



Mapping the AI Product Ecosystem: An Empirical Analysis of Market Concentration, Organizational Diversity, and Pricing Metadata Limitations

Pit Pichappan
Digital Information Research Labs
Chennai. Tamil Nadu. India
pichappan@dirf.org

ABSTRACT

The rapid proliferation of artificial intelligence has created a complex ecosystem of products, platforms, and organizations, yet empirical understanding of its structural characteristics remains limited. This study empirically maps the observed AI product ecosystem by examining 3,027 product records, focusing on market concentration, organizational diversity, startup-product relationships, and pricing metadata completeness. The analysis reveals a dual structural pattern. Within the subset of products with identifiable organizational relationships (42 products across 17 startups), product representation is highly concentrated: four companies account for 57.14% of linked products, with a Herfindahl Hirschman Index of 1,281 and a Gini coefficient of 0.434. However, the ecosystem simultaneously demonstrates organizational diversity, with a Shannon diversity index of 2.450 and Simpson diversity measure of 0.872. Critically, the study identifies substantial metadata limitations: only 1.39% of product records contain identifiable company associations, while 97.72% lack explicit pricing classifications. These findings demonstrate that ecosystem structure cannot be assessed independently of data quality observed concentration among major companies coexists with broad organizational participation, but incomplete entity linkage and pricing metadata severely constrain inference. The study contributes methodologically by distinguishing between substantive ecosystem characteristics and data quality artifacts, and empirically by characterizing the AI product ecosystem as organizationally diverse yet product concentrated. Results highlight the need for improved entity resolution, standardized pricing taxonomies, and multi source data integration to support robust future analyses of the evolving AI product landscape.

Keywords: Artificial Intelligence Ecosystem, Product Concentration, Organizational Diversity, Market Structure, Pricing Metadata, Entity Linkage, Data Quality, Ecosystem Mapping

Received: 12 May 2026, Revised: 10 July 2026, Accepted: 27 July 2026

Copyright: DLINE

1. Introduction and Background

Artificial intelligence (AI) is rapidly becoming a foundational infrastructure of modern society. However, much existing analysis of AI development has focused primarily on technological capability growth, model performance, and investment flows, while comparatively little attention has been given to the broader societal values embedded in AI systems [1].

The expanding application of AI across diverse sectors further demonstrates its growing economic and societal significance. Recent studies confirm AI's strong applicability in digital soil mapping [2], urban planning [3], agriculture [4], industrial process optimisation [5, 6], and energy systems [7]. Researchers have also paid increasing attention to AI's role in mitigating climate risks [8], supporting the digital transformation of food systems [9, 10], and contributing to ecological meteorology and ecosystem services monitoring [11, 12]. Furthermore, there is growing interest in integrating AI into carbon flow assessment, environmental risk prediction, and the adaptation of agroecosystems to climate change [13, 14, 15].

These expanding applications illustrate that AI is no longer confined to isolated technological domains. Instead, its development and deployment increasingly depend on interconnected infrastructures, organizations, data resources, platforms, and complementary technologies. Understanding AI therefore requires attention not only to individual applications and technical capabilities but also to the broader ecosystems within which AI products and services are developed, distributed, and commercialized.

1.1 Digital Platforms and the Emergence of AI Ecosystems

The growing importance of AI coincides with the increasing significance of digital platforms in contemporary economic systems. In platform economy geography, recent scholarly contributions [16, 17, 18, 19, 20] have highlighted the profound transformations generated by digital platforms. These transformations extend across market competition, the organization of global value networks, and the structure of work processes [21]. Consequently, digital platforms exert substantial influence over power dynamics [22] (Bearson et al. 2021), affecting economic and organizational relationships both within individual nations and across international borders while simultaneously reshaping the physical and conceptual spaces in which economic activity occurs.

Within this broader transformation, an emerging stream of research has begun to examine the place based ecosystems in which AI innovations emerge and develop. This literature builds on the contextual turn in management research and, more specifically, on scholarship concerning entrepreneurial ecosystems. Nevertheless, because AI ecosystem research remains relatively nascent, the literature is fragmented across academic and practitioner boundaries, as well as across previously disconnected disciplines. The presence of disparate and insufficiently defined concepts has contributed to a disorganized body of knowledge in which important insights remain latent [23].

The increasing strategic importance of AI also reinforces the need to understand these ecosystems more systematically. AI applications have become an important mechanism through which enterprises seek to enhance value creation and strengthen competitive positioning. However, existing literature provides limited direct empirical analysis of how data elements influence AI adoption. Addressing this gap requires more systematic approaches to conceptualizing and measuring both the data resources underlying AI systems and the AI applications that emerge from them [24].

1.2 The Multi-layer Structure of Generative AI Platform Ecosystems

The ecosystem perspective becomes particularly important in the context of generative AI. Digital platform ecosystems built on generative AI cannot be understood as single or homogeneous markets. Instead, they form complex, interconnected value chains across multiple layers, including semiconductors, cloud infrastructure, foundation model development, inference services, platform orchestration, and applications. These layers differ substantially in capital requirements, barriers to entry, and value capture mechanisms. Consequently, the economic structure of the AI ecosystem is inherently heterogeneous, with different types

of organizations occupying distinct positions and performing complementary functions. Research examining this networked business ecosystem has analyzed its economic structure through a systematic literature review of 36 studies, evaluated six structural hypotheses, and advanced three formal propositions that extend platform theory and complementary-asset frameworks to multi-layer AI ecosystems [25].

This layered structure has important implications for analysing AI product ecosystems. Market concentration, organizational diversity, technological interdependence, and access to critical infrastructure may vary substantially across different layers of the ecosystem. Accordingly, an empirical mapping of AI products and organizations must move beyond treating the AI sector as a single, unified market and instead account for the diverse actors, relationships, and forms of value creation embedded within the broader ecosystem.

1.3 Challenges in Mapping AI and Business Ecosystems

Despite the growing importance of ecosystem analysis, conventional approaches to mapping complex ecosystems still face substantial methodological limitations. Ecosystem mapping frequently relies on Social Network Analysis (SNA), reflecting the inherently interconnected nature of network structures [26-28]. Such approaches typically depend on manual or semi automated content analysis to identify relevant actors and relationships.

However, existing ecosystem mapping approaches are frequently time and labor intensive, even when applied to relatively small datasets [29, 30]. These limitations restrict their scalability and may reduce their practical appeal to executives and other decision makers. In addition, the subjectivity associated with analysts' interpretations, the absence of standardized taxonomies, and limitations in reproducibility and validity further constrain the utility of traditional Business Ecosystem Theory (BET) based approaches in academic research [31].

These challenges are particularly significant in rapidly evolving sectors such as AI, where new organizations, products, technologies, and platform relationships emerge continuously. The complexity and dynamism of these environments make manual ecosystem mapping increasingly difficult and create a need for more scalable, systematic, and reproducible analytical approaches.

1.3.1 AI and Natural Language Processing for Ecosystem Mapping

Recent research has begun to explore AI as a tool for mapping complex business ecosystems. One stream of research explores using artificial intelligence to map business ecosystems, particularly in the context of the energy transition. This context is especially relevant because the energy transition is a globally significant phenomenon that requires interactions among multiple industries, nations, technologies, innovations, and regulatory frameworks [32].

