Mapping the Twin Transition: Technology Communities and Green Business Penetration

Levan Bzhalava
Bratislava University of Economics and Business,
Dolnozemská cesta 1, 852 35, Bratislava, Slovak Republic
e-mail: levani.bzhalava@euba.sk, phone: +421 904 073 080

Jari Kaivo-oja
Finland Futures Research Centre, Turku School of Economics, University of Turku, Unit of Tampere,
Hämeenkatu 15 B 11. 33100 Tampere, Finland.
e-mail: jari.kaivo-oja@utu.fi, phone: +358-41-753 0244

Miroslav Šipikal
Bratislava University of Economics and Business,
Dolnozemská cesta 1, 852 35, Bratislava, Slovak Republic
e-mail: miroslav.sipikal@euba.sk

Adesuwa Omorede
School of Innovation, Design and Engineering, Mälardalen University
Hamngatan 15, 632 17 Eskilstuna, Sweden
e-mail: adesuwa.omorede@mdu.se

Abstract: The twin transition—the convergence of digitalization and environmental sustainability—is reshaping entrepreneurial activity, yet the patterns through which digital technologies are adopted across green business domains remain poorly understood. This study examines 6,580 green ventures founded between 2016 and 2025 using Crunchbase data. Drawing on digital artifact theory and layered modular architecture, we combine technology co-occurrence networks, community detection, venture-level classification, and topic modeling to map digital technology penetration across green business domains. The findings reveal substantial heterogeneity in how digital technologies are integrated into green entrepreneurship. AI and machine learning demonstrate the broadest penetration across green business domains, consistent with their general-purpose character, whereas robotics and 3D technologies are more strongly concentrated in urban sustainability and industrial operations. Enterprise software and data analytics are particularly prominent in corporate sustainability and supply-chain activities, while IoT and sensor technologies are strongly associated with buildings and energy infrastructure. These patterns indicate that digitalization in green entrepreneurship is not uniform but follows distinct technology–business configurations, with some technologies diffusing broadly and others remaining closely tied to specific sustainability domains. The study provides a data-driven mapping of the technological architecture of the twin transition and shows how digital capabilities intersect with different forms of green entrepreneurial activity.

Keywords: twin transition; green entrepreneurship; digital technologies; technology communities; green business; network analysis; topic modeling

1. Introduction

The transition toward a more sustainable economy is increasingly intertwined with rapid advances in digital technologies. Artificial intelligence (AI), cloud computing, data analytics, the Internet of Things (IoT), robotics, and blockchain are becoming embedded in business activities ranging from energy management and industrial production to mobility, agriculture, circular economy solutions, and carbon accounting. This simultaneous transformation is commonly described as the twin transition, reflecting the growing interdependence between digital transformation and the transition toward environmental sustainability (Bianchini et al., 2023; Calvino et al., 2025). Rather than constituting two independent processes, the digital and green transitions can reinforce one another when digital technologies enable firms to monitor environmental impacts, optimize resource use, improve energy efficiency, coordinate complex systems, and develop sustainability-oriented products and business models (Bianchini et al., 2023; Siedschlag et al., 2024; Ben Youssef, 2025).

Entrepreneurial ventures represent an important setting in which this convergence takes place. Sustainability transitions involve not only technological change but also transformations in industries, organizations, markets, institutions, and patterns of innovation (Köhler et al., 2019). New ventures can contribute to these processes by introducing novel technologies, products, services, and business models and by exploiting opportunities created by emerging sustainability challenges (Bzhalava et al., 2026). At the same time, entrepreneurial firms constitute important vehicles for the commercialization and diffusion of emerging digital technologies. The intersection of these processes creates distinctive digitalized green ventures: firms whose environmental orientation is combined with identifiable digital technological capabilities. Examining such ventures therefore provides a micro-level perspective on how the twin transition materializes through concrete combinations of digital technologies and sustainability-oriented business activities.

Existing research provides strong theoretical foundations for understanding sustainability transitions and digital transformation separately. Sustainability-transition research has examined technological innovation, institutional change, emerging industries, socio-technical systems, and the processes through which new technologies and practices challenge established regimes (Köhler et al., 2019). Digital innovation research, meanwhile, emphasizes that digital technologies differ fundamentally from conventional technologies because of properties such as programmability, data-based operation, generativity, and their ability to be recombined across products and services. The layered modular architecture perspective, for example, conceptualizes digital innovation as occurring through loosely coupled device, network, service, and content layers, allowing digital capabilities to be recombined across organizational and sectoral boundaries (Yoo et al., 2010).

A growing literature has begun to connect these streams through the concept of the twin transition. Recent studies examine relationships between digitalization and environmental innovation, firm-level sustainability practices, regional emissions, and sustainable innovation (Bianchini et al., 2023; Siedschlag et al., 2024; Aklilu et al., 2025; Calvino et al., 2025). This research indicates that digital technologies can support environmental innovation and sustainability, but also that their effects vary across technologies, sectors, firms, and regions. For example, firm-level evidence suggests complementarities between the adoption of technologies such as AI and IoT and environmental sustainability practices (Siedschlag et al., 2024), while regional evidence demonstrates that digital and environmental technological capabilities can interact in more complex ways and that the environmental consequences of

digital technologies differ substantially across technological categories (Bianchini et al., 2023). Recent work therefore increasingly emphasizes the need to understand the specific mechanisms and technological configurations underlying the twin transition rather than assuming an automatically positive relationship between digitalization and sustainability (Aklilu et al., 2025; Calvino et al., 2025).

Despite these advances, comparatively little is known about how this intersection is structured within green entrepreneurial ventures. In particular, there remains limited systematic evidence regarding which digital technologies are actually embedded in green ventures, how these technologies combine into broader technological communities, and how those communities penetrate different areas of green business activity. Existing studies frequently examine digitalization through aggregate indicators, individual technologies, patent classifications, or firm-level measures of technology adoption (Bianchini et al., 2023; Siedschlag et al., 2024; Aklilu et al., 2025). Such approaches provide important evidence on the relationship between digital and green transformations but offer less insight into the internal technological structure of digitalized green entrepreneurship.

This distinction is important because digital technologies should not be treated as a homogeneous category. AI and machine learning, for example, can potentially be applied across a wide variety of sectors and organizational functions, whereas robotics or sensor-based systems may require closer integration with particular physical assets and production environments. Recent conceptual work on the twin transition similarly emphasizes that digital transformation involves technologies with different characteristics and channels of influence on environmental outcomes (Calvino et al., 2025). Cloud and enterprise software may support sustainability reporting and supply-chain coordination, while IoT technologies can facilitate real-time monitoring and control of buildings, energy systems, and environmental conditions. Digital technologies may therefore differ not only in their prevalence among green ventures but also in their breadth of penetration across sustainability domains. Some may function as horizontal or general-purpose capabilities, while others may exhibit relatively narrow sectoral or application-specific patterns.

A related methodological problem concerns what should count as a digital technology in the first place. Entrepreneurial databases frequently combine technologies, industrial sectors, infrastructures, business practices, and end-products within the same classification systems. Categories such as artificial intelligence or cloud computing coexist with broader categories such as information technology, hardware, autonomous vehicles, or cryptocurrency. Yet digital innovation theory suggests that different layers of a digital system should be analytically distinguished. Yoo et al. (2010), for example, differentiate device, network, service, and content layers within a layered modular architecture. Treating all technology-related categories as equivalent indicators of digitalization may therefore conflate core digital artifacts with enabling infrastructures or downstream applications. A theoretically informed classification is required before the penetration of digital technologies into green business can be meaningfully analyzed.

Building on Digital Artifact Theory (Ekbia, 2008), the layered modular architecture perspective (Yoo et al., 2010), and Berger et al.'s (2018) taxonomy of digital technologies, this study distinguishes core digital technologies from digital infrastructure and enabling technologies and from applications and practices of digital technologies. Core technologies are defined as programmable and data-driven digital artifacts and provide the basis for identifying digitalized green ventures. This approach produces a more restrictive measure of digitalization than

classifications based on broad ICT-related categories and allows the technological foundations of the twin transition to be examined more systematically.

The study subsequently moves beyond individual technologies by conceptualizing digitalization as a system of technology communities. Technologies rarely occur independently within entrepreneurial ventures. AI may be combined with machine learning and computer vision; software may coexist with SaaS and cloud computing; and IoT applications may integrate sensing and cloud-based data services. Such technological combinations are consistent with the recombinable and layered nature of digital innovation described by Yoo et al. (2010). Patterns of technological complementarity can be empirically captured through co-occurrence networks, in which technologies that frequently appear together form identifiable communities. A community-based perspective therefore offers an intermediate level of analysis between individual technologies and the broad concept of digitalization, allowing the internal technological architecture of digital green entrepreneurship to become visible.

