CTU Journal of Innovation and Sustainable Development
ISSN 2588-1418 | e-ISSN 2815-6412

DOI:10.22144/ctujoisd.2026.058

Patent network analysis of core technology groups and convergence hubs in generative-AI-enabled digital transformation

Chitaranjan Laishram Singh, and Priyanka Chand Bhatt*

Department of Information Management, Yuan Ze University, Taiwan

**Corresponding author (pia@saturn.yzu.edu.tw)*

Article info.

Received 28 Jun 2026
Revised 7 Sep 2026
Accepted 15 Oct 2026

Keywords

CPC domain analysis, digital transformation, generative artificial intelligence, patent co-classification network, technology convergence

ABSTRACT

Generative AI (GenAI) is redefining the technological architecture of enterprise digital transformation, yet its domain structure and convergence topology remain uncharacterized at the patent level. This study applies a three-stage patent analytics pipeline- Cooperative Patent Classification domain mapping, keyword-rule topic assignment across 15 GenAI application archetypes, and weighted co-occurrence network analysis- to 9,692 Lens.org patent records from ten jurisdictions, of which 7,623 published from 2018 onward form the analytical corpus. Filings grew at a compound annual rate of 42% over 2018–2025. Domain mapping identifies Computing & Information Systems (69.5%), AI & Machine Learning (36.5%) and Business & Commerce (25.2%) as a three-layer infrastructure, model and application stack. Large language model development and retrieval-augmented generation together account for 64.5% of filings, while agentic AI and edge inference clusters remain under 1% each. The co-occurrence network (26 nodes, 151 edges, density 0.465) identifies Computing & Information Systems as the dominant convergence hub (composite score 0.759) and Healthcare Informatics as a structural broker (betweenness 0.138) disproportionate to its filing volume. Cross-domain patenting peaked at 2.34 domains per patent in 2021 and moderated to 1.65 by 2025. The findings extend general-purpose technology convergence theory to GenAI and provide a structural map for technology managers.

1. INTRODUCTION

The rise of generative AI, systems producing human-level outputs across data types, marks one of the most consequential shifts in applied artificial intelligence since deep learning. The technical precondition for this inflexion was the publication of the Transformer architecture by Vaswani et al. (2017) in 2017, which introduced scaled self-attention as a parallelizable alternative to recurrent sequence models and enabled the training of foundation models on corpora of previously intractable size. The subsequent scaling of these Transformer-derived architectures into Large Language Models (LLMs), exemplified by the GPT series, BERT, T5, and their successors, endowed these systems with emergent capabilities in reasoning and domain-adaptive inference that had not been anticipated from their architectural properties alone (Bubeck et al., 2023). GenAI's significance is not confined to performance improvements (Chkirbene et al., 2024) on established benchmarks but extends to its capacity to restructure the technological architecture of Digital Transformation (DT) itself. GenAI is a distinct wave because it functions as a horizontal

capability layer, augmenting knowledge work across domains, from coding and legal review to healthcare and industry, blurring boundaries between technology application areas. One-third of organisations are already regularly deploying GenAI in at least one business function within months of public model availability (McKinsey & Company, 2023). By early 2024, that figure had nearly doubled to 65%, with the average adopting organisation using GenAI across two distinct business functions (McKinsey & Company, 2024). Three-quarters of respondents anticipated significant or disruptive competitive change in their industries within three years, an executive-level consensus that the technology constitutes a systemic rather than incremental force. Furthermore, the growth in GenAI patent family publications from fewer than 800 in 2014 to more than 14,000 in 2023, an increase exceeding 800%, with China producing more than 38,000 inventor-attributed patent families over the decade, outpacing the United States by a factor of six (World Intellectual Property Organization [WIPO], 2024).

Situated within the research literature on technology convergence, the process by which previously distinct technological domains progressively merge at their boundaries to generate novel innovation clusters (Curran et al., 2010), GenAI exhibits the hallmark characteristics of a General-Purpose Technology (GPT) as formalised by Gambardella and Torrisi (1998): pervasiveness across sectors, continuous performance improvement, and innovational complementarities that stimulate derivative innovation across the full extent of the host economy. Curran and Leker (2011) established that GPTs characteristically produce patent co-classification convergence detectable through CPC code co-occurrence analysis, applying this framework to trace ICT's convergence with nutraceuticals and functional foods. Whether GenAI now exhibits a comparable convergence pattern across the enterprise digitalization landscape, and through which specific domain pathways, remains an empirically open question that the existing literature has not resolved.

1.1. Motivation

Despite the documented pace of GenAI adoption and the maturity of patent-based technology convergence analysis, the intersection of these two bodies of knowledge remains underexplored in the literature. Patent bibliometrics and CPC co-classification network analysis have been applied to ICT (Jung et al., 2021), CRISPR-based biotechnology (Zhu & Motohashi, 2022), and artificial intelligence at the productivity-macroeconomic interface (Parteka & Kordalska, 2023), consistently revealing the topology of innovation diffusion across sectoral boundaries. Yet no prior study has directed this analytical framework specifically at the GenAI-enabled DT context. Existing work adjacent to this gap has proceeded along two largely separate tracks. The first comprises patent analytics studies that characterize GenAI innovation at a structural level. Mohammadi et al. (2025) employed text mining and patent network modelling, identifying innovation community structures within citation networks. Afifuddin and Seo (2024) developed a machine-learning voting classifier for convergence prediction using CPC co-occurrence networks, demonstrating superior accuracy over link-prediction baselines. Ma and Wu (2024) examined whether AI drives technology convergence in Chinese manufacturing firms using firm-level data, finding positive evidence of convergence effects. The second track comprises application-domain surveys that catalogue LLM use cases across industries. Sindhu et al. (2024) systematically reviewed LLM applications across finance, healthcare, legal services, and education, providing a detailed taxonomy of functional capabilities. Three gaps in the extant literature collectively motivate the present study. First, no published work has constructed a CPC co-classification convergence network specifically for the GenAI-enabled DT patent corpus and computed domain-level structural brokerage metrics. Second, a patent-level topic taxonomy for GenAI enterprise application archetypes, grounded in abstract-level text analysis rather than CPC codes alone, has not been produced or validated at scale. Third, the temporal dynamics of multi-domain patent activity have not been empirically characterised for this corpus. The present study addresses all three gaps through a unified analytics pipeline applied to 7,623 post-2018 patents from ten jurisdictions.