Similarly, Ribeiro [31] proposed an innovative methodology that harnesses Natural Language Processing (NLP) techniques to systematically identify digital platform companies globally. Such approaches demonstrate the potential of computational methods to reduce dependence on manual data collection and interpretation while enabling ecosystem analysis at larger scales.

AI-enabled ecosystem mapping can also generate insights into relationships among established organizations and emerging technological clusters. By detailing relationships between leading entities and emerging technology clusters, N. Mohammadi provided a comprehensive view of the retail innovation landscape. This approach supported stakeholder decision making related to technology adoption, collaborative opportunities, and competitive positioning within an increasingly digital retail ecosystem [33].

Together, these studies demonstrate the potential of AI and NLP techniques to improve the scalability, consistency, and analytical depth of ecosystem mapping. They also provide a foundation for examining AI product ecosystems as complex organizational and technological networks rather than as collections of independent firms or products.

1.4 Metadata and the Limits of AI Product Ecosystem Analysis

Despite advances in computational approaches to ecosystem mapping, the quality and availability of structured metadata remain important constraints. A fundamental tension exists between the persistent costs associated with producing and refining structured metadata under conditions of deep technological friction and the differentiated demands that distinct communities place on data quality, focus, and granularity.

This tension can be understood through the concept of the innovation annulus, defined as the zone between freely available structured data and the advancing frontier of commercially refined knowledge products [34]. Within this zone, organizations face significant challenges in determining how much information should be standardized, refined, and commercialized, particularly when different users require different levels of detail and data quality.

These metadata limitations are especially relevant for AI product ecosystems, where information regarding products, organizations, capabilities, pricing, market positioning, and technological relationships may be incomplete, inconsistently structured, or commercially restricted. Consequently, the empirical mapping of AI ecosystems depends not only on the availability of advanced analytical techniques but also on the accessibility, quality, and granularity of the underlying metadata.

1.5 Research Context

Taken together, the existing literature highlights several interconnected developments. First, AI is becoming increasingly embedded across economic, environmental, industrial, and societal domains. Second, the development and commercialization of AI increasingly occur through complex digital platform ecosystems characterized by multiple organizational layers and heterogeneous mechanisms of value creation. Third, conventional approaches to ecosystem mapping face important limitations related to scalability, subjectivity, reproducibility, and the absence of standardized taxonomies. Finally, although AI and NLP offer promising approaches for systematically mapping complex ecosystems, these methods remain constrained by the availability and quality of structured metadata.

These developments provide a basis for examining the AI product ecosystem empirically. In particular, systematic ecosystem mapping can contribute to a better understanding of market concentration, organizational diversity, the relationships among different AI actors and product categories, and the limitations associated with incomplete or unevenly structured pricing and product metadata. This approach can reveal the broader structure of the AI marketplace while also identifying the methodological and informational constraints that shape empirical analyses of rapidly evolving digital ecosystems.

2. Research Statement

The rapid expansion of artificial intelligence has produced a complex and increasingly interconnected ecosystem of products, startups, digital platforms, technologies, and complementary infrastructures. Although AI is now widely recognized as an important technological and economic infrastructure, empirical understanding of the structure of the AI product ecosystem remains limited. Existing research has primarily focused on AI capabilities, technological performance, adoption, investment, or specific application domains, while comparatively less attention has been directed toward systematically mapping the relationships between AI products and the organizations that develop or provide them.

The empirical examination of AI ecosystems is further complicated by the fragmented and uneven nature of available data. Product information is distributed across multiple digital platforms and directories, while key metadata, including company affiliation, pricing model, technology classification, and market characteristics, may be incomplete or inconsistently represented. Conventional ecosystem mapping approaches based on manual or semi automated analysis are also difficult to scale and may be affected by subjectivity, limited reproducibility, and the absence of standardized taxonomies [26–31].

The central research problem addressed in this study is therefore the limited empirical understanding of how AI products are distributed across organizations and the extent to which available product metadata can support reliable analysis of market structure, organizational diversity, product concentration, and pricing models. In particular, incomplete company-product entity linkage and substantial gaps in pricing metadata create uncertainty regarding whether observed patterns represent actual characteristics of the AI market or limitations in the underlying data.

Accordingly, this study empirically maps the observed AI product ecosystem using a structured multi-source dataset containing AI startups, products, research, employment, news, and entity relationships. The study focuses specifically on the product company relationships available within the dataset and examines the distribution of products across identified organizations, the degree of concentration and inequality in product representation, organizational diversity, and the completeness of pricing metadata. The study therefore contributes to AI ecosystem research by distinguishing between substantive ecosystem characteristics and limitations arising from incomplete metadata.

The research statement can be summarized as follows:

This study investigates the structural characteristics of the observed AI product ecosystem by examining product concentration, organizational diversity, startup product relationships, and pricing model metadata, while explicitly assessing how incomplete entity linkage and missing pricing information constrain the interpretation of empirical ecosystem patterns.

2.1 Research Issues

The study addresses several interconnected research issues associated with empirical AI ecosystem analysis.

2.1.1 Limited visibility of AI product-company relationships

A fundamental issue concerns the incomplete identification of relationships between AI products and the organizations associated with them. Of the 3,027 product records available in the dataset, only 42 records contain an identifiable startup or company association. Consequently, most products cannot be directly linked to a specific organization. This creates uncertainty regarding the extent to which the observed organizational structure represents the broader AI product ecosystem.

2.1.2 Unequal product representation among identified organizations

Within the subset of products for which organizational information is available, product representation is unevenly distributed. A small number of major companies account for a substantial proportion of the linked products, whereas many organizations are represented by only one or a few products. The research issue is therefore whether the observed AI ecosystem demonstrates meaningful organizational concentration despite the presence of multiple participating companies.

2.1.3 The coexistence of concentration and diversity

Market ecosystems cannot be adequately characterized using a single structural measure. An ecosystem may contain many participating organizations while simultaneously exhibiting substantial concentration in the distribution of products. This creates an important analytical issue concerning the relationship between organizational diversity and product concentration. The present study therefore examines whether the AI product ecosystem is simultaneously diverse in organizational participation and concentrated in product representation.

2.1.4 Incomplete pricing metadata

A major research issue concerns the availability and reliability of product pricing information. The analysis shows that 2,958 of 3,027 products are classified as OTHER/UNKNOWN, representing 97.72% of the product dataset. Consequently, the available information is insufficient to establish the actual prevalence of specific pricing strategies such as freemium, subscription, open-source, paid, or free models.

2.15 Distinguishing market structure from metadata structure

An important methodological issue is the possibility that observed empirical patterns may partly reflect the structure and quality of the underlying data rather than the actual structure of the AI market. For example, the large OTHER/UNKNOWN pricing category should not be interpreted as evidence that most AI products lack established pricing models. Similarly, concentration among identified companies cannot automatically be generalized to all AI products because the overwhelming majority of product records lack company attribution.

2.16 Single-source limitations

Although the broader dataset integrates multiple types of AI ecosystem information, all product records used in the pricing analysis originate from a single source, AI123Box. This prevents meaningful statistical comparison across independent product directories and raises the possibility that observed pricing patterns are influenced by source specific data collection and categorization practices.

3. Research Gaps

The study addresses four principal gaps in the existing literature.