At the same time, identifying technology communities alone does not reveal where these technologies are being deployed. To examine the green business penetration of digital technologies, this study combines technology-community analysis with topic modeling of venture business descriptions. Green business penetration is understood here as the distribution of digital technology communities across distinct sustainability-oriented business domains. This enables the technological dimension of ventures—*what digital capabilities they possess*—to be linked with their substantive sustainability orientation—*what environmental or green business problems they address*. Comparing these distributions also makes it possible to distinguish relatively general-purpose technology communities from those displaying stronger sectoral specialization.

The analysis identifies six technology communities: AI & Machine Learning, Robotics & 3D Technologies, Enterprise Software & Cloud Computing, Data & Analytics, Blockchain & Web3, and IoT & Sensor Systems. Only 20.6% of ventures in the overall green venture population exhibit identifiable adoption of the core digital technologies included in the analysis, indicating that the digital dimension of the twin transition remains concentrated within a subset of green entrepreneurship. Among digitalized ventures, Enterprise Software & Cloud Computing and AI & Machine Learning dominate the technological landscape. The analysis further identifies nine green business domains and demonstrates substantial differences in the penetration profiles of the six technology communities. AI & Machine Learning exhibits the broadest distribution across sustainability domains, whereas Robotics & 3D Technologies is substantially more specialized. Enterprise Software & Cloud Computing and Data & Analytics are particularly associated with Corporate Sustainability and Supply Chain Management, while IoT & Sensor Systems exhibits a stronger orientation toward Buildings and Energy Infrastructure.

2. Literature Review

2.1. Green Entrepreneurship and Sustainability Transitions

The increasing urgency of climate change, resource depletion, biodiversity loss, and other environmental challenges has intensified scholarly interest in entrepreneurship as a means of advancing sustainability-oriented economic transformation. Green entrepreneurship refers to entrepreneurial activities that seek to create economic value while simultaneously addressing

environmental objectives through new products, technologies, services, production processes, or business models (Dean & McMullen, 2007; Cohen & Winn, 2007; Schaltegger & Wagner, 2011).

Green entrepreneurship overlaps with several related concepts, including sustainable entrepreneurship, environmental entrepreneurship, ecopreneurship, and circular entrepreneurship. Although the boundaries among these concepts remain debated, they share a common emphasis on entrepreneurial responses to environmental and broader societal challenges (Gast et al., 2017). Recent reviews likewise highlight considerable conceptual fragmentation across these forms of sustainability-oriented entrepreneurship, indicating that the field has yet to reach a consensus on a unified definition of green entrepreneurship (Carst, 2026).

From the perspective of sustainability transitions theory, green entrepreneurship can contribute to broader systemic transformations by introducing alternatives to established technologies, practices, and business models (Bzhalava et al., 2026). Sustainability transitions involve long-term and multidimensional changes in socio-technical systems, including energy, mobility, food, buildings, and industrial production (Geels, 2002; Markard et al., 2012; Köhler et al., 2019). Because incumbent systems are characterized by established technologies, infrastructures, institutions, regulations, and user practices, transitions depend partly on the emergence, development, and diffusion of alternatives that can challenge, disrupt, or reconfigure existing systems.

Green ventures can play an important role in this process. Compared with established firms, new ventures may face fewer organizational constraints arising from existing assets, competencies, and business models, potentially enabling them to explore emerging technological and market opportunities more flexibly (Bzhalava et al., 2026). Green ventures can commercialize renewable energy technologies, develop resource-efficient production systems, facilitate circular business models, introduce environmental monitoring solutions, and provide new services related to climate mitigation and adaptation. However, the technological foundations of green entrepreneurship are also evolving. Sustainability-oriented ventures increasingly operate in an economic environment shaped not only by environmental transformation but also by rapid digitalization. Understanding contemporary green entrepreneurship therefore requires attention to how sustainability-oriented entrepreneurial activity intersects with ongoing digital technological change.

2.2. From Parallel Transformations to the Twin Transition

Digitalization and environmental sustainability have often been examined through distinct research traditions, with digital transformation research focusing primarily on the diffusion and application of information and communication technologies, software, platforms, data, and automation, while sustainability-transition research has focused on decarbonization, renewable energy, resource efficiency, circularity, and broader changes in socio-technical systems. Increasingly, however, scholars have emphasized the interdependence between digital and green transformations, giving rise to the concept of the twin transition (Bianchini et al., 2023; Rehman et al., 2023; Calvino et al., 2025).

The twin transition refers to the interconnected development of digital and green transformations, in which the two processes can potentially reinforce and accelerate one another. Digital technologies can support environmental innovation and sustainability

transitions by enabling greater resource efficiency, improved monitoring, data-driven decision-making, process optimization, and the development of new sustainable products and services. At the same time, environmental challenges and sustainability objectives can stimulate demand for new digital technologies and applications (Bianchini et al., 2023; Calvino et al., 2025). Recent research similarly conceptualizes twin transitions as parallel yet mutually reinforcing processes that can contribute to sustainable competitiveness at the organizational level (Tabares et al., 2025).

Empirical evidence increasingly supports complementarities between digitalization and environmental innovation. At the firm level, Siedschlag et al. (2024) find evidence that digital technology adoption and environmental sustainability practices can reinforce one another, particularly in relation to technologies such as AI and IoT. At the regional level, Bianchini et al. (2023) show that digital and environmental technological capabilities interact in shaping environmental outcomes. Patent-based research similarly indicates relationships between digital innovation and subsequent environmental innovation, while demonstrating that the strength of these relationships varies across digital technology categories (Aklilu et al., 2025). This observation has important implications for twin-transition research. It suggests that treating digitalization as a single explanatory variable may obscure substantial technological heterogeneity. The relevant question is therefore not simply whether green firms are becoming digitalized, but which digital technologies are involved and how they connect to specific forms of green economic activity.

2.3. Digital Technologies as Heterogeneous Technological Capabilities

A central premise of this study is that digital technologies should not be treated as a homogeneous category. AI, software, cloud computing, blockchain, the Internet of Things (IoT), sensors, robotics, and data analytics differ in their technological properties, infrastructural requirements, complementary assets, and potential applications. These differences are important for understanding how digital technologies contribute to green entrepreneurship and the twin transition.

Digital innovation theory helps explain this heterogeneity. Yoo et al. (2010) conceptualize digital innovation as a layered architecture comprising devices, networks, services, and content. Because digital functionality can be recombined across these layers, digital technologies are highly flexible and generative. Their reprogrammability and reliance on digitally encoded information also allow functionality to be modified and recombined across organizational and sectoral boundaries (Ekbia, 2008).

These characteristics give some digital technologies relatively broad, general-purpose potential. AI, software, and data analytics, for example, can support diverse sustainability applications, including renewable-energy forecasting, resource optimization, carbon accounting, industrial efficiency, supply-chain management, and environmental monitoring (Calvino et al., 2025). Other technologies, however, depend more strongly on sector-specific infrastructures and complementary assets. Robotics requires integration with physical production systems, IoT and sensors depend on connected physical assets, and digital twins typically require detailed representations of physical processes or infrastructure. Blockchain, similarly, may be particularly relevant where traceability, verification, or decentralized transactions are required.

This heterogeneity suggests that digital technologies should not be treated as interchangeable indicators of "digitalization." Recent research increasingly distinguishes between specific digital technologies when examining their relationships with environmental innovation and sustainability outcomes (Aklilu et al., 2025). For green entrepreneurship, the relevant question is therefore not simply whether a venture is digital, but which digital technologies it uses, how these technologies are configured, and how they contribute to particular environmental objectives.

2.4. Distinguishing Digital Technologies, Infrastructure, and Applications

Measuring digitalization presents an important conceptual challenge because industry databases, patents, surveys, and business descriptions often combine technologies with infrastructure, sectors, organizational practices, and end-products. As a result, firms may be classified as "digital" based on indicators that represent fundamentally different technological phenomena. A useful distinction is between core digital technologies, digital infrastructure and enabling technologies, and technology applications. AI, machine learning, and cloud computing represent relatively identifiable technological capabilities, whereas ICT and information technology are broad categories encompassing multiple technologies. Hardware and telecommunications primarily provide enabling physical and network infrastructure, while autonomous vehicles, drones, cryptocurrency, and decentralized finance represent applications or systems that combine multiple underlying technologies. For example, an autonomous-vehicle venture may rely on AI, machine learning, sensors, software, and cloud computing; treating "autonomous vehicles" as equivalent to AI therefore conflates an application with an underlying technological capability (Yoo et al., 2010).

This distinction is particularly important for measuring the twin transition, where existing indicators often capture different technological levels and dimensions of digital and green transformation (Bruno et al., 2026). Accordingly, this study distinguishes between core digital technologies, digital infrastructure and enabling technologies, and digital applications and practices. This framework helps identify the specific digital capabilities embedded in green ventures while reducing the risk of overstating digitalization through broad ICT or technology-enabled categories.