1.2. Research question and objectives

Despite this transition, the scholarly literature lacks a systematic, data-driven characterization of which technology domains GenAI is penetrating and how it is catalyzing convergence across previously distinct innovation clusters. Patent corpora represent the earliest measurable proxy of applied technological intent, and structured analysis of patent classification codes alongside textual

abstracts offers a tractable route toward filling this gap. This study addresses the following research questions:

- RQ1: Which core technology domains constitute the primary innovation landscape of GenAI-enabled digital transformation, as evidenced by international patent filings?
- RQ2: What latent topical clusters emerge from the unstructured patent abstract corpus, and what GenAI application archetypes do they represent?
- RQ3: How is GenAI mediating technological convergence across domains, and which domains function as structural convergence hubs within the co-occurrence network?

This study, thus, pursues three interrelated objectives to address the research questions. The first objective is descriptive: to characterise the core technology domain landscape of GenAI-enabled digital transformation by applying CPC-to-domain mapping and producing domain frequency tables, jurisdictional profiles, and year-over-year growth statistics. The second objective is topical: to identify and semantically label the latent application clusters constituting the GenAI-enabled DT innovation landscape through a keyword-rule topic assignment. The third objective is convergence-analytic: to construct a weighted technology co-occurrence network, compute betweenness centrality, weighted degree, and a composite convergence score for each domain node, and identify which domains function as structural hubs mediating the diffusion of GenAI capabilities into sector-specific enterprise applications.

The principal contributions of this study are threefold. It delivers the first CPC co-classification convergence network analysis specific to the GenAI-enabled DT patent corpus. It provides a 15-topic patent taxonomy applied to 7,623 post-2018 patents from ten jurisdictions and checked against a manually inspected sample. It reveals counterintuitive structural findings, including the disproportionately high betweenness centrality of Healthcare Informatics and the consolidation of convergence activity post-2021, that challenge assumptions derived from volume-based domain rankings alone.

2. EXISTING THEORIES AND PREVIOUS WORKS

This section surveys the three theoretical pillars that underpin the present study and positions prior empirical work against the gaps addressed herein. The first pillar is the conceptual and empirical literature on digital transformation (DT) as an organisational phenomenon. The second is the technology convergence tradition, which provides both theoretical grounding and methodological precedent for patent co-classification analysis. The third is the rapidly expanding body of applied GenAI research, spanning foundation model architectures, retrieval-augmented systems, and sector-specific deployments.

2.1. Theoretical foundations of digital transformation

The scholarly conceptualization of digital transformation has undergone substantial refinement, evolving from narrow IT-adoption studies toward systemic accounts of how digital technologies restructure value creation and competitive positioning. The most widely cited formulation is Vial's (2019) process-oriented framework, which defines DT as a process wherein digital technologies create disruptions that trigger strategic responses from organizations seeking to alter their value creation paths. The role of a specific technology is not to define the transformation but to constitute one input into a complex socio-technical system (Hanelt et al., 2021); this framing directly motivates the present study's focus on which domains GenAI enters and through which structural channels it propagates, rather than on its direct organizational outcomes.

Hanelt et al. (2021) extended this framework, showing firms that have progressed furthest in their transformations are embedded in platform ecosystems with increasingly indistinct boundaries, and DT is driven not by isolated technology adoption but by digital business ecosystem co-evolution. Their framework provides the organisational-level counterpart to the innovation supply-side analysis conducted here. The observation of Plekhanov et al. (2023) that firms most advanced in DT operate within platform ecosystems with blurred business boundaries is consistent with the convergence network topology the present study uncovers at the technology domain level: a densely connected graph anchored by a universal hub through which capabilities flow into sector-specific applications.

The general-purpose technology framework, formalised by Gambardella and Torrisi (1998) and subsequently applied to earlier AI waves, provides the macro-theoretical anchor for understanding

GenAI's structural role in DT. A GPT is defined by three properties: pervasiveness across sectors, continuous technical improvement, and innovational complementarities that stimulate derivative innovation disproportionate to the GPT itself. The framework predicts that GPTs characteristically generate co-invention activity, hybrid patents combining the GPT with sector-specific knowledge, which is precisely what the co-classification network in this study captures.

2.2. Technology convergence theory and patent co-classification methods

2.2.1. Convergence theory

Technology convergence, the blurring of boundaries between previously distinct technology sectors, has been conceptualized as both a firm-level challenge and a system-level innovation dynamic. Cross-industry convergence signals are detectable in publicly available patent data before they manifest in product markets (Curran et al., 2010). This temporal precedence property of patent co-classification, well-established in the technology forecasting literature, is a key justification for analysing GenAI convergence at the patent level rather than through market or product data, which would reflect a structurally lagged signal.

Jung et al. (2021) applied social network theory to IPC-coded patent data to characterize the structural role of Information and Communication Technology (ICT) in technology convergence networks. Technology convergence networks expand over time, and industrial actors engage in convergence earlier than universities or public research institutes. This temporal dimension informs the present study's interpretation of the post-2021 moderation in per-patent domain breadth: a transition from exploratory early-stage convergence, characteristic of the 2019–2021 GPT-3-era, when diverse actors experimented with GenAI across domains, to a consolidation phase consistent with established platform-centric development architectures.

2.2.2. Patent-Based convergence methodology

The methodological precedent for co-classification network analysis in technology convergence research is substantial. Previous studies (Zhu & Motohashi, 2022) employed Graph Convolutional Network (GCN) methodology to show the semantic content of patent descriptions alongside structural co-classification information. Afifuddin and Seo (2024) developed a voting classifier for convergence prediction from CPC co-occurrence networks, demonstrating that structural network features are the most predictive features for identifying future convergence events. Mohammadi et al. (2025) applied text mining and patent network modelling for citation-network-based community detection to detect the emergence of multimodal and healthcare sub-communities from 2020 onward. The present study is therefore complementary rather than redundant, addressing the technology-domain convergence dimension that citation network analysis does not directly measure.

2.3. GenAI: Foundation models, retrieval architectures, and sector applications

The technical foundation of the present corpus, the Transformer architecture introduced by Vaswani et al. (2017), has been surveyed from an application and industry perspective by Sindhu et al. (2024). The second dominant topic cluster in the present corpus, RAG & Knowledge-Augmented Systems at 26.1% of patents, is grounded in the architecture introduced by Lewis et al. (2020) at NeurIPS 2020, which coupled pre-trained language models with non-parametric external retrieval modules to address hallucination, knowledge staleness, and domain specificity. Gao et al. (2024) systematically surveyed the subsequent proliferation of RAG paradigms, documenting the rapid architectural sophistication of the field from 2020 to 2024. Fan et al. (2024) reviewed RAG specifically at the enterprise KDD 2024 venue, identifying enterprise knowledge management and document automation as the primary deployment contexts. Previous studies identify that state-of-the-art clinical GenAI systems simultaneously invoke medical imaging, signal processing, structured EHR reasoning, and model fine-tuning, precisely the multi-domain co-classification pattern that elevates Google's MedPaLM-2, cited in their review as achieving expert-level performance on medical licensing examinations.