3.1 Gap in empirical mapping of AI product ecosystems

Existing AI research has extensively examined technological capabilities, model performance, adoption, investment, and application across multiple sectors. However, there remains limited empirical research systematically examining the structural organization of the AI product ecosystem, particularly the distribution of products across organizations and the relationships between product representation and organizational participation.

Previous research recognizes the importance of AI ecosystems and multi-layer platform structures, but empirical evidence regarding the organizational structure of product-level AI ecosystems remains limited. This study addresses this gap by examining product-company relationships in an empirical AI ecosystem dataset.

3.2 Gap in the simultaneous measurement of concentration and diversity

Existing ecosystem analyses frequently emphasize either market concentration or the diversity of participating actors. These perspectives are complementary but should not be treated as mutually exclusive. An ecosystem can contain numerous organizations while product activity remains concentrated among a relatively small group.

The present study addresses this analytical gap by combining multiple complementary indicators, including the Herfindahl Hirschman Index (HHI), Gini coefficient, Shannon diversity index, Simpson diversity measure, cumulative concentration shares, and the startup to product ratio. This multidimensional approach provides a more comprehensive characterization of ecosystem structure than reliance on a single concentration or diversity measure.

3.3 Gap concerning metadata quality in AI ecosystem research

Research on AI ecosystems has increasingly recognized the potential of computational and Natural Language Processing approaches for ecosystem mapping. However, less attention has been given to the extent to which incomplete metadata constrains the validity of empirical ecosystem analysis.

The present study addresses this gap by treating missing pricing and entity information as substantive methodological findings rather than simply removing incomplete observations. The analysis explicitly evaluates how metadata incompleteness affects the interpretation of product concentration, organizational

diversity, and pricing structures.

3.4 Gap between observed data and actual market characteristics

A further gap concerns the distinction between patterns observed in available datasets and the underlying characteristics of the actual AI market. Incomplete data may produce apparent patterns that are artifacts of classification practices, source coverage, or entity resolution limitations.

This study addresses this gap by explicitly distinguishing between the observed AI product ecosystem represented in the dataset and the broader global AI product market. The findings are therefore interpreted as descriptive evidence concerning the observed and linked dataset rather than as definitive estimates of global AI market structure.

This study adopts a quantitative, descriptive, and exploratory empirical research design.

The quantitative component is appropriate because the study analyzes structured product and organizational records using numerical indicators of distribution, concentration, inequality, diversity, and metadata completeness. The descriptive component characterizes the composition and structure of the observed AI product ecosystem without attempting to establish causal relationships. The exploratory component is appropriate because systematic empirical research on the structural characteristics of AI product ecosystems remains relatively underdeveloped and because the study investigates patterns emerging from a large, heterogeneous ecosystem dataset.

The study follows an ecosystem mapping research design in which AI products constitute the principal units of observation and identified startups or companies constitute the organizational entities to which products are linked. The analysis therefore examines the structure of the ecosystem through the distribution of product records across identified organizations.

The research design is organized around three analytical dimensions:

- 1. Product metadata structure**, focusing on the availability and distribution of pricing information.
- 2. Organizational structure**, focusing on relationships between AI products and identified startups or companies.
- 3. Ecosystem structure**, focusing on concentration, inequality, diversity, and cumulative product representation.

The study does not attempt to estimate causal effects or test whether particular organizational characteristics cause differences in pricing or product development. Instead, the purpose is to empirically characterize the observed structure of the AI product ecosystem and identify the limitations that incomplete metadata impose on interpretation.

The unit of analysis is primarily the AI product record. However, for organizational concentration and diversity analyses, the unit of analysis is aggregated to the startup/company level, based on identifiable product-company relationships.

The study also adopts a data-quality-aware analytical design. This is important because missing data are not treated as a minor technical problem. Instead, the completeness of pricing and entity linkage information is itself analyzed as an important characteristic affecting the reliability and interpretability of ecosystem level conclusions.

3.5 Research Method

3.5.1 Data source and dataset

The study uses a structured multi source dataset representing the AI ecosystem. The dataset integrates information relating to AI startups, products, research publications, employment opportunities, news, and entity relationships. It contains six CSV files, 45 variables, and more than 6,000 records. The dataset provides a structured representation of the AI landscape and supports quantitative and empirical analysis.

The principal files include:

- startups.csv, containing information on AI startup companies;
- products.csv, containing 3,027 AI product and service records;
- research_papers.csv, containing academic and technical AI publications;
- jobs.csv, containing AI-related employment information;
- news.csv, containing recent AI industry news; and
- entity_mapping_log.csv, containing relational mappings among ecosystem entities.

The present empirical analysis focuses primarily on the products.csv data and its available organizational linkage information.

3.5.2 Data collection and processing

The underlying dataset was compiled from publicly accessible web sources, with original URLs retained to support data provenance and traceability. The data preparation process involved systematic data engineering procedures, including data cleaning, normalization, deduplication, and schema validation.

For the present study, the analytical dataset was processed through the following stages:

- 1. Data selection:** Product-level records were selected as the primary analytical population.
- 2. Variable identification:** Pricing model and startup/company fields were identified as the principal variables for structural ecosystem analysis.
- 3. Data completeness assessment:** Missing and unclassified observations were quantified before substantive interpretation.
- 4. Entity linkage:** Products containing identifiable startup or company information were extracted for organizational analysis.
- 5. Aggregation:** Linked product records were aggregated by startup/company to determine the number of products associated with each organization.
- 6. Classification:** Pricing information was categorized into FREE, FREEMIUM, PAID, SUBSCRIPTION, OPEN SOURCE, and OTHER/UNKNOWN.
- 7. Validation:** Descriptive totals and percentages were examined for internal consistency.

This processing approach ensured that the analysis distinguished between valid substantive observations and missing or incomplete metadata.

3.5.3 Pricing-model analysis

The first stage of analysis examined the distribution of product pricing classifications across all 3,027 product records.

We calculated frequency counts and percentages for each pricing category. The primary objective was to determine how much pricing information was available and sufficiently structured for ecosystem analysis. Because 97.72% of products were classified as OTHER/UNKNOWN, the analysis interpreted this result primarily as evidence of limited metadata completeness rather than as evidence concerning the actual distribution of AI industry pricing strategies.

The pricing analysis was therefore conducted as both:

- a descriptive analysis of observed pricing classifications; and
- an assessment of pricing-metadata completeness.

3.5.4 Startup-product relationship analysis

The second stage examined relationships between AI products and identified startups or companies.

Only product records containing an identifiable organizational association were included in this analysis. The linked records were grouped by company, and the number of associated products was calculated for each organization.

The following descriptive indicators were used:

- number of identified startups;
- number of linked products;
- products per startup;
- company level product counts; and
- percentage share of linked products.

3.6 Product concentration analysis

Product concentration was examined to determine the extent to which linked AI products were distributed unequally among identified organizations.

The analysis used cumulative concentration shares, including the proportion of linked products represented by the largest four, five, and ten companies.

The Herfindahl–Hirschman Index (HHI) was calculated as:

$$[\text{HHI} = \sum_{i=1}^n s_i^2]$$

where (s_i) represents the proportional share of linked products associated with startup (i).

The HHI was used to quantify the degree to which product representation was concentrated among a limited number of organizations.