2.5. From Individual Technologies to Technology Communities

Although distinguishing individual digital technologies is important, examining them in isolation can provide an incomplete picture. Digital innovation is inherently combinatorial, with technologies typically operating through complementary relationships. AI, for example, depends on software and data infrastructures, while IoT combines sensors, connectivity, data processing, and cloud services. Similarly, robotics integrates software, sensing, computational capabilities, and physical systems. The layered and recombinable nature of digital innovation (Yoo et al., 2010) therefore suggests that, in some contexts, the relevant unit of analysis is a configuration of technologies rather than an individual technology. This perspective is particularly relevant to entrepreneurial ventures, which often create new products and services by combining existing technological components in novel ways.

Technology co-occurrence provides a means of identifying such configurations empirically. When technologies repeatedly appear within the same ventures, their co-occurrence may indicate technological complementarity, shared knowledge bases, or common application contexts. Network analysis can represent these relationships as technologies and their co-

occurrences, while community-detection methods can identify groups with stronger internal associations. This approach shifts the analysis from individual technologies toward broader digital capability domains. For example, AI and machine learning may form one capability domain, while cloud computing and enterprise software may form another. These domains can then be compared in terms of their prevalence, structure, and application across green entrepreneurial activities.

Despite the combinatorial nature of digital innovation, such a meso-level perspective remains relatively underdeveloped in twin-transition research. Existing studies commonly examine aggregate digitalization, individual technologies, firm-level adoption, or patent categories, while less attention has been given to how digital technologies combine and cluster within sustainability-oriented ventures.

2.6. Digital Technologies Across Green Business Domains

A second dimension of technological heterogeneity concerns where digital technologies are applied. Sustainability spans diverse domains, including energy, transportation, agriculture, manufacturing, buildings, water, waste, biodiversity, and carbon management. Thus, the significance of a digital technology depends not only on its technical properties but also on the sustainability problems and application contexts in which it is deployed. Some technologies have relatively broad, horizontal applicability. AI, software, and cloud computing, for example, can support renewable-energy forecasting, industrial optimization, climate-risk assessment, waste sorting, precision agriculture, and environmental analytics. Other technologies are more closely tied to particular physical or organizational environments. IoT and sensing technologies are especially relevant where continuous monitoring is required, while robotics depends on interaction with physical systems and may therefore be concentrated in manufacturing, logistics, and waste processing. Blockchain may be particularly relevant to applications involving traceability, verification, transactions, or decentralized coordination.

This distinction between horizontal diffusion and vertical specialization provides a useful framework for understanding digitalization in green entrepreneurship. General-purpose technologies should be distributed across a wider range of sustainability domains, whereas more specialized technologies should exhibit more concentrated application patterns. Recent twin-transition research supports the view that digital technologies operate through different environmental mechanisms and exhibit distinct relationships with environmental innovation (Calvino et al., 2025; Aklilu et al., 2025).

However, limited research has systematically examined these diffusion patterns among green ventures. Identifying whether digital technologies are broadly distributed across sustainability domains or concentrated in specific applications can therefore reveal whether they function primarily as general enabling capabilities or as components of specialized green technological trajectories.

2.7. Green Business Penetration and the Twin Transition

Building on the preceding discussion, this study conceptualizes **green business penetration** as the extent and pattern through which digital technology communities are represented across sustainability-oriented business domains. This perspective shifts attention from whether digitalization supports sustainability to how digital and green activities intersect. It therefore complements twin-transition research by identifying the specific technology–

business configurations through which digital capabilities are applied to sustainability challenges.

Three gaps motivate this study. First, twin-transition research often examines digitalization at an aggregate level or focuses on individual technologies, providing limited insight into how technologies combine into distinct capability communities. Second, although the environmental applications of technologies such as AI, IoT, blockchain, cloud computing, and robotics are increasingly documented, less is known about how these configurations are distributed across green business domains. Third, existing measurement approaches—such as patents and firm surveys—provide only partial views of the relationship between technological capabilities and entrepreneurial activity. Venture-level data offer an opportunity to connect these dimensions more directly.

3. Data and Methodology

This study utilizes data on green ventures obtained from Crunchbase, a leading database of technology startups and entrepreneurship. Crunchbase provides detailed information on ventures, including business activities, location, founding date, employment, and financing history. The platform aggregates information from venture capital firms, incubators, accelerators, entrepreneurs, and publicly available online sources. Since its establishment in 2007, Crunchbase has become a prominent source of entrepreneurial data and it has been widely used in academic research (Dalle et al., 2017).

The sample is limited to ventures founded between 2016 and 2025 and classified by Crunchbase under the categories “Green Tech,” “Clean Tech,” or “Sustainability.” The selected period reflects entrepreneurial activity in the aftermath of the Paris Agreement in 2015, a landmark policy initiative that significantly strengthened the institutional framework for sustainability transitions and climate-oriented innovation (Geels, 2020; Köhler et al., 2019). After applying these selection criteria and removing missing values, the final dataset comprises 6,580 green ventures. Table 1 reports the distribution of these ventures across the different time periods.

Table 1. Number of green ventures in the period 2016-2025

Time period	Number of green ventures
2016	758
2017	785
2018	914
2019	604
2020	1015
2021	995
2022	594
2023	486
2024	305
2025	124

Source: CrunchBase.

Indusrtry and Technology Tag Classification

To identify digital technologies embedded within green ventures, we analyze the business and technology classification tags assigned to each company in the dataset. Crunchbase assigns each venture one or more descriptive tags capturing its industry focus, operational domain, and technological orientation. Table A1 (Annex I) presents the full distribution of industry and technology tags observed across the sample of green ventures, ranked by frequency of occurrence. The most prevalent tags—Renewable Energy (n = 754), Energy (n = 749), Manufacturing (n = 669), and Software (n = 655)—reflect the sectoral composition of the green entrepreneurship landscape, while tags such as Artificial Intelligence (AI) (n = 334), Mobile Apps (n = 81), Analytics (n = 97), Hardware (n = 73), and E-Commerce Platforms (n = 40) capture the digital technology dimension central to our statistical analysis.

Classification of Digital Technologies - Theoretical Framework

To classify the digital technology tags identified in our Crunchbase-derived dataset, we draw on Berger et al.'s (2018) multi-layer taxonomy of digital technologies, which characterizes digital artifacts according to three defining properties: re-programmability, data homogenization, and self-referential capacity for generating and processing digital information. This framework is complemented by the Digital Artifact Theory (Ekbia, 2008), which distinguishes digital artifacts from their physical instantiations, and by the Layered Modular Architecture perspective (Yoo et al., 2010), which conceptualizes digital innovation as occurring across distinct but interdependent layers—device, network, service, and content.

Building on these established frameworks, we manually classify all digital-technology-relevant Crunchbase tags into three conceptually distinct tiers: (1) Core Digital Technologies (Pure Digital Artifacts), (2) Digital Infrastructure and Enabling Technologies, and (3) Applications and Practices of Digital Technologies. This tiered structure allows us to distinguish digital technologies in the strict, artifact-based sense from the infrastructural substrates that support them and from the downstream products and practices in which they are embedded—a distinction that is critical for avoiding conflation of technology adoption with mere technology-enabled output.

Core Digital Technologies (Table A4)

The first tier comprises technologies that satisfy Berger et al.'s (2018) criteria for digital artifacts in the strict sense: they are re-programmable, operate on homogenized (digitally encoded) data, and exhibit self-referential properties characteristic of computational systems. This tier includes 33 core digital technology tags spanning software (Software, Consumer Software, Enterprise Software), computational intelligence (Artificial Intelligence, Machine Learning, and their specialized subdomains—Generative AI, Natural Language Processing, Computer Vision), data-engineering technologies (Database, Data Management, Data Integration, Data Mining, Big Data), cloud-based computational architectures (Cloud Computing, Cloud Infrastructure, SaaS, PaaS), geospatial and remote sensing systems, distributed ledger technologies (Blockchain, Web3), automation technologies (RPA, Robotics), sensing and connectivity technologies (Internet of Things, Sensors), and immersive and three-dimensional computational technologies (Virtual Reality, 3D Technology). All these digital technologies are often linked to the Industry 4.0 transformation.

Two classification decisions warrant explicit justification. First, although Sensors and Remote Sensing technologies possess physical hardware components, we retain them within the Core Digital Technologies tier rather than reclassifying them as enabling infrastructure. This decision follows environmental innovation literature, which treats sensing technologies as fundamentally digital data-acquisition systems—critical to smart irrigation, air quality monitoring, precision agriculture, and biodiversity tracking—rather than as passive hardware components (Klerkx et al., 2019; Wolfert et al., 2017; Liakos et al., 2018; Weng, 2019).. Second, Blockchain and Web3 technologies, despite relying on distributed network infrastructure, are classified as core digital technologies because their defining function is the decentralized validation, storage, and execution of digital transactions and smart contracts, consistent with Berger et al.'s self-referential and data-homogenization criteria.