2.4. Identification of research gaps

The existing literature establishes three co-existing bodies of knowledge that have not been previously integrated. Digital transformation theory provides organizational frameworks but lacks a supply-side, IP-based account of the technological architecture driving convergence. Technology convergence theory provides rigorous co-classification methodology and validated metrics but has not been applied to the GenAI-enabled DT context. GenAI application research provides detailed functional

taxonomies but relies on literature review or market survey data rather than structured patent analysis and does not produce structural convergence maps.

The present study occupies a position not previously filled in this literature: it is the first work to (i) apply CPC co-classification network analysis with domain-level betweenness centrality metrics specifically to the GenAI-enabled DT patent corpus, (ii) combine structural network analysis validated against enterprise DT application archetypes, and (iii) produce temporal convergence dynamics across the 2018–2025 window that straddles the pre-ChatGPT and post-GPT-4 regimes. In doing so, it operationalizes the convergence measurement within the theoretical DT framework, applied to the specific enabling technology that both frameworks identify as the current primary driver of enterprise-level structural change.

3. METHODOLOGY

3.1. Research methods and models

The study employs a mixed-method quantitative framework combining patent bibliometrics, keyword-rule topic modelling, and network-theoretic convergence analysis. The methodological architecture comprises three interlocking analytical layers, as summarised in Table 1 and detailed in the research process shown in Figure 1.

Table 1. Analytical framework

Layer	Method	Instrument	Output
Taxonomy	CPC Domain Mapping	CPC prefix-to-domain	Domain frequency & trend matrix
Topics	Keyword-Rule Topic Assignment	Rule taxonomy derived from GenAI literature; title + abstract text mining	Per-patent topic labels; topic distribution
Network	Co-occurrence Network Analysis	Weighted undirected graph; NetworkX; betweenness, convergence score	Convergence hub ranking; network

CPC Technology Domain Mapping: Each patent's CPC classification string was parsed into individual codes, and each code was mapped by its subclass prefix to one of 26 technology domains. For instance, codes beginning with G06N were assigned to AI & Machine Learning, G16H to Healthcare Informatics, and G05B to Control & Automation. Patents carrying codes from several prefixes received multiple domain assignments, which is the basis for the co-occurrence analysis in Layer 3. The 26 domains of this taxonomy constitute the node set of the convergence network. The complete prefix-to-domain mapping is available from the authors on request.

Keyword-Rule Topic Assignment: Fifteen topic classes were defined a priori from the GenAI application literature reviewed in Section 2, each representing a distinct application archetype, from LLM & Foundation Model Development to AI Agents & Autonomous Workflows. Each class is operationalised by a rule set of 8 to 14 domain-specific n-grams matched against the concatenated title and abstract of each patent. Because a patent may match the rules of more than one class, assignment follows a fixed first-match priority ordering in which the more semantically specific classes are tested before the broader ones, so that each patent receives exactly one label. Patents whose text matched no rule were assigned through a domain-informed fallback based on their CPC domains. The assignment was validated through a structured manual inspection: a stratified random sample of 50 patents was drawn from the three most populated classes (LLM & Foundation Model Development, RAG & Knowledge-Augmented Systems, and AI-Driven Business Process Automation) in proportion to class size, and two authors independently judged whether each assigned label was correct, resolving disagreements by discussion with reference to the rule set. Agreement with the assigned label exceeded 90% in each of the three classes. Patents bridging LLM development and RAG architectures were the most frequent edge cases; these informed the final priority ordering, so that patents with overlapping keyword signals are assigned to the most specific class rather than to the first rule matched. The complete rule sets and priority order are available from the authors on request.

Technology Convergence Network: A weighted undirected graph was constructed in which nodes represent technology domains and edges represent patents that simultaneously carry CPC codes from both endpoint domains. Edge weight equals the raw co-occurrence count. Three network-centrality metrics were computed for each node: (i) weighted degree, the sum of co-occurrence weights across all adjacent edges; (ii) betweenness centrality, the

proportion of shortest paths between all domain pairs that traverse the node, reflecting structural brokerage; and (iii) a composite convergence score defined as the equal-weighted mean of the normalised weighted degree and normalised betweenness centrality:

$$CS(v) = 0.5 \times norm(WD(v)) + 0.5 \times norm(BC(v)) \quad (1)$$

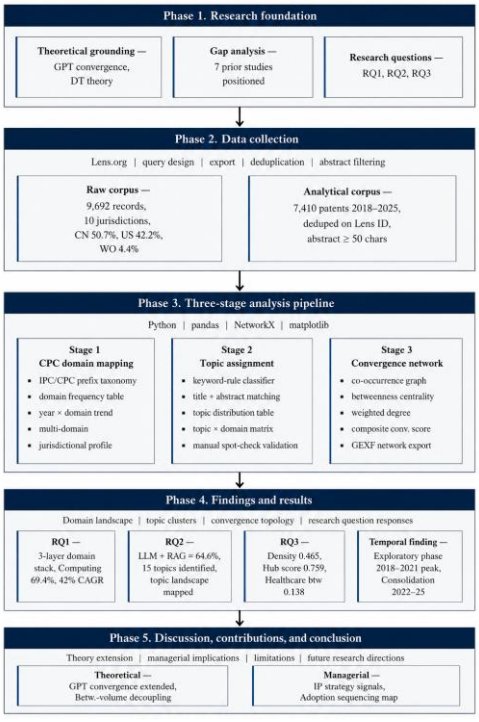


Figure 1. Research process

where $WD(v)$ is the weighted degree and $BC(v)$ is the betweenness centrality of node v, both normalised to [0, 1] over the full node set. Edges with co-occurrence counts below three were pruned to remove noise from incidental multi-domain coding.

The threshold of three was chosen so that a domain pair is connected only when it recurs in at least three patents, which removes edges produced by a single or incidental multi-domain coding while retaining the low-volume, sector-specific domains whose brokerage the analysis is designed to detect; a higher threshold would remove those peripheral nodes before their bridging position could be observed. Betweenness centrality was computed on the thresholded graph with NetworkX. The sensitivity of the centrality ranking to this threshold was not tested and is noted as a limitation in Section 6.

3.2. Data collection

Source and retrieval: The patent corpus was retrieved from Lens.org, an open-access scholarly and patent search platform. The query combined GenAI-specific terminology, including large language model, generative artificial intelligence, foundation model, and retrieval-augmented generation, applied to patent titles, abstracts, and claims. Records were exported in CSV format, including full metadata: Lens ID, title, abstract, publication date, jurisdiction, CPC and IPCR classifications, applicant, citation counts, and legal status. The dataset is summarized in Table 2.

