power Supply Company. Rapid Fault Positioning System for Power Distribution Network. Patent CN 118980884 A, 2024.
131. China Electric Power Res Inst. System Integration Interface Adapter Based on Common Information Model (CIM). Patent CN 102710435 A, 2012.
132. State Grid Corp. China. Province-Level Power Distribution Network Operation Monitoring System Based on Network-Wide Data. Patent CN 104124756 A, 2014.
133. China South Power Grid Electric Power Res Inst Co. Ltd. Information Integration System and Method Based on Electric Power PaaS Cloud Platform. Patent CN 106230987 A, 2016.
134. University of Hangzhou Dianzi. Power Distribution Network Voltage Regulation Control System and Method Based on Edge Calculation And Air Conditioner Demand Response. Patent CN 110190605 A, 2019.
135. Gen Electric. Wireless Communication Systems and Methods for Intelligent Electronic Devices. Patent US 11662760 B2, 2023.
136. Sas Inst Inc. Electric Power Grid Supply and Load Prediction Using Cleansed Time Series Data. Patent US 10418811 B2, 2019.
137. State Grid Jiangsu Electric Power Co. Ltd. Suzhou Power Supply Branch, Multivariate Complex Urban Power Grid Urban Power Grid Multi-Source Data Driven Fault Prediction Method and System. Patent CN 112149877 A, 2020.
138. Suzhou Weixun Optoelectronics Tech Co. Ltd. Intelligent Fault Positioning Method and System Based on Multi-Source Electric Power Information Integration. Patent CN 118940150 A, 2024.
139. State Grid Shanghai Municipal Electric Power Co. Photovoltaic Output Prediction Method and System Based on Improved MTGNN Model. Patent CN 119864807 A, 2025.
140. China Utone Construction Consulting Co. Ltd. IFC-Based Data Interaction Method and Corresponding Web-BIM Engineering Information Integration Management System. Patent CN 105912651 A, 2016.
141. Shanghai Municipal Eng Design Inst Group Co. Ltd. BIM Information and Traffic Simulation Information Integration System and Integration Method Thereof. Patent CN 107844635 A, 2018.
142. Xuzhou Zm Besta Heavy Steel Structure Co. Ltd. BIM-Based Bottom-Assembled Building Information Extraction Method. Patent CN 109214068 A, 2019.
143. Henan No 8 Construction Group Co. Ltd. Building Engineering Construction Optimization System and Method Based on BIM, Terminal and Storage Medium. Patent CN 112651126 A, 2021.
144. Guangzhou Gaoxin Project Man Co. Ltd. BIM-Based Whole-Process Cost Consultation Service Method and System. Patent CN 115034911 A, 2022.
145. China Chemical Lifting Transp Co. Ltd. Large Prefabricated Part Hoisting Safety Control Method Based on BIM. Patent CN 115271269 A, 2022.
146. Lau Wai Man Raymond. A System for Managing the Progress of Construction Based on Lidar Technology. Patent US 2023/0267240 A1, 2023.
147. Shenzhen Urban Transp Planning Center Co. Ltd. BIM (Building Information Modeling)-Based Modeling Method for Computable Multi-Mode Network Atlas Topology Model. Patent CN 116611155 A, 2023.
148. Zhongyifeng Digital Tech Group Co. Ltd. Information Model Data Integration Method and System Based on Building Visualization. Patent CN 117807691 A, 2024.
149. Yinzhu Suzhou Building Tech Co. Ltd. Method and Device for Engineering Data Management through Virtual Building Interaction. Patent CN 119248777 A, 2025.
150. University of Illinois. Four-Dimensional Augmented Reality Models for Interactive Visualization and Automated Construction Progress Monitoring. Patent US 9070216 B2, 2015.
151. China Trancomm Technologies Co. Ltd. Traffic Information Resource Integration Processing Method and System. Patent CN 102054025 A, 2011.
152. Anhui Changtong Traffic Information Service Co. Ltd. Regional Traffic Intelligent Management System Based on Big Data. Patent CN 107729413 A, 2018.
153. Mao, G. Vehicle Tracking and Road Traffic Information Collection Method and System Thereof. Patent CN 107888877 A, 2018.
154. Universal Automobile Ltd. Company. Vehicle Information Integration System and Method, and Storage Medium. Patent CN 114120625 A, 2022.