3.7 Inequality analysis

The Gini coefficient was used as a complementary measure of inequality in the distribution of products across identified startups.

The coefficient ranges from zero, representing perfect equality, to one, representing maximum inequality. In the present study, the Gini coefficient was used to determine whether linked products were distributed relatively evenly among participating organizations or concentrated disproportionately among a smaller group of companies.

Using both the HHI and the Gini coefficient enabled the analysis to distinguish concentration from broader inequality in product representation.

3.8 Organizational diversity analysis

Organizational diversity was assessed using the Shannon diversity index and Simpson diversity measure.

The Shannon diversity index was calculated as:

$$[H = - \sum_{i=1}^n p_i \log(p_i)]$$

where (p_i) represents the proportion of linked products associated with startup (i) .

The Simpson diversity measure was used to provide a complementary assessment of organizational diversity based on the distribution of products across companies.

These measures were interpreted alongside concentration indicators because organizational diversity and product concentration can coexist. An ecosystem may contain numerous participating organizations while still exhibiting substantial inequality in the distribution of products among them.

3.9 Startup-level pricing analysis

For products containing both startup/company and pricing information, pricing classifications were cross-tabulated by organization.

The objective was to identify descriptive variation in observed pricing classifications across startups.

However, inferential statistical testing was not used because the number of observations per company was extremely small and the dominance of the OTHER/UNKNOWN category would create substantial expected-frequency problems. Accordingly, startup level pricing patterns were interpreted descriptively rather than as statistically established differences in corporate pricing strategies.

3.9 Visualization

Four principal visualizations were used to support interpretation:

- Pricing distribution chart, showing the dominance of the OTHER/UNKNOWN category;
- Startup product count ranking, showing the distribution of linked products across companies;
- Cumulative concentration curve, illustrating the proportion of products accounted for by leading organizations; and
- Startup level pricing visualization, illustrating observed heterogeneity in pricing classifications.

The visualizations complement the numerical indicators by making the structural patterns of concentration, diversity, and metadata incompleteness more readily interpretable.

3.10 Interpretation and analytical boundaries

The study adopts a cautious interpretation strategy because of two major limitations: incomplete company-product entity linkage and incomplete pricing metadata.

Consequently, the concentration and diversity results reflect the linked subset of the dataset, whereas the pricing analysis is interpreted primarily as evidence of metadata availability and classification quality.

The findings are therefore not generalized directly to the complete global AI market. Instead, the study provides a descriptive empirical characterization of the AI product ecosystem represented in the available dataset and demonstrates how data completeness influences ecosystem level inference.

4. Dataset

This dataset presents a structured, multi-source compilation of the artificial intelligence (AI) landscape, integrating information on startups, products, research, employment, and news. Designed to support quantitative analysis and empirical research, it provides a comprehensive snapshot of the AI ecosystem as of 2026 and addresses the fragmentation of publicly available industry data.

4.1 Data Collection and Processing

Data were aggregated from publicly accessible web sources, with original URLs retained to ensure provenance and traceability. The compilation followed a systematic data engineering pipeline encompassing data cleaning, normalization, deduplication, and schema validation, resulting in a consistent and analyzable structure.

5. Results and Discussion

5.1 AI Product Pricing-Model Distribution

The analysis of the AI product ecosystem identified 3,027 product records in the product dataset. However, the distribution of pricing models reveals a substantial limitation in pricing metadata availability. As presented in Table 1, 2,958 products, corresponding to 97.72% of the complete product dataset, were classified as OTHER/UNKNOWN. Only 60 products (1.98%) were explicitly identified as FREE, while nine products (0.30%) were classified as PAID. No products were explicitly classified as FREEMIUM, SUBSCRIPTION, or OPEN SOURCE. Consequently, only 69 of the 3,027 products, or 2.28%, had an explicitly identifiable pricing model.

Pricing-model analysis

Pricing model	Products	Percentage
FREE	60	1.98%
FREEMIUM	0	0.00%
PAID	9	0.30%
SUBSCRIPTION	0	0.00%
OPEN SOURCE	0	0.00%
OTHER/UNKNOWN	2,958	97.72%
Total	3,027	100%

Table 1. Pricing Model

This distribution is important because the very large OTHER/UNKNOWN category should not be interpreted as evidence that AI products predominantly operate without established pricing structures. Rather, it indicates that pricing information was unavailable, incompletely represented, or not sufficiently standardized in the underlying source data. The results therefore demonstrate a clear distinction between the *observed pricing distribution* and the *actual pricing strategies of AI products*. While the observed data contain slightly more FREE products than PAID products, the difference is too small and the amount of missing information too large to support a substantive conclusion that free pricing dominates the AI product ecosystem.

The absence of FREEMIUM, SUBSCRIPTION, and OPEN SOURCE classifications is similarly more appropriately interpreted as a limitation of the available metadata than as evidence that these business models are absent from the AI product market. In particular, subscription based and freemium models are widely applicable to digital and AI services, meaning that their complete absence from the classification strongly suggests that the source taxonomy did not adequately capture these pricing arrangements. The pricing analysis should therefore be understood primarily as an assessment of pricing information availability and classification, rather than as a definitive measurement of AI industry pricing strategy.

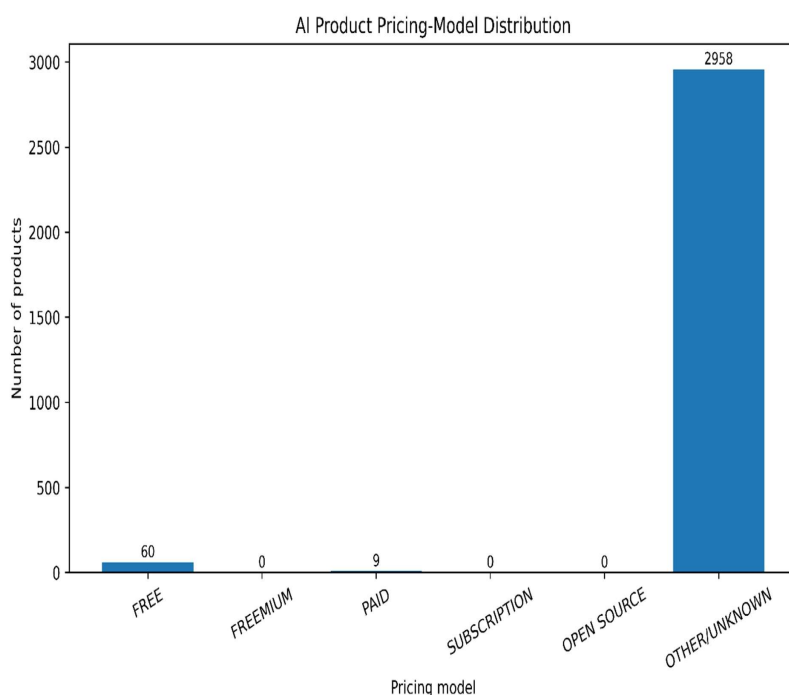


Figure 1. Pricing Distribution

Figure 1 visually represents the pricing distribution and makes the dominance of the OTHER/UNKNOWN category immediately apparent. The large gap between the UNKNOWN category and all explicitly classified categories shows that the main characteristic of the pricing data is not variation among pricing models but a lack of sufficiently granular pricing information. The small FREE and PAID categories appear visually as minor components of the overall product population. Thus, Figure 1 reinforces the conclusion that pricing based comparisons should be interpreted cautiously and that future data collection should prioritise systematic

Table 3 provides the same pricing information in percentage form and confirms this pattern. FREE products account for 1.982% of the dataset, PAID products for 0.297%, and OTHER/UNKNOWN products for 97.721%, while FREEMIUM, SUBSCRIPTION, and OPEN SOURCE each account for 0%. The close correspondence between the counts and percentages confirms the internal consistency of the pricing distribution. The result also establishes that more than 97% of the product observations cannot currently be assigned to one of the substantive pricing categories considered in the analysis.