Table 2. Dataset summary

Parameter	Value
Retrieved records (Lens.org export; deduplicated on Lens ID; abstract ≥ 50 characters)	9,692 (1987–2026)
Domain-mapping corpus (Figure 3)	9,692 records
Pre-2018 records (trend analysis only)	2,069
Topic-assignment corpus, 2018–2026 (Figure 4, Figure 5)	7,623 patents
Growth, multi-domain and network corpus, 2018–2025 (Figure 2, Figure 6, Figure 7, Table 3)	7,410 patents
Primary jurisdictions (2018–2025 corpus)	CN (50.7%), US (42.2%), WO (4.4%), KR (1.0%), EP (1.0%)
Topic classes	15
CPC domain categories (= network nodes)	26
Network edges	151 (min. co-occurrence ≥ 3)

Corpus construction: The Lens.org export, after deduplication on Lens ID and removal of records with absent or trivially short abstracts (fewer than 50 characters), comprised 9,692 records from ten jurisdictions, led by China (CN), the United States (US), the WIPO Patent Cooperation Treaty route (WO) and the European Patent Office (EP). This full corpus is the basis for the CPC domain mapping and

the domain frequency profile (Figure 3). Records published before 2018 (n = 2,069) were retained only for the long-run filing trend. The 7,623 patents published from 2018 onward form the topic-assignment corpus (Figure 4 and Figure 5). Because publication in 2026 was only partially observed at the time of retrieval, the annual growth, multi-domain rate, jurisdictional profile and convergence network statistics are computed on the 7,410 patents published between 2018 and 2025. Each figure and table states the corpus version based, and Table 2 summarises the flow.

Temporal scope: The 2018 lower bound was selected to capture the period coinciding with the emergence of transformer-based architectures, marked by the publication of Attention Is All You Need (Vaswani et al., 2017) in 2017, and the subsequent commercialization wave.

3.3. Analysis methods

The analysis proceeded in four sequential phases executed in Python (pandas, NetworkX, matplotlib). Phase 1 performed CPC parsing and domain frequency tabulation on the full corpus of 9,692 records. Phase 2 applied the keyword-rule topic classifier to the 7,623 patents published from 2018 onward and produced a per-patent topic label and a Topic × Domain co-occurrence matrix. Phase 3 constructed the convergence network from the 2018–2025 patents and computed all node-level metrics. Phase 4 computed the temporal statistics on the same 2018–2025 patents: a year-stratified multi-domain rate, defined as the mean number of distinct CPC domains per patent per publication year,

$$MDR(y) = (1/N_y)\ \Sigma\ D_i \quad (2)$$

where N_y is the total number of patents published in year y, D_i is the number of distinct CPC domains assigned to patent i, and the summation is taken over all patents published in year y.

This metric operationalizes the rate at which individual patents bridge multiple technology boundaries, a direct measure of convergence intensity over time. The Compound Annual Growth Rate (CAGR) of annual patent filings was computed across the period 2018–2025 to quantify the acceleration of GenAI-related innovation activity.

$$CAGR = (\text{End Value} / \text{Start Value})\text{^}(1/n) - 1 \quad (3)$$

where End Value is the filing count in 2025 (2,541), Start Value is the filing count in 2018 (218), and n is the number of years (7).

4. FINDINGS AND RESULTS

4.1. Collected data and corpus characteristics

Patent growth trajectory: Annual patent filings grew from 218 in 2018 to 2,541 in 2025, a 1,066% increase and a CAGR of 42.0% (Figure 2). The most pronounced single-year acceleration occurred between 2022 and 2023, when filings rose from 586 to 968 (+65.2%), coinciding with the mass-market release of large-scale generative models.

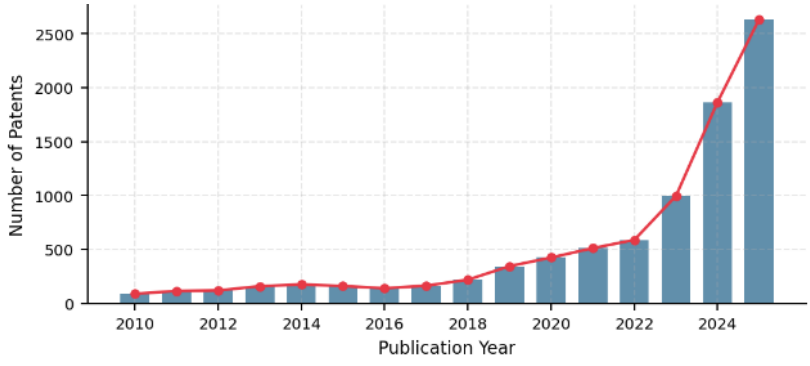


Figure 2. Annual GenAI-Related patent filings

Jurisdictional distribution: Within the 2018–2025 corpus (n = 7,410), China (CN) contributed 50.7% of patents (n = 3,755), followed by the United States (US) at 42.2% (n = 3,127), PCT-route filings (WO) at 4.4% (n = 328), the Republic of Korea at 1.0% (n = 73), and the European Patent Office at 1.0% (n = 72). China's leading share is consistent with the AI patenting trends documented by the OECD AI Policy Observatory (OECD, n.d.), while the near parity with US filings signals an intensely competitive bilateral innovation landscape in generative AI applications.

Domain distribution: Across the full corpus of 9,692 records, Computing & Information Systems (G06F)

was the dominant domain, present in 69.5% of records (n = 6,737), reflecting the infrastructural role of software systems in deploying GenAI capabilities. AI & Machine Learning (G06N) appeared in 36.5% (n = 3,542), indicating that a substantial share of filings involve not merely the application of pre-existing models but the development of model architectures. Business & Commerce (G06Q) ranked third at 25.2% (n = 2,445), indicating that enterprise-facing applications constitute a structurally significant innovation cluster. The remaining domains each account for under 13%: Networks & Communications (H04L) at 12.6% (n = 1,224), Speech & Audio Processing (G10L) at 6.4% (n = 625) and Computer Vision (G06V) at 5.9% (n = 571) are the principal secondary clusters, while Healthcare Informatics (G16H) appears in 3.1% (n = 300) and Control & Automation (G05B) in 2.6% (n = 253). Figure 3 reports the fourteen most frequent domains; all percentages use 9,692 as the denominator.