Digital Infrastructure and Enabling Technologies (Table A5)

The second tier comprises technologies that constitute the physical and architectural substrate enabling digital computation, but which do not themselves constitute digital artifacts under Berger et al.'s (2018) definition. This includes hardware-layer technologies (Hardware, Electronics, Semiconductors), macro-scale physical infrastructure (Data Centers, IT Infrastructure), and communication infrastructure (Telecommunications, Satellite Communication, Communications Infrastructure). Consistent with the Layered Modular Architecture framework (Yoo et al., 2010), these technologies occupy the device and network layers that support, but are conceptually distinct from, the service- and content-layer digital artifacts captured in Table A4.

We additionally classify Information Technology and Information and Communications Technology (ICT) as foundational rather than specific digital technologies, and exclude both from our core digital technology dictionary. This decision follows the OECD's (2017) distinction between specific technologies and broad sector categories: umbrella terms such as IT and ICT carry minimal signal regarding which underlying digital technologies are actually present in a venture, and their inclusion risks artificially inflating measured digital technology adoption rates.

Applications and Practices of Digital Technologies (Table A6)

The third tier comprises tags that represent digital capabilities, practices, or downstream application domains rather than digital technologies per se. Software Engineering is classified as a development practice rather than a technology artifact. Industrial Automation is excluded for being overly broad, conflating software, control systems, and physical machinery without specifying the underlying digital technology. Autonomous Vehicles and Drones are classified as application platforms or end-products that embed digital technologies (e.g., AI, sensors, robotics) rather than representing core digital technologies in themselves. Similarly, Cryptocurrency, Non-Fungible Tokens (NFTs), and Decentralized Finance (DeFi) are classified as blockchain applications—business models and digital assets built atop blockchain infrastructure—rather than as core digital technologies.

Construction of the Digital Technology Dictionary

Our final digital technology dictionary, used to identify digital technology adoption among green ventures, comprises the 33 tags classified as Core Digital Technologies in Table A4. We exclude all tags classified as Digital Infrastructure and Enabling Technologies (Table A5) and

Applications and Practices of Digital Technologies (Table A6) from the adoption dictionary, on the grounds that their presence reflects either supporting infrastructure, overly broad sectoral categorization, or downstream application rather than direct adoption of a specific digital technology. This tiered exclusion strategy ensures that our measure of digital technology adoption captures genuine engagement with programmable, data-driven digital artifacts, rather than incidental association with hardware, infrastructure, or technology-enabled products.

Identification of Digital Technology Communities

To identify meaningful groups of digital technologies, we first extracted all digital technology keywords associated with ventures in the Crunchbase database. Following the digital technology classification presented in Table A4, we retained only keywords corresponding to Core Digital Technologies (Pure Digital Artifacts) and excluded all other venture industry and technology tags. Next, we constructed a technology co-occurrence network based on the simultaneous appearance of digital technology tags within individual venture profiles. Co-occurrence analysis assumes that technologies frequently mentioned together are more likely to represent related technological capabilities or complementary knowledge domains. To identify robust relationships among technologies, we compared four alternative graph specifications that differed in the treatment of co-occurrence strength and frequency thresholds (Table A7). Specifically, we evaluated networks based on raw co-occurrence counts and normalized pointwise mutual information (NPMI), a measure that adjusts for differences in keyword frequency and captures the strength of association between technology tags.

Table 2. Technology Communities Identified Through Network Analysis

Community Number	Community Name	Digital Technology Tag (Frequency)	Community Frequency
Community 0	AI & Machine Learning	Artificial Intelligence (AI) (242), Machine Learning (113), Generative AI (16), Computer Vision (13), Data Collection and Labeling (7), Remote Sensing (5)	396
Community 1	Robotics & 3D Tech	Robotics (35), 3D Technology (11), Robotic Process Automation (RPA) (5)	51
Community 2	Enterprise Software & Cloud Computing	Software (330), SaaS (182), Enterprise Software (49), Cloud Computing (18), PaaS (8), Consumer Software (8), Cloud Management (7), Cloud Infrastructure (6), Mapping Services (4), Developer Platform (4)	616
Community 3	Data & Analytics	Analytics (60), Big Data (47), Predictive Analytics (24), Data Visualization (22), Business Intelligence (18), Geospatial (15), Database (13), Data Integration (10)	209
Community 4	Blockchain & Web3	Blockchain (67), Web3 (24), Data Mining (5)	96
Community 5	IoT & Sensor Systems	Internet of Things (82), Sensor (15), Cloud Data Services (7)	104

Note. *Communities were identified using the NPMI > 0 and co-occurrence frequency ≥ 3 network variant (Q = 0.589). Community frequency represents the cumulative occurrence count of technologies assigned to each community, while* n *denotes the number of technology tags within the community.*

The quality of each network specification was assessed using modularity (Q), which measures the extent to which a network can be partitioned into internally cohesive and externally distinct communities. Among the four variants, the network constructed using positive NPMI values and a minimum co-occurrence frequency of three (Variant D) achieved the highest modularity

score (Q = 0.6003), indicating the clearest community structure. Consequently, this specification was selected for subsequent analyses.

Community detection applied to the selected network identified six distinct digital technology communities (Table 2): (1) AI and Machine Learning, (2) Robotics and 3D Technologies, (3) Enterprise Software and Cloud Computing, (4) Data and Analytics, (5) Blockchain and Web3, and (6) IoT and Sensor Systems (see Table 2). These communities represent clusters of technologies that frequently co-occur within entrepreneurial ventures and therefore reflect broader technological capability domains rather than isolated technologies.

Assignment of Ventures to Technology Communities

After establishing the technology-community structure, we assign individual ventures to the identified communities. Because many ventures contained multiple technology tags and some ventures lacked tags represented in the technology network, a multi-method assignment procedure was employed. First, a direct overlap approach assigned ventures to communities based on the presence of technology tags belonging to a given community. Second, an NPMI-based assignment method incorporated the strength of associations among technologies identified in the co-occurrence network. Third, a Naive Bayes classifier[1] was trained on the subset of ventures containing at least one matched technology tag (n = 1,356) and subsequently applied to the full sample of 6,580 ventures. The classifier leveraged the observed technology-community relationships to predict the most likely community membership of ventures.

To assess classification reliability, pairwise agreement was calculated across the three assignment methods (Table A8). The highest agreement was observed between the overlap and Naive Bayes approaches (98.3%), while agreement between the NPMI-based and alternative methods was approximately 79%. Overall, 1,068 ventures received consistent assignments across methods, whereas only four ventures (0.1% of the sample) exhibited ambiguous classifications.

Table 3. Community assignment results

Community	Description	Ventures	% of Total
Community 0	AI & Machine Learning	354	5.4%
Community 1	Robotics & 3D Tech	50	0.8%
Community 2	Enterprise Software & Cloud Computing	666	10.1%
Community 3	Data & Analytics	106	1.6%
Community 4	Blockchain & Web3	98	1.5%
Community 5	IoT & Sensor Systems	82	1.2%
Unassigned	No matching graph tags	5,224	79.4%
Total		6,580	100.0%

The final classification results indicate that Enterprise Software and Cloud Computing represents the largest technology community (666 ventures), followed by AI and Machine Learning (354 ventures). Blockchain and Web3, IoT and Sensor Systems, Data and Analytics,

[1] A Naive Bayes classifier is a very simple but powerful probabilistic machine learning algorithm used for classification tasks. It predicts the probability that an observation belongs to a particular class by applying Bayes' theorem under the assumption that all input features are conditionally independent given the class. This strong independence assumption is why it is called naive. See e.g. Lewis (1998)

and Robotics and 3D Technologies account for smaller but distinct segments of the entrepreneurial landscape (see Table 3). A substantial proportion of ventures (79.4%) remained unassigned because they did not contain technology tags represented in the final co-occurrence network. Unassigned ventures are creating the biggest community. Second biggest community is Enterprise Software & Cloud Computing. Third biggest community is AI & Machine Learning. The smallest community is Robotics & 3D Tech.

Validation of Community Structure

To evaluate whether the identified technology communities represented distinct technological domains rather than overlapping categories, we conducted a community pair analysis (Table A9). Three complementary measures were employed: the phi coefficient (φ), which captures the correlation between venture memberships across communities; the Jaccard coefficient, which measures the proportion of shared ventures relative to the combined membership of two communities; and inter-community NPMI, which captures the average strength of cross-community technology associations.

The results indicate generally low levels of overlap between communities. While moderate co-adoption patterns were observed between AI & Machine Learning and Enterprise Software & Cloud Computing ($\varphi = 0.223$), as well as between AI & Machine Learning and Data & Analytics ($\varphi = 0.174$), no community pair simultaneously exceeded the predefined thresholds for potential merging ($\varphi > 0.30$ and Jaccard > 0.20). Consequently, all six communities were retained as analytically distinct categories.

Taken together, the community detection, classification, and validation procedures provide a data-driven taxonomy of digital technologies that captures the major technological domains represented among green ventures.