Pricing Model Interpretation

Pricing model	Products	Percentage
Free	60	1.98%
Freemium	0	0
Paid	9	0.297%
Subscription	0	0
Open Source	0	0
Other/Unknown	2,958	97.72%

Table 3. Pricing Model Interpretation

5.2 Pricing-Model Differences Across Startups

The startup-level pricing analysis provides a more detailed view of the 42 product records with a startup or company name. These 42 linked products were associated with 17 unique startups. Table 2 demonstrates considerable descriptive variation in pricing classifications across these companies, although the small number of observations per company prevents reliable statistical inference.

Among the 42 products with an identified startup, the distribution is:

Startup	Products	Free	Paid	Other/ Unknown
Google	12	2	0	10
OpenAI	6	1	2	3
Anthropic	3	1	0	2
Stability AI	3	1	0	2
ByteDance	2	0	0	2
xAI	1	0	1	
Runway	2	1	0	1
DeepSeek	2	2	0	0
Kuaishou	2	1	0	1
Other identified Startups	8	8	0	0

Table 2. Pricing Model by Startup

Google accounted for the most linked products, with 12. Of these, two were classified as FREE and ten as OTHER/UNKNOWN. Thus, 83.33% of Google's linked products had no explicit pricing classification, while 16.67% were identified as FREE. OpenAI was associated with six products: one FREE, two PAID, and three OTHER/UNKNOWN. Consequently, OpenAI had the broadest observed pricing variation among the major companies represented in the table, with FREE, PAID, and UNKNOWN classifications all present.

Anthropic and Stability AI each contributed three linked products. In both cases, one product was classified as FREE and two as OTHER/UNKNOWN. ByteDance contributed two products, both classified as OTHER/UNKNOWN. xAI, Runway, and Kuaishou each contributed two products, with one FREE and one OTHER/UNKNOWN. DeepSeek also contributed two products, both classified as FREE, giving it the highest observed proportion of explicitly FREE products among companies with more than one linked product. The remaining eight identified products, grouped as other identified startups, were classified as FREE.

The results therefore indicate descriptive differences in pricing classification across startups, but these differences should not be interpreted as statistically established differences in business models. Several startups have only one or two linked products, creating extremely small cell counts. Furthermore, the dominance of the OTHER/UNKNOWN category means that the effective sample available for comparing explicit pricing models is very small. Under these circumstances, a conventional chi-square test of startup by pricing model would have substantial expected frequency problems. The absence of a formal inferential claim is therefore methodologically appropriate.

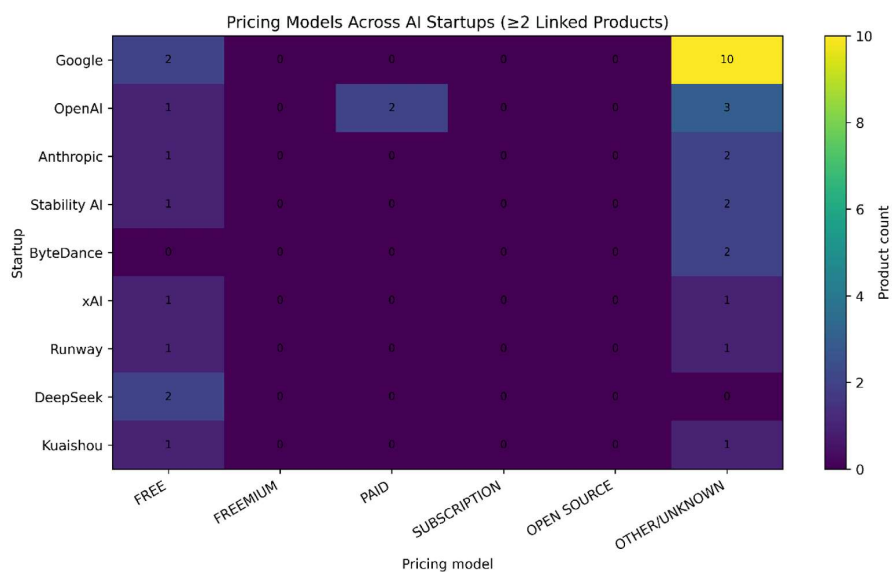


Figure 4. Pricing model by startup

Figure 4 complements Table 2 by visually displaying pricing classifications across startups. The figure shows that Google has the most linked products, but most remain unclassified by pricing. OpenAI displays greater pricing heterogeneity because its products include both FREE and PAID classifications alongside UNKNOWN observations. DeepSeek provides a contrasting pattern because both linked products are classified as FREE. Anthropic and Stability AI occupy intermediate positions, each having one FREE and two UNKNOWN products. The visual pattern therefore supports the interpretation that startup level pricing heterogeneity exists in the observed data, but incomplete classification strongly constrains it.

An additional observation concerns Midjourney. The results identify its single linked product as PAID. Because only one product is represented for this company, this observation is useful descriptively but cannot be generalized to the company's broader product portfolio or used as evidence of a statistically meaningful company level pricing strategy.

5.3 Startup–Product Relationships

The startup–product relationship analysis reveals a substantial limitation in entity linkage but also identifies a meaningful concentration pattern within the subset of products for which company information is available. Of the 3,027 product records, only 42, or 1.39%, contain a populated startup/company name. Conversely, 2,985 records, representing 98.61% of the product dataset, do not have an identified startup association. This means the startup-level results describe the linked product subset, not the full AI product population.

The 42 linked products correspond to 17 unique startups, averaging 2.47 linked products per identified startup. This value indicates that, among products with company attribution, each identified startup is associated with about 2.5 products on average. However, the mean masks a strongly unequal distribution. Product representation is concentrated among a small number of companies rather than being evenly distributed across the 17 startups.

As shown in the startup ranking, Google is the most prolific company in the linked dataset, with 12 products. Google alone therefore accounts for 28.57% of all startup linked products. OpenAI follows with six products, representing 14.29% of the linked product population. Anthropic and Stability AI each have three products, while ByteDance, xAI, Runway, DeepSeek, and Kuaishou each have two products. The remaining ten-to-seventeen ranked positions collectively include eight other identified startups, each with one product.