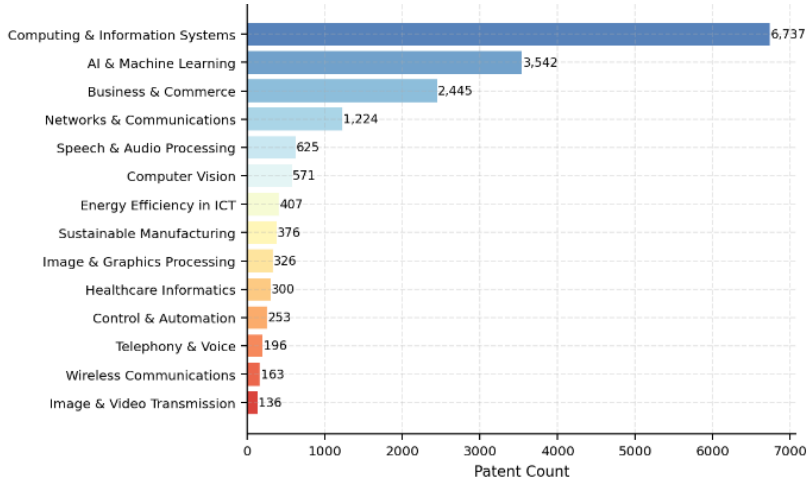


Figure 3. Core technology domains in the GenAI patent corpus

Multi-domain rate: Of the 7,410 patents published between 2018 and 2025, 64.1% (n = 4,749) carried CPC codes spanning two or more distinct technology domains, and 26.2% (n = 1,940) spanned three or more, evidencing pervasive multi-domain coding within the corpus. The mean number of domains per patent reached its peak at 2.34 in 2021, a year of rapid application diversification following the initial GPT-3 deployment cycle, and moderated to 1.65 by 2025 as filings increasingly concentrated within established platform stacks (Figure 7).

4.2. Data analysis

Topic Distribution: The keyword-rule classifier assigned each of the 7,623 patents published from 2018 onward to one of 15 GenAI application topics (Figure 4). Two topics account for 64.5% of all filings: LLM & Foundation Model Development (n = 2,927; 38.4%) and RAG & Knowledge-Augmented Systems (n = 1,991; 26.1%); adding AI-Driven Business Process Automation (n = 633; 8.3%) extends the concentration to 72.8% across the top three classes. This right-skewed distribution is consistent with the Pareto-like structure documented in general-purpose technology patent portfolios and indicates that the foundation model layer and its retrieval-augmented derivative are the dominant vectors of documented intellectual property activity in the current generation of GenAI innovation. Because the retrieval query itself contained the terms large language model, foundation model and retrieval-augmented generation, this share describes concentration within the GenAI patent corpus as delimited by the query and should not be read as an estimate over AI patenting in general (see Section 6).

The remaining 12 topics each account for 5.7% or less. Enterprise Knowledge Management (n = 40; 0.52%), AI Agents & Autonomous Workflows (n = 51; 0.67%) and Edge AI & Distributed Intelligence (n = 62; 0.81%) are the smallest clusters, consistent with the nascent commercialisation timelines of agentic, on-device and enterprise knowledge GenAI deployments at the time of data collection.

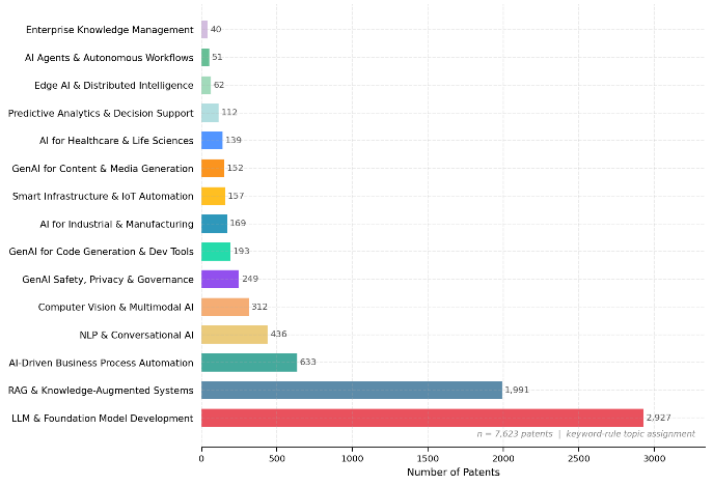


Figure 4. Topic distribution

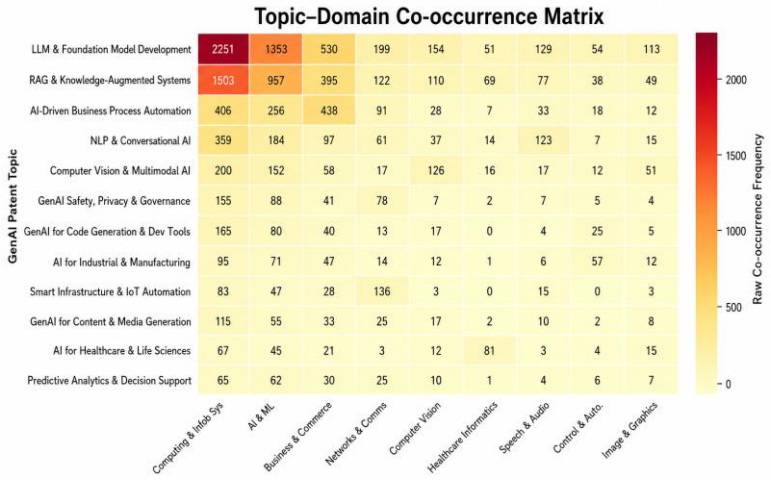


Figure 5. Topic–Technology domain co-occurrence

Topic–Domain Convergence Patterns: The Topic × Domain co-occurrence matrix (Figure 5; twelve largest topics and nine domains, 2018–2026 corpus) reveals structurally distinct clustering patterns. The two dominant topics both exhibit broad multi-domain footprints: LLM patents are co-classified under Computing & Information Systems (n = 2,251), AI & Machine Learning (n = 1,353) and Business & Commerce (n = 530), while RAG patents similarly span Computing & Information Systems (n = 1,503), AI & Machine Learning (n = 957) and Business & Commerce (n = 395). This horizontal co-classification pattern indicates that LLM and RAG innovation is not domain-siloed but integrated across the technology stack.

In contrast, AI for Healthcare & Life Sciences (n = 139) exhibits a concentrated profile anchored in Healthcare Informatics (n = 81), with Computing & Information Systems (n = 67) and AI & Machine Learning (n = 45) as secondary linkages, the most domain-specific pattern in the matrix. AI for Industrial & Manufacturing is co-classified most often with Computing & Information Systems (n = 95) and AI & Machine Learning (n = 71), and records the highest Control & Automation count of any topic (n = 57), reflecting the integration of GenAI reasoning into industrial control pipelines. Smart Infrastructure & IoT Automation shows a dominance of Networks & Communications (n = 136) over Computing & Information Systems (n = 83). Together with AI-Driven Business Process Automation, anchored in Business & Commerce (n = 438) ahead of Computing & Information Systems (n = 406), and the healthcare topic, these are the only topics in which a non-computing domain is the primary classification.