Topic Modeling and Sustainability Domain Identification

To further examine the substantive activities of digitalized green ventures, we applied topic modeling to venture business descriptions. Topic modeling is an unsupervised machine-learning approach for identifying latent thematic structures in large text corpora by grouping words that frequently co-occur across documents (Blei et al., 2003; Dieng et al., 2020). We used this approach to identify the primary sustainability domains addressed by digitalized green ventures and examine how these domains vary across technology communities. The analysis was conducted on ventures classified as digitalized using the technology-community identification procedure described above. Venture descriptions were preprocessed using standard natural language processing procedures, including text cleaning, tokenization, lowercasing, stop-word removal, and lemmatization, before being transformed into a document-term matrix.

We estimated models with 2–34 topics and evaluated them using topic coherence, perplexity, and topic diversity (Table A10). Coherence assesses the semantic interpretability of topics, while perplexity evaluates predictive performance; topic diversity captures the uniqueness of words across topics and helps identify redundant or overlapping topics (Dieng et al., 2020). The 34-topic model achieved the highest coherence (0.562), while maintaining high topic diversity (0.800) and favorable perplexity (−7.444), and was therefore selected as the optimal specification. We then manually reviewed the 34 machine-generated topics and consolidated them into nine broader sustainability domains based on representative keywords and venture

descriptions: (1) Food Waste and Circular Economy, (2) Water and Environmental Management, (3) Industrial Operations and Manufacturing, (4) Buildings and Energy Infrastructure, (5) Renewable Energy Infrastructure, (6) Sustainable Manufacturing, (7) Climate Action and Carbon Management, (8) Urban Sustainability, and (9) Corporate Sustainability and Supply Chain Management (Table 14). Combining topic modeling with technology-community analysis provides a multidimensional characterization of digitalized green entrepreneurship, capturing both the technologies ventures employ and the sustainability challenges they address.

4. Results and Discussion

Network analysis of digital-technology tag co-occurrences identified six distinct technology communities under the optimal network specification (NPMI > 0; co-occurrence frequency ≥ 3; Q = 0.589). Table 2 presents their composition, and Table 3 reports venture distributions. Enterprise Software & Cloud Computing is the largest community, comprising 616 tag occurrences and 666 ventures (10.1% of the sample), followed by AI & Machine Learning, with 396 tag occurrences and 354 ventures (5.4%). The remaining communities are substantially smaller: Data & Analytics (106 ventures; 1.6%), IoT & Sensor Systems (82; 1.2%), Blockchain & Web3 (98; 1.5%), and Robotics & 3D Tech (50; 0.8%).

Overall, 5,224 ventures (79.4%) were not assigned to a digital technology community, leaving 1,356 ventures (20.6%) with identifiable adoption of core digital technologies. Enterprise Software & Cloud Computing and AI & Machine Learning account for approximately three-quarters of these digitalized ventures, indicating that digitalization in green entrepreneurship is concentrated in software- and data-intensive capabilities rather than more specialized technologies such as robotics, blockchain, and IoT. This pattern suggests that digital innovation in green entrepreneurship is primarily oriented toward data processing, process optimization, decision-making, and organizational capabilities.

Topic Modeling and Technology–Topic Alignment

To examine the substantive domains addressed by digitalized green ventures and their relationship with the technology-community structure, we applied LDA to venture descriptions from 2016–2025. The thematic distribution is concentrated in governance, infrastructure, and industrial applications. T9: Corporate Sustainability and Supply Chain Management is the most prevalent topic (19.9%), characterized by terms such as sustainability, data, platform, ESG, SaaS, and supply chain, reflecting ventures focused on ESG reporting, sustainability analytics, and supply-chain solutions. T8: Urban Sustainability (14.6%) and T4: Buildings & Energy Infrastructure (13.2%) follow, alongside T3: Industrial Operations & Manufacturing and T5: Renewable Energy Infrastructure (11.1% each). By contrast, T6: Sustainable Manufacturing (5.3%), T1: Food Waste and Circular Economy (7.8%), and T2: Water & Environmental Management (7.8%) account for smaller shares, indicating comparatively less digitalized activity in these domains (see Table 4).

Table 4. Topic-based classification and trend identification for digitalized green ventures (2016-2025)

Topic label	Top ten words	Topic prevalence
T1: Food Waste and Circular Economy	Waste, food, intelligence, artificial, circular, management, market, platform, mobile, recycling	7.8%
T2**:** Water & Environmental Management	Technology, solution, cloud, datum, water, monitoring, sensor, finance, nature, edge	7.8%
T3: Industrial Operations & Manufacturing	System, iot, cost, robotic, energy, network, autonomous, smart, industry, control	11.1%
T4: Buildings & Energy Infrastructure	Energy, solution, real, building, smart, time, power, solar, insight, electric	13.2%
T5: Renewable Energy Infrastructure	Intelligence, energy, asset, saas, sustainable, renewable, artificial, infrastructure, analysis, data	11.1%
T6: Sustainable Manufacturing	Impact, social, efficiency, material, environmental, solution, air, manufacturing , quality, measure	5.3%
T7: Climate Action & Carbon Management	Carbon, emission, climate, footprint, action, impact, reduction, transformation empower, natural	8.9%
T8: Urban Sustainability	Sustainable, product, design, financial, home, space, environment, mobility, urban, tool	14.6%
T9: Corporate Sustainability and Supply Chain Management	Sustainability, datum, platform, esg, saas, management, solution, chain, analytic, supply	19.9%

Table 5. Share (%) of Green Ventures by Dominant Topic Across Technology Communities.

Green Ventures	AI & Machine Learning	Robotics & 3D Tech	Enterprise Software & Cloud Computing	Data & Analytics	Blockchain & Web3	IoT & Sensor Systems
T1: Food Waste & Circular Economy	10.8	0.0	8.5	3.6	7.6	1.4
T2: Water & Environmental Management	7.1	3.4	6.0	15.2	9.8	17.8
T3: Industrial Operations & Manufacturing	13.3	32.8	8.2	8.9	9.8	15.1
T4: Buildings & Energy Infrastructure	14.2	6.9	12.7	8.0	13.0	27.4
T5: Renewable Energy Infrastructure	10.5	0.0	11.7	9.8	22.8	5.5
T6: Sustainable Manufacturing	5.9	12.1	4.0	1.8	9.8	8.2
T7: Climate Action & Carbon Management	9.1	0.0	10.2	8.9	9.8	2.7
T8: Urban Sustainability	13.0	41.4	13.9	8.9	15.2	15.1
T9: Corporate Sustainability & Supply Chain Management	16.1	3.4	24.7	34.8	2.2	6.8

Table 5 cross-tabulates technology communities by their dominant topics and reveals substantial variation in thematic concentration. We measure this heterogeneity using normalized Shannon entropy, where higher values indicate a more even distribution across topics and lower values indicate greater concentration (Shannon, 1948). AI & Machine Learning has the highest normalized entropy (0.981), with venture shares distributed relatively evenly across all nine topics and no topic exceeding 16.1%. This broad distribution suggests that AI/ML operates as a general-purpose digital capability across diverse green-business

domains rather than being tied to a specific sustainability niche. In contrast, Robotics & 3D Tech has the lowest entropy (0.637) and the strongest thematic concentration. Urban Sustainability (T8) accounts for 41.4% of ventures and Industrial Operations & Manufacturing (T3) for 32.8%, together representing 74.2% of the community. The absence of ventures in Food Waste and Circular Economy (T1), Renewable Energy Infrastructure (T5), and Climate Action and Carbon Management (T7) further indicates a relatively narrow application profile centered on spatial, urban, and manufacturing contexts.

The remaining communities occupy intermediate positions. Blockchain & Web3 and Enterprise Software & Cloud Computing both show high topical diversity (entropy = 0.946), but their dominant applications differ: Blockchain & Web3 is most concentrated in Renewable Energy Infrastructure (T5; 22.8%), whereas Enterprise Software & Cloud Computing peaks in Corporate Sustainability and Supply Chain Management (T9; 24.7%). IoT & Sensor Systems (0.882) is most prevalent in Buildings & Energy Infrastructure (T4; 27.4%), while Data & Analytics (0.874) shows its strongest concentration in Corporate Sustainability and Supply Chain Management (T9; 34.8%).

Overall, the results reveal a continuum from horizontally applicable to domain-specific digital technologies. AI/ML exhibits the broadest cross-domain applicability, whereas Robotics & 3D Tech is the most specialized. The intermediate communities display distinct application affinities, with Blockchain & Web3 oriented toward renewable energy, Enterprise Software & Cloud Computing and Data & Analytics toward corporate sustainability and supply chains, and IoT & Sensor Systems toward buildings and energy infrastructure. These patterns suggest that the functional characteristics of digital technologies shape the sustainability domains in which green ventures deploy them.

5. Conclusion

This study examined how digital technologies intersect with green entrepreneurship by developing a data-driven taxonomy of digital technologies and tracing their application across sustainability domains. Using 6,580 green ventures founded between 2016 and 2025, we combined theoretically grounded technology classification, co-occurrence network analysis, venture-level community assignment, and topic modeling. This approach provides a multidimensional view of digital green entrepreneurship by linking the technological capabilities adopted by ventures to the sustainability challenges they address.