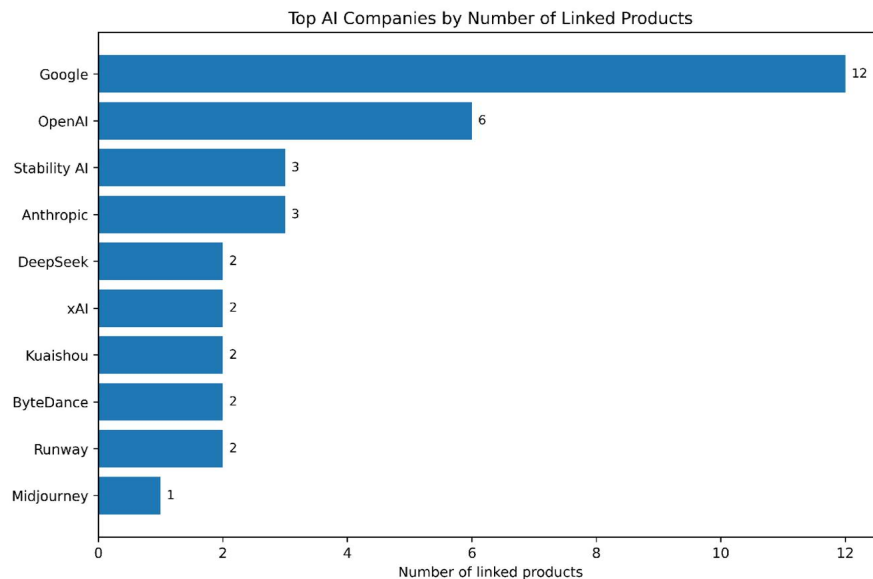


Table 2. Top startups

Figure 2 clearly illustrates this unequal distribution. Google has a substantially longer product count than the other companies, followed by OpenAI, while the remaining companies form a much smaller group. The figure therefore demonstrates that the linked AI product ecosystem is characterized by a long tailed distribution, in which a limited number of companies account for a disproportionately large number of identified products and many companies have only a single linked product.

5.4 Product Concentration

The concentration analysis further confirms the unequal distribution of products among startups. The four largest companies Google, OpenAI, Anthropic, and Stability AI—collectively account for 24 of the 42 linked products. Their combined share is therefore 57.14%. This means that more than half of all startup linked products are associated with only four companies.

Concentration increases as more companies are included. The five largest startups account for 61.90% of

linked products, while the ten largest startups account for 83.33%. Thus, the majority of the observed product ecosystem is concentrated among a relatively small group of companies. The remaining startups collectively account for only 16.67% of linked products. This pattern demonstrates that the ecosystem contains considerable diversity in the number of participating companies, but product representation is highly uneven. The Herfindahl Hirschman Index (HHI) quantifies this concentration. The startup level HHI is 0.1281, equivalent to approximately 1,281 on the conventional 0–10,000 scale. This value is consistent with a moderately concentrated distribution within the linked product subset. The HHI should nevertheless not be interpreted as a measure of concentration across the entire AI product market because the calculation is based on only 42 products with identified startup associations.

The Gini coefficient provides a complementary measure of inequality in product representation and is approximately 0.434. This value indicates a noticeable degree of inequality in how products are allocated among startups. A perfectly equal distribution would yield a value of zero, while higher values indicate greater inequality. The observed value therefore confirms that products are not evenly distributed across the 17 identified companies. Instead, companies such as Google and OpenAI account for far more products, while startups are represented by only one product.

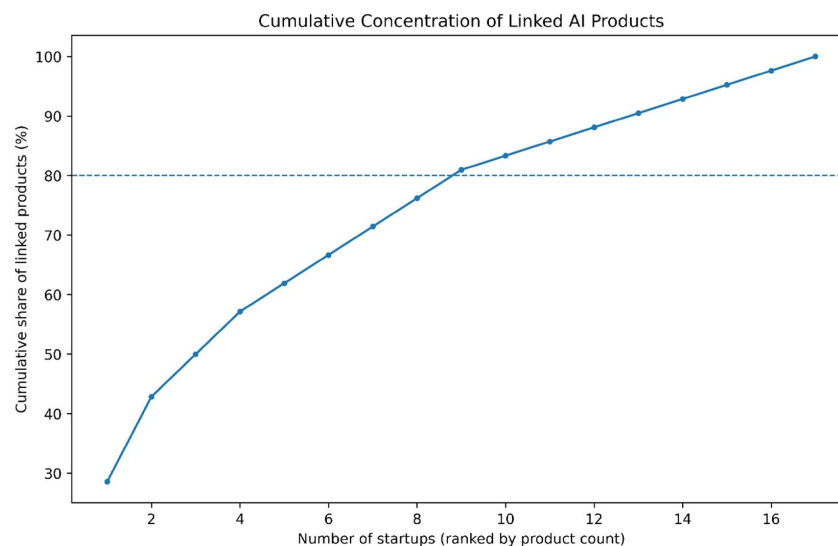


Figure 3. Product Concentration

Figure 3 provides the cumulative concentration curve and visually interprets this inequality. The curve rises relatively rapidly because a relatively small number of highly represented startups contribute a large proportion of the linked products. The figure consequently reinforces the numerical finding that the product ecosystem is concentrated rather than evenly distributed. In substantive terms, the ecosystem appears to contain a combination of highly diversified product organizations and a much larger number of companies represented by only a small number of products.

5.5 Product Ecosystem Diversity

Although the concentration measures demonstrate inequality, the diversity measures show that the ecosystem is not restricted to a small number of companies. Seventeen unique startups are represented among the 42 linked products. The Shannon diversity index is 2.450, while the Simpson diversity measure is 0.872. These values indicate a relatively diverse set of participating startups while also reflecting unequal product contributions.

The apparent contrast between diversity and concentration is substantively important. A product ecosystem can contain many participating companies without distributing products equally among them. That is precisely

the pattern observed here. The 17 identified startups indicate a relatively broad range of organizational participation, whereas the concentration measures show that a small group of companies contributes a disproportionate share of the products.

The combination of the Shannon and Simpson measures with the HHI and Gini coefficient therefore provides a more complete characterization than any single metric. Shannon diversity captures the variety of participating startups, Simpson diversity reflects the probability that randomly selected products belong to different startups, HHI emphasizes concentration, and the Gini coefficient captures inequality in product representation. Taken together, the measures indicate an ecosystem that is organizationally diverse but product concentrated.

Figure 3 is particularly useful in explaining this relationship because it transforms the numerical concentration results into a cumulative representation. The steep early rise in cumulative product share indicates that the leading startups contribute disproportionately to the total number of linked products. The relatively flatter portion associated with the remaining startups reflects the large number of companies contributing relatively few products.

5.6 Startup-to-Product Relationship

The startup-to-product ratio provides a concise representation of the observed organizational structure. Within the linked subset, 42 products are associated with 17 startups, for 2.47 products per startup, or about one startup for every 2.47 products. This ratio best characterises the linked ecosystem because both the numerator and denominator refer to observations with an established company association.

A second calculation would divide 3,027 total product records by 17 identified startups, yielding approximately 178.06 product records per identified startup. However, this value should not be interpreted as the actual average product portfolio of the 17 companies. The calculation is statistically inappropriate because 2,985 of the 3,027 products have no startup assignment. Consequently, attributing all 3,027 products to the 17 identified companies would artificially inflate the startup level product count. The meaningful ratio remains 2.47 linked products per identified startup.

The difference between the two ratios highlights the importance of entity completeness in ecosystem analysis. The 1.39% linkage rate indicates that company-product relationships are sparsely represented in the available product metadata. Therefore, the observed concentration among Google, OpenAI, Anthropic, Stability AI, and other identified startups may accurately describe the linked portion of the dataset but cannot automatically be generalized to the entire population of 3,027 products.