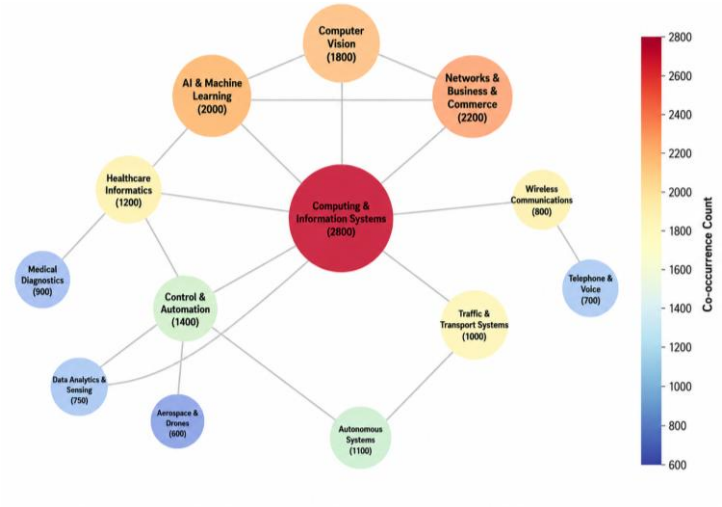


Figure 6. Technology convergence network

Technology Convergence Network: The co-occurrence network built on the 2018–2025 corpus comprised 26 nodes and 151 weighted edges after application of the minimum co-occurrence threshold of three patents. Figure 6 gives a simplified view of the principal hubs; node-level metrics for the ten highest-scoring domains are reported in Table 3.

Computing & Information Systems emerged as the primary convergence hub, recording a composite convergence score of 0.759, a weighted degree of 6,540 and a d egree of 23, connecting to all but two of the other domain nodes. Its betweenness centrality of 0.081 indicates that it lies on a substantial share of the shortest inter-domain paths, positioning it as the primary technological intermediary through which GenAI capabilities propagate into sector-specific applications. AI & Machine Learning recorded the second-highest weighted degree (5,244) and an equal degree (23), but a considerably lower betweenness centrality (0.008), indicating a role as a co-innovator that frequently appears alongside other domains rather than as a structural bridge.

A counter-intuitive finding concerns Healthcare Informatics, which achieved the third-highest convergence score (0.476) despite modest patent volume (n = 300; 3.1% of the full corpus). Its betweenness centrality of 0.138, the second highest in the network, indicates that healthcare-related patents disproportionately bridge otherwise weakly connected technology clusters, consistent with the interdisciplinary nature of clinical AI systems that combine medical imaging (G06V), signal processing (G10L) and model training (G06N). Autonomous Systems recorded the highest betweenness centrality (0.156), reflecting the cross-domain character of autonomous navigation patents that span control, computer vision and communications (Table 3).

Table 3. Centrality and convergence scores (top ten domains; 2018–2025 corpus)

Rank	Domain	Degree	Betweenness	Conv. Score
1	Computing & Information Systems	23	0.081	0.759
2	Autonomous Systems	14	0.156	0.507
3	Healthcare Informatics	14	0.138	0.476
4	Aerospace & Drones	9	0.134	0.430
5	AI & Machine Learning	23	0.008	0.428
6	Data Analytics & Sensing	5	0.128	0.410
7	Business & Commerce	21	0.041	0.355
8	Wireless Communications	15	0.095	0.323
9	Networks & Communications	20	0.038	0.229
10	Sustainable Manufacturing	13	0.054	0.222

In sum, patent activity is concentrated in Computing & Information Systems (69.5%) and AI & Machine Learning (36.5%), yet diffusion into sector-specific domains is evident. Healthcare Informatics appears in 3.1% of records and Control & Automation in 2.6%, and the 139 patents in the AI for Healthcare & Life Sciences topic, together with the high betweenness of Healthcare Informatics, point to convergence that is qualitatively significant rather than large in volume. The multi-domain rate peaked at 2.34 domains per patent in 2021, and Computing & Information Systems holds the highest convergence score (0.759) by a substantial margin, consistent with a bridging role for the infrastructure layer.

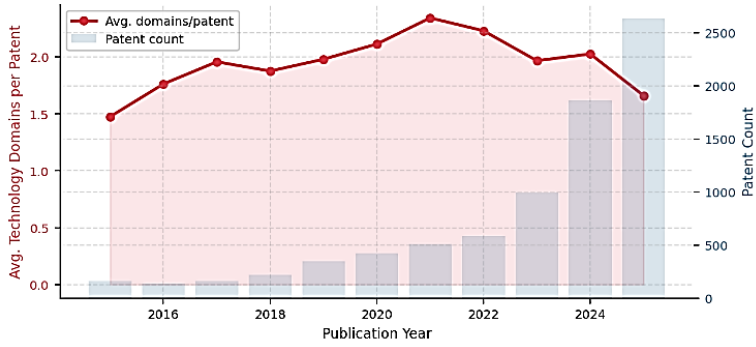


Figure 7. Temporal technology convergence trend: Mean CPC domains per patent and annual patent count

Two results warrant emphasis. First, LLM development and RAG systems together constitute nearly two-thirds of the topic-assignment corpus, establishing these as the foundational technology layer of the current innovation cycle. Second, the post-2022 acceleration in filing volume coincides with a moderation in per-patent domain breadth (Figure 7), a pattern consistent with a transition from exploratory cross-domain experimentation toward platform-centric development, although the analysis is descriptive and alternative explanations cannot be excluded (Section 6).

5. DISCUSSION

5.1. Answering research questions

5.1.1. RQ1: Core technology domain landscape

The evidence answers RQ1: the core technology domain landscape of GenAI-enabled DT is dominated by Computing & Information Systems (G06F; 69.5% of the corpus), AI & Machine Learning (G06N; 36.5%), and Business & Commerce (G06Q; 25.2%). These three domains form a layered architecture: G06F as the infrastructure layer on which GenAI capabilities are deployed, G06N as the model layer where those capabilities are developed and fine-tuned, and G06Q as the application layer through which they are monetised and integrated into processes. This three-layer stack is the dominant structural motif of the GenAI-enabled DT innovation system. Eleven further domains each account for between 1.4% and 12.6% of records, collectively representing the sector-specific diffusion frontier of GenAI. The 65.2% single-year surge in filings between 2022 and 2023, from 586 to 968 patents, coincides with the mass-market release of GPT-3.5-turbo and the public availability of associated APIs, the period in which GenAI became deployable at enterprise scale. This inflexion is consistent with the S-curve diffusion dynamics characteristic of general-purpose technologies.