Three main findings emerge. First, identifiable adoption of core digital technologies remains limited but significant: 1,356 ventures (20.6%) were assigned to one of six digital technology communities, while 79.4% had no identifiable representation in the final digital-technology network. Digitalization is therefore an important, but far from universal, characteristic of green entrepreneurship. This distinction is important because participation in a sustainability-related sector does not necessarily imply the adoption of core digital technologies. Second, digitalization is concentrated in a small number of technological domains. Enterprise Software & Cloud Computing and AI & Machine Learning account for the majority of digitalized ventures, while Data & Analytics, IoT & Sensor Systems, Blockchain & Web3, and Robotics & 3D Tech are considerably smaller. The prominence of software-, cloud-, and AI-based technologies suggests that digital green entrepreneurship is driven primarily by scalable computational and information capabilities that support monitoring, optimization, prediction,

coordination, and decision-making across sustainability contexts. Third, technologies differ substantially in the breadth of their sustainability applications. Corporate Sustainability and Supply Chain Management is the largest thematic domain, followed by Urban Sustainability and Buildings & Energy Infrastructure. AI & Machine Learning exhibits the highest topic diversity, with applications distributed broadly across sustainability domains, consistent with its role as a general-purpose technology. By contrast, Robotics & 3D Tech is strongly concentrated in urban and industrial applications. Other communities display distinct application affinities: Enterprise Software & Cloud Computing and Data & Analytics are closely associated with corporate sustainability and supply chains, IoT & Sensor Systems with buildings and energy infrastructure, and Blockchain & Web3 with renewable energy infrastructure. These patterns demonstrate that digital technologies should not be treated as a homogeneous category in studies of green entrepreneurship.

The study makes three contributions. Conceptually, it distinguishes core digital technologies from enabling infrastructure and downstream digital applications, reducing the risk of equating any technology-related activity with digital technology adoption. Methodologically, it demonstrates how venture-level classification, network analysis, community detection, and topic modeling can be integrated to construct an empirically grounded taxonomy of digital green entrepreneurship. Empirically, it reveals a heterogeneous landscape in which broadly applicable technologies coexist with more specialized technology–sustainability configurations.

The findings also have implications for policy and entrepreneurial ecosystems. General-purpose technologies such as AI, software, and cloud computing may benefit from horizontal policies supporting digital skills, data access, computational infrastructure, and technology diffusion. More specialized technologies, including robotics, IoT, and blockchain, may require sector-specific interventions such as demonstration projects, standards, complementary physical infrastructure, and coordination between technology providers and sustainability actors. Supporting the digital and green transitions therefore requires both broad digital capabilities and targeted technology–sector complementarities.

Several limitations should be acknowledged. The analysis relies on Crunchbase classifications and publicly reported venture descriptions, which may be incomplete or inconsistently reported. Consequently, the absence of a digital technology tag does not imply that a venture does not use digital technologies internally; rather, it indicates the absence of an identifiable digital orientation in the available data. The technology communities also depend on observed tag co-occurrences and methodological choices in network construction, while the consolidation of machine-generated topics into broader sustainability domains involves researcher interpretation. Finally, the analysis identifies technological and thematic associations but does not establish causal effects on venture performance, environmental outcomes, innovation, or survival.

Future research could examine how technology communities evolve over time and whether institutional, technological, or economic changes reshape digital green entrepreneurship. Linking the taxonomy to financing, growth, patenting, survival, and exit outcomes could reveal whether particular digital capabilities generate entrepreneurial advantages. Further research could also examine geographic variation and the influence of innovation systems, environmental policy, digital infrastructure, and venture capital on technology–sustainability specialization. Combining venture databases with patents, websites, investment records, and

product-level data would provide additional opportunities to validate and extend the classification.

Overall, the findings show that the digital and green transitions converge unevenly at the venture level. Digital green entrepreneurship combines broadly applicable computational technologies, particularly software, cloud computing, and AI, with more specialized configurations such as robotics, IoT, and blockchain. Understanding this technological heterogeneity is essential for explaining how digital capabilities are mobilized to address sustainability challenges and how the digital and green transformations become interconnected within entrepreneurial ecosystems.

Acknowledgment

This project is funded by the EU NextGenerationEU through the recovery and resilience plan for Slovakia under the project No.09I03-03-V04-00509.

References:

Aklilu, A. Z., Dussaux, D., & Verrier, A. (2025). Quantifying digital innovation for the twin transition: Historical trends, current landscape and links with environmental innovation. *OECD Environment Working Papers, No. 261*. OECD Publishing. https://doi.org/10.1787/8cc4cff0-en

Ardito, L. (2023). The influence of firm digitalization on sustainable innovation performance and the moderating role of corporate sustainability practices: An empirical investigation. *Business Strategy and the Environment, 32*(8), 5252–5272. https://doi.org/10.1002/bse.3415

Ardito, L., Raby, S., Albino, V., & Bertoldi, B. (2021). The duality of digital and environmental orientations in the context of SMEs: Implications for innovation performance. *Journal of Business Research, 123*, 44–56. https://doi.org/10.1016/j.jbusres.2020.09.022

Ben Youssef, A. (2025). Twin transition: Digital transformation pathways for sustainable innovation. *Sustainability, 17*(21), 9491. https://doi.org/10.3390/su17219491

Berger, S., Denner, M.-S., & Röglinger, M. (2018). The nature of digital technologies—Development of a multi-layer taxonomy. In *Proceedings of the 26th European Conference on Information Systems (ECIS 2018)*.

Bianchini, S., Damioli, G., & Ghisetti, C. (2023). The environmental effects of the "twin" green and digital transition in European regions. *Environmental and Resource Economics, 84*(4), 877–918. https://doi.org/10.1007/s10640-022-00741-7

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. In *Proceedings of the Biennial GSCL Conference 2009* (pp. 31–40).

Bzhalava, L., Makkonen, T., Šipikal, M., Omorede, A. & Hassan, S.S. (2026). Detecting promising green ventures with foresight and data mining**.** *Sustainable Futures*, 12, 101966. https://doi.org/10.1016/j.sftr.2026.101966

Calvino, F., Dechezleprêtre, A., & Haerle, D. (2025). *How can the digital transformation affect the net-zero transition? A conceptual framework for the "twin" transition*. OECD Science, Technology and Industry Working Papers, No. 2025/17. OECD Publishing. https://doi.org/10.1787/4ee71f6d-en

Cohen, B., & Winn, M. I. (2007). Market imperfections, opportunity and sustainable entrepreneurship. *Journal of Business Venturing*, *22*(1), 29–49. https://doi.org/10.1016/j.jbusvent.2004.12.001

Dalle, J.-M., den Besten, M., & Menon, C. (2017). Using Crunchbase for economic and managerial research. *OECD Science, Technology and Industry Working Papers, No. 2017/08*. OECD Publishing. https://doi.org/10.1787/6c418d60-en

Dean, T. J., & McMullen, J. S. (2007). Toward a theory of sustainable entrepreneurship: Reducing environmental degradation through entrepreneurial action. *Journal of Business Venturing*, *22*(1), 50–76. https://doi.org/10.1016/j.jbusvent.2005.09.003

Ekbia, H. R. (2009). Digital artifacts as quasi-objects: Qualification, mediation, and materiality. *Journal of the American Society for Information Science and Technology*, *60*(12), 2554–2566. https://doi.org/10.1002/asi.21189

Gast, J., Gundolf, K., & Cesinger, B. (2017). Doing business in a green way: A systematic review of the ecological sustainability entrepreneurship literature and future research directions. *Journal of Cleaner Production*, *147*, 44–56. https://doi.org/10.1016/j.jclepro.2017.01.065

Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: A multi-level perspective and a case-study. *Research Policy*, *31*(8–9), 1257–1274. https://doi.org/10.1016/S0048-7333(02)00062-8

Geels, F. W. (2020). Micro-foundations of the multi-level perspective on socio-technical transitions: Developing a multi-dimensional model of agency through crossovers between social constructivism, evolutionary economics and neo-institutional theory. *Technological Forecasting and Social Change*, *152*, 119894. https://doi.org/10.1016/j.techfore.2019.119894

Holzmann, P., & Gregori, P. (2023). The promise of digital technologies for sustainable entrepreneurship: A systematic literature review and research agenda. *International Journal of Information Management*, *68*, 102593. https://doi.org/10.1016/j.ijinfomgt.2022.102593

Kim, J., & Jin, W. (2024). Impact of digital capabilities on entrepreneurial performance in SMEs. *Journal of Innovation & Knowledge, 9*(4), 100609. https://doi.org/10.1016/j.jik.2024.100609

Köhler, J., Geels, F. W., Kern, F., Markard, J., at al. (2019). An agenda for sustainability transitions research: State of the art and future directions. *Environmental Innovation and Societal Transitions, 31,* 1–32. https://doi.org/10.1016/j.eist.2019.01.004

Markard, J., Raven, R., & Truffer, B. (2012). Sustainability transitions: An emerging field of research and its prospects. *Research Policy, 41*(6), 955–967. https://doi.org/10.1016/j.respol.2012.02.013

McCallum, A., & Nigam, K. (1998). A comparison of event models for Naive Bayes text classification. In *Proceedings of the AAAI-98 Workshop on Learning for Text Categorization* (pp. 41–48).