155. Guangzhou Zhihong Tech Co. Ltd. Perceptual Network Information Reconstruction Method. Patent CN 109257422 A, 2019.
156. University of Shanghai. Information Integration Method in Road Anomaly Monitoring System. Patent CN 111737524 A, 2020.

157. Beijing Yushu Tech Co. Ltd. Relational Data Processing Method and Device Based on Three-Dimensional Model and Electronic Equipment. Patent CN 110232235 A, 2019.
158. Handan Meteorological Service. Internet of Things Service System and Method of Risk Analysis and Early-Warning of Urban Road Rainstorm and Waterlogging. Patent CN 107239575 A, 2017.
159. China Railway Seventh Group Co. Ltd. Tunnel Intelligent Monitoring Measurement Information Integration and Early Warning System. Patent CN 112254660 A, 2021.
160. China Academy Railway Sciences Corp. Ltd. Railway Infrastructure Monitoring and Information Security Protection System Based on 5G. Patent CN 115883789 A, 2023.
161. University of Wuhan Tech. Method for Monitoring Underway Ship in Real Time Based on AIS and VTS Information Integration. Patent CN 103353756 A, 2013.
162. Ouyang, L. Intelligent Bus Charging System and Method and Medium. Patent CN 111452669 A, 2020.
163. Taiyuan Inst China Coal Tech & Eng Group. Intelligent Robot Group for Fully-Mechanized Mining Advance Support. Patent CN 112746864 A, 2021.
164. University of Xian Sci & Technology. Regional Prediction Method for Periodic Weighting of Large-Dip-Angle Working Face. Patent CN 120373537 A, 2025.
165. Cosco Shipping Technology Co. Ltd. Maritime Document Knowledge Graph Construction Method and System. Patent CN 119938916 A, 2025.
166. Oracle Int Corp. Use of Projector and Selector Component Types for ETL Map Design. Patent US 9507838 B2, 2016.
167. Oracle Int Corp. Supporting Combination of Flow Based ETL and Entity Relationship Based ETL. Patent US 10216814 B2, 2019.
168. Oracle Int Corp. System and Method for Code Generation from a Directed Acyclic Graph Using Knowledge Modules. Patent US 10073867 B2, 2018.
169. Oracle Int Corp. Optimizing Incremental Loading of Warehouse Data. Patent US 11847131 B2, 2023.
170. Informatica Corp. Querying, Versioning, and Dynamic Deployment of Database Objects. Patent US 7421458 B1, 2008.
171. Singh, A.K. Graph Querying, Graph Motif Mining and the Discovery of Clusters. Patent US 8396884 B2, 2013.
172. Yantai Oolanduo Information Tech Co. Ltd. Big Data Platform Data Storage Management Method. Patent CN 118643012 A, 2024.
173. China Unicom. Data Scheduling Method, Device and System Based on Edge Computing. Patent CN 108924198 A, 2018.
174. Luo, H. Industrial Internet of Things Information Integration Platform. Patent CN 113518127 A, 2021.
175. University of Nanjing Posts & Telecommunications. Satellite Internet of Things Random Access Adaptive Method and Device, and Storage Medium. Patent CN 114867128 A, 2022.
176. Geological Surveying And Mapping Inst. GIS Cloud Management Method and System. Patent CN 113094601 A, 2021.
177. Sichuan Xw Bank Co. Ltd. Target Node Key Information Filling Method and System Based on Association Network. Patent CN 110706095 A, 2020.
178. Phacil LLC. System and Method for Continuous AI Management and Learning. Patent US 11399060 B2, 2022.
179. University of Hefei Technology. Carbon Emission Prediction Method and System Based on Artificial Intelligence. Patent CN 119647696 A, 2025.
180. Jinan Inst Of Survey & Mapping. Geographic Information Mining Method and System Based on Big Data. Patent CN 119884215 A, 2025.
181. University of Jiangnan. Verifiable Large Model Retrieval Enhancement Generation System and Method Based on Evidence Chain. Patent CN 121301554 A, 2026.
182. Electric Power Res Inst Co. Ltd. Csg. Network Safety Big Data Analysis Method and System, and Related Device. Patent CN 109587125 A, 2019.
183. Shenzhen Hongtu Tech Co. Ltd. Data Tracing Method and Device, Computer Equipment and Storage Medium. Patent CN 114817340 A, 2022.