5.1.2. RQ2: Latent application topic clusters

The keyword-rule topic assignment answers RQ2. The corpus exhibits a highly right-skewed distribution: the top two topics, LLM & Foundation Model Development (38.4%) and RAG & Knowledge-Augmented Systems (26.1%), collectively account for 64.5% of all patent filings, and the addition of AI-Driven Business Process Automation (8.3%) extends concentration to 72.8% across the top three classes. This concentration is measured within a corpus retrieved with LLM-, foundation-model- and RAG-related terms, so the share of the two leading classes is partly a property of the retrieval frame; within that frame, however, managers who benchmark competitive R&D positioning exclusively against application-domain peers risk missing the IP concentration at the model and retrieval infrastructure layers, where the largest volume of proprietary capability is being filed. The

bottom three topics, Enterprise Knowledge Management (0.52%), AI Agents & Autonomous Workflows (0.67%) and Edge AI & Distributed Intelligence (0.81%), are collectively represented by only 153 patents. The disparity between their current patent share and their forecast commercial significance is notable, since both McKinsey & Company (2024) and WIPO (2024) project agentic AI and edge deployment as primary growth vectors through 2030.

5.1.3. RQ3: Technology convergence network structure

The co-occurrence network answers RQ3 at both topological and node levels. At the network level, the 26-node, 151-edge graph has a density of 0.465, i.e., 46.5% of all possible domain pairs are connected by at least three co-classified patents, indicating that GenAI-enabled DT is not an isolated technology phenomenon but a pervasive convergence pattern that reaches nearly half of all domain pairs in the taxonomy at measurable intensity. The average node degree of 11.6 further indicates that each domain co-occurs with nearly half of all other domains. At the node level, Computing & Information Systems achieves a composite convergence score of 0.759, a weighted degree of 6,540 and a degree of 23, connecting to all but two of the other domain nodes, consistent with a role as the universal integration substrate of GenAI-enabled DT. Its betweenness centrality of 0.081 reflects a structural distinction between volume-based centrality and brokerage centrality. The domains with the highest betweenness, Autonomous Systems (0.156), Healthcare Informatics (0.138), Aerospace & Drones (0.134) and Data Analytics & Sensing (0.128), are not high-volume nodes but structural bridges connecting peripherally positioned domain clusters that share few other pathways through the network.

5.2. Theoretical contributions

This study makes four contributions to the intersection of digital transformation theory, technology convergence theory, and GenAI research. First, it provides patent-level evidence that a Transformer-derived GPT produces CPC co-classification convergence patterns consistent with Curran and Leker's (2011) anticipatory convergence framework, extending the framework's applicability from manufacturing-proximate GPTs (electrification, ICT) to cognitive-computation GPTs. The density of the observed convergence network (0.465) exceeds the values of 0.12 to 0.28 reported by Jung et al. (2021) for ICT convergence networks; this is consistent with a more pervasive convergence pattern than prior GPTs at a comparable stage, although density also depends on the granularity of the domain taxonomy and the pruning threshold, so the comparison is indicative rather than conclusive. Second, the study introduces the betweenness-volume decoupling as a convergence diagnostic: a node with disproportionately high betweenness centrality relative to its co-occurrence volume represents a structural seam rather than a hub. This diagnostic had not been explicitly operationalized in the patent convergence literature and constitutes a methodological contribution applicable beyond the GenAI context. Third, the two-phase pattern observed, exploratory (2018–2021) followed by consolidation (2022–2025), is descriptive evidence consistent with phase-structured convergence in the GenAI domain, and suggests that theoretical models derived from ICT-era convergence may require revision for cognitive-AI GPTs, whose deployment cycles are compressed relative to electrification or the internet and whose consolidation phase may therefore be discernible within a single decade. Fourth, the present study maps the technology disruption layer of this process at the domain level, identifying which technology domains occupy bridging positions (via convergence score and betweenness) and which remain structurally peripheral (Semiconductor Hardware and Switching & Routing, which record zero betweenness and are isolated from the convergence network).

5.3. Managerial implications

The findings carry implications for industry and decision makers. The concentration of 64.5% of GenAI DT patent activity in LLM development and RAG systems means that organizations building GenAI capabilities primarily through application-layer patents, filing in domain-specific application topics while licensing model and retrieval infrastructure, are accumulating competitively exposed IP portfolios. Organizations seeking durable protection should evaluate whether their IP strategy addresses the infrastructure layer (G06F, G06N) and the retrieval-augmented architecture layer alongside application-specific filings. The 42.0% CAGR of the corpus (2018–2025) further implies that the window for establishing foundational IP positions is contracting: each year's delay in filing corresponds to a denser prior-art environment. The betweenness-high, volume-low nodes in the convergence network, Autonomous

Systems (betweenness 0.156, CS 0.507), Healthcare Informatics (0.138, CS 0.476) and Data Analytics & Sensing (0.128, CS 0.410), mark the domains in which GenAI-enabled competitive disruption is structurally positioned to occur. Healthcare Informatics illustrates the mechanism: clinical systems such as Google's MedPaLM-2 combine medical imaging, signal processing, structured EHR reasoning and model fine-tuning within a single system, and this multi-domain footprint is what places Healthcare Informatics in its bridging position. Autonomous Systems shows the same pattern, with navigation patents spanning control theory, computer vision and communications engineering. Firms whose competitive position depends on clinical AI, autonomous navigation or sensor-driven manufacturing intelligence should treat these betweenness signals as forward-looking indicators rather than current-state assessments and initiate cross-domain capability building accordingly.

The three smallest topic clusters, Enterprise Knowledge Management (40 patents), AI Agents & Autonomous Workflows (51) and Edge AI & Distributed Intelligence (62), represent the most asymmetric risk-return profiles in the current IP landscape. Their present scarcity reflects nascent commercialization rather than limited value: Microsoft's Copilot agent architecture and Apple's on-device inference work indicate that early IP positions are being established while filing density is still low. Organizations that establish positions in these clusters during the low-density phase may capture option value proportionate to the gap between current and anticipated filing rates, although the present data cannot indicate when or how rapidly these clusters will grow.

6. CONCLUSION

This study sets out to provide the first patent bibliometric and co-classification network analysis of the technology domain structure and convergence topology underlying GenAI-enabled digital transformation. Applied to a Lens.org corpus of 9,692 records from ten jurisdictions, of which 7,623 patents published from 2018 onward form the analytical corpus, with annual filings growing at a CAGR of 42.0% between 2018 and 2025 and a structural break in 2022–2023 (+65.2% in a single year), the three-stage pipeline of CPC domain mapping, keyword-rule topic assignment, and weighted co-occurrence network analysis produces a set of findings that are empirically specific, theoretically grounded, and managerially actionable.