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (pp. 262–272).

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences, 103*(23), 8577–8582. https://doi.org/10.1073/pnas.0601602103

OECD. (2017). *OECD science, technology and industry scoreboard 2017: The digital transformation*. OECD Publishing. https://doi.org/10.1787/9789264268821-en

Schaltegger, S., & Wagner, M. (2011). Sustainable entrepreneurship and sustainability innovation: Categories and interactions. *Business Strategy and the Environment, 20*(4), 222–237. https://doi.org/10.1002/bse.682

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379–423; *27*(4), 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Siedschlag, I., Mohan, G., & Yan, W. (2024). Twin transitions across enterprises: Do digital technologies and sustainability go together? *Journal of Cleaner Production, 481*, 144025. https://doi.org/10.1016/j.jclepro.2024.144025

Sussan, F., & Acs, Z. J. (2017). The digital entrepreneurial ecosystem. *Small Business Economics, 49*(1), 55–73. https://doi.org/10.1007/s11187-017-9867-5

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). The new organizing logic of digital innovation: An agenda for information systems research. *Information Systems Research, 21*(4), 724–735. https://doi.org/10.1287/isre.1100.0322

Annex I

Table A1. Green Entrepreneurship Industry Tags

Keyword	Frequency	Keyword	Frequency	Keyword	Frequency
Renewable Energy	754	Green Consumer Goods	100	Architecture	49
Energy	749	Social Impact	98	Wholesale	48
Manufacturing	669	Logistics	98	Power Grid	47
Software	655	Packaging Services	98	Innovation Management	47
Environmental Consulting	592	Natural Resources	97	Biomass Energy	47
Waste Management	528	Analytics	97	Mining	45
Recycling	458	Real Estate	94	Lifestyle	45
Clean Energy	416	Health Care	94	Janitorial Service	45
Environmental Engineering	365	Carbon Capture	91	Pollution Control	44
E-Commerce	355	Advanced Materials	86	Facilities Support Services	44
Consulting	339	Water Purification	84	Nutrition	43
Artificial Intelligence (AI)	334	Apps	82	Alternative Protein	42
Agriculture	331	Mobile Apps	81	Product Research	42
Information Technology	330	Communities	79	Social	42
Solar	287	Food Processing	78	Infrastructure	41
Food and Beverage	271	Oil and Gas	78	Management Consulting	41
Biotechnology	269	Business Development	78	Life Science	41
AgTech	269	Education	75	Social Entrepreneurship	41
SaaS	259	Internet	74	Smart Building	40
Energy Efficiency	226	Hardware	73	Electrical Distribution	40
Electric Vehicle	221	Commercial	73	Aquaculture	40
Farming	211	Finance	70	E-Commerce Platforms	40
Fashion	184	Organic	70	Professional Services	40
Non Profit	175	Product Design	67	Consumer Electronics	39
Retail	169	Training	67	Machinery Manufacturing	39
Transportation	167	Smart Cities	67	Information Services	38
Energy Management	162	Aerospace	65	Marketing	38
Construction	161	Impact Investing	65	Home and Garden	37
Water	157	Building Material	62	Travel	37
FinTech	155	Electronics	61	Wind Energy	36
Consumer Goods	146	Robotics	61	Compliance	36
Marketplace	145	Service Industry	59	Events	34
Automotive	139	Big Data	59	Industrial Engineering	33
Financial Services	137	Organic Food	57	Risk Management	33
Internet of Things	134	Plastics and Rubber Manufacturing	57	Customer Service	33
Blockchain	134	Textiles	57	Smart Home	32
Machine Learning	125	Sharing Economy	57	Nanotechnology	32

Industrial	117	Wellness	57	Marine Technology	32
Energy Storage	113	Forestry	56	Venture Capital	31
Supply Chain Management	112	Charging Infrastructure	55	Hospitality	31
B2B	108	Enterprise Software	54	Consumer	31
Green Building	108	Project Management	54	Property Management	31
Battery	106	Retail Technology	51	Industrial Manufacturing	31
Chemical	101	Advice	50	Chemical Engineering	31

Table A2. Green Entrepreneurship Industry Tags

Keyword	Frequency	Keyword	Frequency	Keyword	Frequency
Residential	31	Shipping	19	Insurance	14
Cosmetics	30	Mobile	19	Charity	14
Apparel	30	Medical	19	Paper Manufacturing	13
Heating	30	Procurement	19	Computer Vision	13
Ventilation	30	Shopping	19	Mechanical Engineering	13
Air Conditioning (HVAC)	30	Public Safety	19	Delivery Service	13
Home Services	30	Delivery	18	Building Maintenance	13
Fuel	28	Laundry and Dry-cleaning	18	Sales	13
Animal Feed	28	Seafood	18	Data Management	13
Industrial Automation	28	Database	18	Property Development	13
Web3	27	Trading Platform	18	Online Portals	13
Data Visualization	27	Commercial Real Estate	17	Landscaping	13
Asset Management	27	Interior Design	17	Government	13
Tourism	27	Coffee	17	Horticulture	13
3D Technology	27	Facility Management	17	Sports	13
Predictive Analytics	26	Civil Engineering	17	Shoes	12
Cloud Computing	26	Cleaning Products	17	Technical Support	12
Rental	26	Geospatial	17	Public Transportation	12
PropTech	26	Advertising	17	Security	12
Data Center	25	Social Media	17	Real Estate Investment	12
Cryptocurrency	25	Mineral	17	Satellite Communication	12
Personal Health	25	Plant-Based Foods	16	Space Travel	12
Air Transportation	24	Lighting	16	Furniture	12
Beauty	23	Ride Sharing	16	Snack Food	11
Sensor	23	Wildlife Conservation	16	Product Management	11
Home Improvement	23	Cannabis	16	Payments	11
Grocery	22	Association	16	Data Integration	11
Semiconductor	22	Collaboration	16	GovTech	11
Subscription Service	22	Test and Measurement	16	EdTech	11
Business Intelligence	22	Telecommunications	16	Digital Media	11
Market Research	22	Small and Medium Businesses	16	Fuel Cell	11
Biofuel	22	Banking	16	Cyber Security	11
Fleet Management	22	Human Resources	16	Art	11
B2C	21	Web Development	16	Geothermal Energy	10
Marine Transportation	21	Pharmaceutical	15	Hydroelectric	10
Last Mile Transportation	21	Recruiting	15	Web Apps	10
Social Network	21	Fitness	15	Pet	10
Hydroponics	21	Medical Device	15	Parking	10

InsurTech	21	Software Engineering	15	Public Relations	10
Autonomous Vehicles	20	Productivity Tools	15	Media and Entertainment	10
Food Delivery	20	Nuclear	14	Event Management	10
Generative AI	20	Accounting	14	Digital Marketing	10
Drones	19	PaaS	14	Handmade	10
Mining Technology	19	Quality Assurance	14	Housekeeping Service	10

Table A3. Green Entrepreneurship Industry Tags

Keyword	Frequency	Keyword	Frequency	Keyword	Frequency
Outdoors	9	E-Learning	7	Non-Fungible Token (NFT)	5
Legal Tech	9	Management Information Systems	6		
Printing	9	Livestock	6		
Information and Communications Technology (ICT)	9	Gamification	6		
Brand Marketing	9	Bitcoin	6		
Cloud Data Services	9	Graphic Design	6		
Hotel	9	Remote Sensing	6		
Cloud Management	9	Water Transportation	6		
Sporting Goods	9	Field Support	6		
Travel Agency	9	Outsourcing	6		
Crowdfunding	9	Adventure Travel	6		
Wood Processing	9	Mapping Services	6		
Consumer Software	9	News	6		
Home Decor	9	Industrial Design	6		
Cycling	8	iOS	6		
Timber	8	Consumer Research	6		
Real Time	8	Car Sharing	6		
Intelligent Systems	8	Decentralized Finance (DeFi)	6		
Incubators	8	Lending	6		
Baby	8	IT Management	6		
Gift	8	Credit	6		
Personal Care and Hygiene	8	Publishing	6		
IT Infrastructure	8	Virtual Reality	6		
Freight Service	8	Railroad	5		
Computer	8	Professional Networking	5		
Corporate Training	8	Developer Platform	5		
Data Collection and Labeling	8	Home Appliances	5		
Precious Metals	8	Local Business	5		
Jewelry	7	Farmers Market	5		
Personal Development	7	Natural Language Processing	5		
Restaurants	7	Collaborative Consumption	5		
Cloud Infrastructure	7	Fruit	5		
Cooking	7	Emerging Markets	5		
Wealth Management	7	Dietary Supplements	5		
Navigation	7	UX Design	5		