5.7 Source-Level Pricing Analysis

The source analysis reveals another important limitation. All 3,027 product records originate from a single source, AI123Box. Because there is no meaningful source level variation, it is not possible to statistically compare pricing distributions across different product directories or data sources.

This limitation is particularly relevant to the original objective of determining whether pricing models differ by source. A source comparison requires at least two independent sources containing sufficiently large and comparable product samples. The current dataset does not meet this requirement. Consequently, we cannot draw a statistically defensible conclusion about source specific pricing preferences.

The single source structure also suggests that the observed pricing distribution may partly reflect the data collection and categorization practices of AI123Box. The 97.72% OTHER/UNKNOWN rate may therefore represent source specific metadata coverage rather than a characteristic of the wider AI product market. Future research should incorporate multiple product directories, official company websites, software marketplaces, and product documentation to improve cross source validity.

5.8 Integrated Interpretation of the AI Product Ecosystem

Altogether, the results reveal two central characteristics of the observed AI product ecosystem: high organizational concentration among linked products and substantial incompleteness of product metadata. The concentration results show that four companies account for 57.14% of linked products, five companies account for 61.90%, and ten companies account for 83.33%. Google is the dominant company in the linked dataset, contributing 28.57% of linked products, followed by OpenAI with 14.29%. The HHI of approximately 1,281 and Gini coefficient of approximately 0.434 further demonstrate that product representation is unevenly distributed.

At the same time, the ecosystem cannot be characterized as lacking diversity. Seventeen unique startups are represented among the 42 linked products, with a Shannon diversity index of 2.450 and Simpson diversity of 0.872. These measures demonstrate that a range of companies participate in the observed ecosystem, but their contributions differ substantially in scale. The resulting structure can therefore be characterized as diverse in organizational participation but concentrated in product representation.

The pricing results reveal a different type of limitation. The fact that 2,958 of 3,027 products, or 97.72%, are classified as OTHER/UNKNOWN means the dataset provides insufficient evidence to draw strong conclusions about AI pricing strategies. Only 60 products are explicitly FREE and nine are explicitly PAID, while FREEMIUM, SUBSCRIPTION, and OPEN SOURCE are absent from the classification. The most appropriate interpretation is therefore that pricing metadata are severely incomplete rather than that these pricing models are genuinely absent from the AI ecosystem.

The startup level results provide some evidence of pricing heterogeneity. OpenAI has FREE, PAID, and UNKNOWN classifications; Google is dominated by UNKNOWN classifications with two FREE products; DeepSeek has two FREE products; and several smaller companies have exclusively FREE classifications within the linked records. However, these patterns remain descriptive because many startups have very few observations. Consequently, treat them as evidence of observed variation in metadata classification, not as statistically demonstrated differences in corporate pricing strategies.

The four figures collectively reinforce these conclusions. Figure 1 demonstrates the extreme dominance of the OTHER/UNKNOWN pricing category. Figure 2 highlights Google's leading position in terms of linked product count and the sharp difference between the leading company and most other startups. Figure 3 demonstrates the cumulative concentration of products among the leading companies, while Figure 4 illustrates the heterogeneous but sparsely observed pricing classifications across startups. Together, the figures provide complementary visual evidence for the numerical findings reported in Tables 1–3.

Overall, the analysis suggests that the dataset is much better suited to examining AI product company concentration and ecosystem structure than to making definitive claims about AI product pricing strategies. The principal empirical contribution of the present analysis is therefore the identification of an ecosystem that combines broad organizational participation with substantial product concentration, while simultaneously exhibiting significant limitations in pricing metadata and company-product entity linkage.

5.9 Implications for Data Quality and Future Analysis

The results identify two specific data quality priorities for future AI ecosystem research. First, pricing information should be collected and normalized from primary product sources using a standardized taxonomy capable of distinguishing FREE, FREEMIUM, PAID, SUBSCRIPTION, OPEN SOURCE, and other models. Second, company product entity resolution should be substantially expanded because only 42 of the 3,027 product records currently have an identifiable startup/company association.

Improving entity linkage would substantially strengthen the concentration, diversity, and startup-level analyses. More identified company product relationships would allow more reliable estimation of product portfolios, concentration indices, and company level pricing distributions. Similarly, incorporating multiple independent product sources would enable genuine source level comparisons and reduce the possibility that the observed pricing distribution is an artifact of a single directory.

Accordingly, the present findings should be interpreted as a descriptive empirical characterization of the observed and linked AI product ecosystem, rather than as a complete representation of the global AI product market. Within that limitation, the results provide consistent evidence that product representation is concentrated among a relatively small group of major AI companies, while the participating company base remains comparatively diverse. The analysis also shows that incomplete pricing and entity metadata are important methodological considerations that future AI ecosystem studies should explicitly address.

6. Discussion

6.1 Overview of the Findings

The findings empirically characterise the observed AI product ecosystem and reveal a structural pattern defined by two simultaneous conditions: unequal product concentration among identified organisations and substantial incompleteness in ecosystem metadata. These findings matter because they show that the analytical value of AI ecosystem datasets depends not only on the number of available records but also on the completeness and quality of the relationships and attributes within those records.

The analysis of 3,027 AI product records shows that the dataset contains substantial product level information. However, two major limitations constrain the ability to draw broader ecosystem level conclusions. First, only 42 product records, representing 1.39% of the total product population, contain an identifiable startup or company association. Second, 2,958 products, representing 97.72% of the dataset, are classified as OTHER/UNKNOWN for pricing. These findings demonstrate that the principal challenge in the empirical analysis of this ecosystem is not the absence of product records but the incompleteness of the metadata required to connect products to organizations and commercial characteristics. The results therefore support an important methodological distinction between dataset size and analytical completeness. A large collection of AI products does not automatically provide sufficient information to analyze organizational structure, pricing strategies, or market relationships. For ecosystem research, the availability of entity-level relationships is particularly important because the ecosystem is defined not only by the presence of individual products but also by the organizational and relational structures connecting those products.

6.2 Product Concentration and Organizational Structure

Within the linked subset of the dataset, product representation is concentrated among a relatively small number of organizations. Four companies account for 57.14% of the 42 linked products, while the ten largest companies account for 83.33%. Google represents the largest organization in the linked sample, followed by OpenAI, with the remaining organizations contributing substantially smaller numbers of linked products.

The concentration measures reinforce this pattern. The HHI of approximately 1,281 indicates moderate concentration within the linked subset, while the Gini coefficient of approximately 0.434 shows noticeable inequality in product distribution across the 17 identified companies. Together, these measures indicate that the observed organizational structure is not characterized by an equal distribution of products. Rather, a relatively small number of companies account for a disproportionately large share of the available product-company relationships.

This finding should nevertheless be interpreted carefully. The concentration measures do not represent the concentration of the global AI product market. They represent concentration within the subset of products for which an organizational relationship is available. Because more than 98% of product records lack an identified startup association, the observed dominance of particular companies may partly reflect differences in entity-linkage coverage rather than differences in actual product portfolios.

This distinction is central to interpreting the results. The analysis provides evidence of observed concentration in the linked dataset, but it does not provide sufficient evidence to claim that the global AI product market is concentrated to the same degree. The study therefore demonstrates the importance of explicitly defining the empirical population to which concentration measures apply.