The principal findings, collectively, advance the field's understanding of GenAI-enabled DT along three dimensions. At the structural dimension, the innovation landscape is organised around a three-layer stack: Computing & Information Systems as infrastructure, AI & Machine Learning as model layer, Business & Commerce as application interface, with a network topology of density 0.465 that is substantially denser than ICT-era convergence networks, consistent with a more pervasive convergence pattern. At the topical dimension, the IP landscape is dominated by LLM development and RAG systems at 64.5%, with sector-specific applications- healthcare, industrial automation, IoT, content generation- each accounting for under 2.5% individually, indicating that the current innovation cycle's energy is concentrated at the foundational and retrieval-augmentation layers rather than at the sector-specific application surface. At the convergence-dynamic dimension, the study observes a two-phase temporal pattern, exploratory convergence (2018–2021, peak multi-domain rate 2.34) followed by consolidation (2022–2025, rate declining to 1.65), and isolates Healthcare Informatics and Autonomous Systems as betweenness-high structural seams that are candidate sites for next-cycle convergence events despite their currently modest patent volumes.

The findings provide the DT research community with an IP supply-side structural map that complements the organizational-process frameworks of Vial (2019) and Hanelt et al. (2021), grounding the technology disruption layer of those frameworks in patent data rather than case study inference, and offer technology managers a convergence diagnostic, the betweenness-volume decoupling, that is not available from volume-ranked domain analyses alone. The limitation of this study is that the keyword-rule assignment uses predefined n-gram rules with first-match priority rather than probabilistic topic models such as LDA or BERTopic, which constrains classification for patents at the boundary between classes; its manual validation covered 50 patents from the three largest classes, so per-class precision and recall for the twelve smaller classes remain unmeasured. Integrating citation-based influence metrics with co-classification convergence scores, testing the temporal pattern formally, and extending the

manual validation to all fifteen classes are the immediate directions for future work.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Afifuddin, M., & Seo, W. (2024). Predictive modeling for technology convergence: A patent data-driven approach through technology topic networks. *Computers & Industrial Engineering*, *188*, 109909. https://doi.org/10.1016/j.cie.2024.109909

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E.,… & Zhang, J. (2023). *Sparks of artificial general intelligence: Early experiments with GPT-4*. arXiv preprint arXiv:2303.12712. https://doi.org/10.48550/arXiv.2303.12712

Chkirbene, Z., Hamila, R., Gouissem, A., & Devrim, U. (2024). Large language models (LLM) in industry: A survey of applications, challenges, and trends. In *Proceedings of the 2024 IEEE 21st International Conference on Smart Communities: Improving Quality of Life Using AI, Robotics and IoT (HONET)* (pp. 229–234). IEEE. https://doi.org/10.1109/HONET63146.2024.10822885

Curran, C.-S., Bröring, S., & Leker, J. (2010). Anticipating converging industries using publicly available data. *Technological Forecasting and Social Change*, *77*(3), *385–395. https://doi.org/10.1016/j.techfore.2009.10.002*

Curran, C.-S., & Leker, J. (2011). Patent indicators for monitoring convergence: Examples from NFF and ICT. *Technological Forecasting and Social Change*, *78*(2), *256–273. https://doi.org/10.1016/j.techfore.2010.06.021*

Fan, W., Ding, Y., Ning, L., Wang, S., Li, H., Yin, D., Chua, T. S., & Li, Q. (2024). A survey on RAG meeting LLMs: Towards retrieval-augmented large language models. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)* (pp. 6491–6501). ACM. https://doi.org/10.1145/3637528.3671470

Gambardella, A., & Torrisi, S. (1998). Does technological convergence imply convergence in markets? Evidence from the electronics industry. *Research Policy*, *27*(5), *445–463. https://doi.org/10.1016/S0048-7333(98)00062-6*

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2024). *Retrieval-augmented generation for large language models: A survey*. arXiv preprint arXiv:2312.10997. https://doi.org/10.48550/arXiv.2312.10997

Hanelt, A., Bohnsack, R., Marz, D., & Antunes Marante, C. (2021). A systematic review of the literature on digital transformation: Insights and implications for strategy and organizational change. *Journal of Management Studies*, *58*(5), 1159–1197. https://doi.org/10.1111/joms.12639

Jung, S., Kim, K., & Lee, C. (2021). The nature of ICT in technology convergence: A knowledge-based network analysis. *PLOS ONE*, *16*(7), e0254424. https://doi.org/10.1371/journal.pone.0254424

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems* (Vol. 33, pp. 9459–9474). https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

Ma, D., & Wu, W. (2024). Does artificial intelligence drive technology convergence? Evidence from Chinese manufacturing companies. *Technology in Society*, *79*, 102715. https://doi.org/10.1016/j.techsoc.2024.102715

McKinsey & Company. (2023). *The state of AI in 2023: Generative AI's breakout year.* McKinsey Global Survey on AI. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year

McKinsey & Company. (2024). *The state of AI in early 2024: Gen AI adoption spikes and starts to generate value. McKinsey Global Survey on AI.* https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024

Mohammadi, N., Heidary Dahooie, J., Bengari, A. A., & Rahimi, A. (2025). Mapping the technological evolution of generative AI: A patent network analysis. *Scientific Reports*, *15*(1), 45690. https://doi.org/10.1038/s41598-025-26810-7

OECD. (n.d.). *OECD AI Policy Observatory*. Organisation for Economic Co-operation and Development. https://oecd.ai

Parteka, A., & Kordalska, A. (2023). Artificial intelligence and productivity: Global evidence from AI patent and bibliometric data. *Technovation*, *125*, 102764. https://doi.org/10.1016/j.technovation.2023.102764

Plekhanov, D., Franke, H., & Netland, T. H. (2023). Digital transformation: A review and research agenda. *European Management Journal*, *41*(6), 821–844. https://doi.org/10.1016/j.emj.2022.09.007

Sindhu, B., Prathamesh, R. P., Sameera, M. B., & KumaraSwamy, S. (2024). The evolution of large language model: Models, applications and challenges. In *Proceedings of the 2024 International Conference on Current Trends in Advanced Computing (ICCTAC)* (pp. 1–8). IEEE. https://doi.org/10.1109/ICCTAC61556.2024.10581180

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 5998–6008). https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, *28*(2), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003

World Intellectual Property Organization. (2024). *Generative artificial intelligence: Patent landscape report.* WIPO. https://doi.org/10.34667/tind.49740

Zhu, C., & Motohashi, K. (2022). Identifying the technology convergence using patent text information: A graph convolutional networks (GCN)-based approach. *Technological Forecasting and Social Change*, 176, 121477. https://doi.org/10.1016/j.techfore.2022.121477