Gaming	7	Leisure	5		
Developer APIs	7	Robotic Process Automation (RPA)	5		
Women's	7	Communications Infrastructure	5		
Wearables	7	Content Creators	5		
Boating	7	Creative Agency	5		
Data Mining	7	Funding Platform	5		
Loyalty Programs	7	Legal	5		

Table A4. Core Digital Technologies (Pure Digital Artifacts)

Digital Technology Keyword	Classification	Theoretical / Operational Rationale
Software / Consumer Software / Enterprise Software	Core Digital Tech	The programmable layer of digital systems, consisting of digital artifacts that enable computation, information processing, automation, and delivery of digital functionality.
Artificial Intelligence (AI) / Machine Learning	Advanced Digital Technology	Computational systems that use algorithms and data to learn patterns, make predictions, generate outputs, and support or automate decision-making.
Generative AI / Natural Language Processing / Computer Vision	Specialized AI Technology	Specialized domains of AI that enable generation, analysis, interpretation, and transformation of textual, visual, auditory, and other digital data.
Analytics / Predictive Analytics / Business Intelligence / Data Visualization	Core Digital Tech	Data-driven technologies using statistical, mathematical, and computational methods to transform data into actionable insights, support decisions, and forecast outcomes.
Database / Data Management / Data Integration / Data Collection and Labeling / Data Mining / Big Data	Core Digital Tech	Data-engineering technologies governing collection, storage, organization, integration, enrichment, and lifecycle management of digital information assets.
Cloud Computing / Cloud Data Services / Cloud Infrastructure	Core Digital Tech	Virtualized computational environments providing scalable storage, processing, networking, and data-management capabilities as digital resources.
SaaS / PaaS / Cloud Management / Developer Platform	Software-as-a-Service & Platform Architectures	Software-based service and platform architectures that support development, deployment, administration, and delivery of digital applications.
Cyber Security	Core Digital Tech	Algorithmic and cryptographic methods designed to protect the integrity, confidentiality, availability, and resilience of digital systems and information assets.
Geospatial / Remote Sensing / Mapping Services	Core Digital Tech	Spatial data-driven systems leveraging digital computation to acquire, process, analyze, and visualize location-based information from earth observation and sensing technologies.
Blockchain / Web3	Distributed Digital Infrastructure	Distributed ledger architectures enabling decentralized storage, validation, and execution of digital transactions and smart contracts.
Robotic Process Automation (RPA) / Robotics	Advanced Digital Technology	Automation technologies that combine software, sensors, control systems, and machine intelligence to execute tasks with limited human intervention.
Internet of Things / Sensor	Core Digital Tech	Networked sensing and communication technologies that capture physical signals, convert them into digital data, and enable machine-to-machine connectivity.

Virtual Reality	Advanced Digital Technology	Immersive digital environment technology that creates computer-generated interactive experiences for simulation, training, visualization, and engagement.
3D Technology	Advanced Digital Technology	The technology enabling the creation, modeling, simulation, visualization, and fabrication of three-dimensional objects and environments through computational design, rendering, digital twins, additive manufacturing, and related digital production processes.

Table A5. Digital Infrastructure and Enabling Technologies

Digital Technology Keyword	Classification	Theoretical / Operational Rationale
Hardware / Electronics / Semiconductor	Enabling Technology	Physical technological infrastructure, including semiconductor devices, electronic components, computing equipment, and communication hardware, that enables generation, processing, storage, and transmission of digital information.
Data Center / IT Infrastructure	Infrastructure	Macro-scale physical facilities housing servers, storage systems, networking equipment, cooling systems, and power architectures required for digital computation and cloud services.
Telecommunications / Satellite Communication / Communications Infrastructure	Infrastructure	Physical and electromagnetic communication networks providing connectivity and data transmission between distributed digital systems.
Information Technology / Information and Communications Technology (ICT)	Foundational Digital Technology	Broad technological domain encompassing digital computing, communications, networking, information processing, and associated software infrastructures. They carry almost no signal about *which* digital technologies are present

Table A6. Applications and Practices of Digital Technologies

Digital Technology Keyword	Classification	Theoretical / Operational Rationale
Software Engineering	Digital Capability / Practice	A discipline and development practice rather than a technology artifact itself.
Industrial Automation	Technology System / Application Domain	Too broad and often mixes software, control systems, and physical machinery.
Autonomous Vehicles	Digital Technology Application	An end-product/application domain that embeds digital technologies rather than representing a core digital technology itself.
Drones	Digital Technology Platform	A product/application platform rather than a foundational digital technology.
Cryptocurrency / Non-Fungible Token (NFT) / Decentralized Finance (DeFi)	Blockchain Applications	Blockchain-based applications, assets, or business models rather than core digital technologies.

Table A7. Modularity Comparison Across Graph Variants

Variant	Nodes	Edges	Density	Modularity (Q)	Communities
A: Raw count ≥3	34	109	0.1943	0.1030	4
B: Raw count ≥5	24	63	0.2283	0.1047	3
C: NPMI > 0	37	147	0.2207	0.4006	5
D: NPMI > 0 + count ≥3	33	57	0.1080	0.6003	6

Note. NPMI = normalized pointwise mutual information. Variant D (NPMI > 0 + count ≥3) produced the highest modularity score (Q = 0.6003) and was therefore selected as the best-performing graph specification for subsequent community detection analyses.

Table A8. Pairwise agreement between methods

Comparison	Agreement
Overlap vs. NPMI	79.3%
Overlap vs. Naive Bayes	98.3%
NPMI vs. Naive Bayes	79.4%

Note. NPMI = normalized pointwise mutual information. Pairwise agreement was calculated among ventures assigned to technology communities by the respective methods.

Table A9. Community Pair Analysis

Community Pair	Phi	Jaccard	Inter-NPMI	Relationship
AI & Machine Learning as focal community				
AI & Machine Learning × Enterprise Software & Cloud Computing	0.223***	0.154	0.041	Moderate co-adoption
AI & Machine Learning × Data & Analytics	0.174***	0.113	0.061	Moderate co-adoption
AI & Machine Learning × Robotics & 3D Tech	0.116***	0.057	0.076	Weak co-adoption
AI & Machine Learning × IoT & Sensor Systems	0.112***	0.070	0.080	Weak co-adoption
AI & Machine Learning × Blockchain & Web3	0.090***	0.058	0.000	No evidence of association
Enterprise Software & Cloud Computing as focal community				
Enterprise Software & Cloud Computing × Data & Analytics	0.164***	0.096	0.055	Moderate co-adoption
Enterprise Software & Cloud Computing × IoT & Sensor Systems	0.091***	0.053	0.029	Limited co-adoption
Enterprise Software & Cloud Computing × Blockchain & Web3	0.059***	0.040	0.070	Limited co-adoption
Remaining pairs				
Data & Analytics × IoT & Sensor Systems	0.116***	0.074	0.081	Weak co-adoption
IoT & Sensor Systems × Robotics & 3D Tech	0.053***	0.034	0.040	Limited co-adoption
Blockchain & Web3 × IoT & Sensor Systems	0.058***	0.041	0.011	Limited co-adoption
Robotics & 3D Tech × Data & Analytics	0.009	0.013	0.202	Limited co-adoption
Data & Analytics × Blockchain & Web3	0.017	0.022	0.000	No evidence of association
Robotics & 3D Tech × Enterprise Software & Cloud Computing	0.017	0.017	0.000	No evidence of association
Robotics & 3D Tech × Blockchain & Web3	0.001	0.009	0.000	No evidence of association

Table A10. Topic Model Evaluation Metrics Across Alternative Numbers of Topics

K	Coherence	Perplexity	Topic Diversity
2	0.229	-5.497	0.950
3	0.239	-5.586	0.733
4	0.280	-5.664	0.725
5	0.273	-5.699	0.700

6	0.297	-5.732	0.733
7	0.333	-5.756	0.700
8	0.337	-5.779	0.750
9	0.372	-5.789	0.744
10	0.369	-5.833	0.740
11	0.387	-5.846	0.727
12	0.388	-5.841	0.733
13	0.413	-5.881	0.738
14	0.443	-5.886	0.757
15	0.455	-5.995	0.767
16	0.468	-6.344	0.781
17	0.474	-6.552	0.812
18	0.470	-6.613	0.778
19	0.487	-6.685	0.795
20	0.478	-6.735	0.770
21	0.506	-6.818	0.829
22	0.478	-6.831	0.786
23	0.490	-6.901	0.787
24	0.487	-6.932	0.800
25	0.494	-6.991	0.800
26	0.491	-7.043	0.785
27	0.487	-7.074	0.789
28	0.495	-7.141	0.786
29	0.491	-7.183	0.807
30	0.529	-7.231	0.777
31	0.555	-7.284	0.813
32	0.544	-7.355	0.800
33	0.551	-7.401	0.824
34	0.562	-7.444	0.800