6.3 The Coexistence of Organizational Diversity and Product Concentration

One of the most important findings is the apparent coexistence of organizational diversity and product concentration. The linked dataset contains 17 unique startups, indicating that multiple organizations participate in the observed ecosystem. The Shannon diversity index of 2.450 and Simpson diversity measure of 0.872 further indicate that the ecosystem includes a relatively broad range of organizational participants.

At the same time, the HHI and Gini coefficient show that product representation is unevenly distributed among these organizations. Thus, diversity in organizational participation does not imply equality in product representation. An ecosystem may include numerous participants while a relatively small number of organizations account for a large proportion of observable products.

This finding illustrates the value of combining concentration and diversity measures in ecosystem analysis. A study relying only on the number of participating companies might conclude that the ecosystem is broadly distributed. Conversely, an analysis based only on concentration might overlook the presence of a diverse organizational base. Combining the two perspectives provides a more nuanced interpretation: the observed AI product ecosystem is organizationally diverse but product concentrated.

This characterization also suggests that ecosystem diversity should not be interpreted solely in terms of the number of organizations present. The relative contribution of those organizations must also be considered. In this study, organizational variety and unequal product representation coexist, making the AI ecosystem structurally heterogeneous rather than uniformly distributed.

6.4 Pricing Metadata as a Major Analytical Constraint

The pricing analysis reveals the most substantial data-quality limitation in the study. Only 69 of the 3,027 products have an explicitly identifiable FREE or PAID classification, while 97.72% of records fall within the OTHER/UNKNOWN category. In addition, no products are classified as FREEMIUM, SUBSCRIPTION, or OPEN SOURCE.

The findings should not be interpreted as evidence that these pricing models are absent from the AI product ecosystem. Rather, the observed distribution indicates that the available data do not provide sufficiently complete or standardized information to classify the pricing structures of most products.

This distinction between observed metadata and actual market characteristics is particularly important. A dataset may accurately record the existence of thousands of AI products while failing to capture important commercial attributes. In such circumstances, the absence of a classification is not equivalent to evidence of the absence of the underlying phenomenon.

The pricing results therefore make a methodological contribution by demonstrating that missing metadata should itself be treated as an empirical characteristic of ecosystem datasets. The 97.72% OTHER/UNKNOWN classification is not simply a technical limitation to ignore or remove. It directly determines the types of questions the dataset can and cannot answer. In the present case, the data can describe the availability of pricing information but cannot support definitive conclusions regarding the actual prevalence of different AI pricing strategies.

6.5 Startup-Level Pricing Variation

The startup-level analysis provides limited descriptive evidence of variation in observed pricing classifications. OpenAI, for example, contains FREE, PAID, and OTHER/UNKNOWN product classifications. Google has the most linked products, although most remain unclassified by pricing. DeepSeek is represented by two products classified as FREE, while several other organizations exhibit combinations of FREE and OTHER/UNKNOWN classifications.

These observations indicate heterogeneity in the available pricing metadata. However, they do not establish statistically significant differences in corporate pricing strategies. The number of products associated with

most companies is extremely small, and the dominance of missing pricing information substantially reduces the effective sample available for comparison.

The absence of inferential statistical testing is therefore methodologically justified. Under such conditions, conventional comparisons across organizations would be affected by small cell counts and highly uneven category distributions. The findings should consequently be understood as descriptive patterns in the observed metadata rather than evidence of systematic differences in organizational pricing strategies.

This result reinforces a broader point about ecosystem analytics: the availability of a categorical variable does not necessarily mean it can support meaningful statistical comparison. Analytical decisions must consider not only the existence of data fields but also category completeness, sample size, distributional balance, and entity coverage.

6.6 Entity Linkage as a Critical Requirement for Ecosystem Mapping

The low rate of product company linkage is one of the study's most important methodological findings. Only 42 of the 3,027 products contain an identifiable startup association. This limitation affects every analysis that depends on organizational information, including product portfolios, concentration, diversity, and company-level pricing comparisons.

The startup-to-product ratio of 2.47 products per identified startup meaningfully describes the linked subset because both the number of products and the number of startups are based on established entity relationships. By contrast, dividing all 3,027 product records by the 17 identified startups would produce an inflated and substantively misleading estimate because most products cannot be attributed to any of these companies.

This result demonstrates that entity resolution is not merely a preprocessing activity but a substantive component of ecosystem research. The ability to identify relationships among products and organizations determines the extent to which structural characteristics of an ecosystem can be measured. Improving company-product linkage would therefore potentially have a greater effect on the quality of future ecosystem analysis than simply increasing the total number of product records.

6.7 Source Dependence and External Validity

The source-level analysis identifies a further limitation: all 3,027 product records originate from AI123Box. As a result, the dataset lacks sufficient source level variation to compare pricing classifications or metadata completeness across independent product directories.

This has implications for external validity. The observed pricing distribution may partly reflect the classification practices, extraction procedures, or metadata coverage of the source rather than the broader structure of the AI product market. Consequently, the extremely high proportion of OTHER/UNKNOWN observations may be source specific.

The findings therefore demonstrate the importance of multi source ecosystem datasets. Incorporating product information from multiple directories, official company websites, software marketplaces, and product documentation would enable future research to evaluate whether observed patterns remain consistent across sources. Such triangulation would also help distinguish genuine ecosystem characteristics from artifacts associated with individual data collection systems.

6.8 Implications for AI Ecosystem Research

The study has several implications for future empirical research on AI ecosystems.

First, ecosystem analysis should distinguish clearly between product-level coverage and relational coverage. Large numbers of product records provide a useful starting point, but ecosystem structure cannot be fully understood without reliable information connecting products to organizations and other ecosystem entities.

Second, concentration and diversity should be analyzed jointly. The present findings show that an ecosystem can contain numerous organizations while product representation remains concentrated among a relatively small number of participants. Combining HHI, Gini, Shannon, and Simpson measures provides a more complete structural characterization.

Third, report and interpret missing metadata explicitly. Incomplete pricing information, for example, affects the substantive conclusions that can be drawn from the data. Treating missingness as a methodological finding improves transparency and prevents unsupported interpretations.

Finally, future AI ecosystem studies would benefit from standardized taxonomies for product characteristics and improved entity-resolution procedures. These improvements would make it possible to conduct more reliable analyses of organizational portfolios, market structure, pricing strategies, technology categories, and relationships among ecosystem actors.

Overall, the findings suggest that current AI ecosystem datasets may be more effective for identifying broad structural patterns than for making precise claims about commercial strategies. The present dataset is therefore particularly useful for exploratory ecosystem characterization, while more complete metadata would be required for robust market-level inference.

7. Conclusion

This study empirically examined the structure of an observed AI product ecosystem through the analysis of 3,027 product records and the subset of products for which identifiable organizational relationships were available. The analysis focused on product concentration, organizational diversity, startup product relationships, and pricing model metadata.

The findings reveal a dual structural pattern. On the one hand, the linked ecosystem is organizationally diverse, with 17 identified startups represented among the available product-company relationships. The Shannon diversity index of 2.450 and Simpson diversity measure of 0.872 indicate that the observed ecosystem includes a relatively broad range of organizational participants. However, product representation is concentrated and unequal. Four companies account for 57.14% of linked products, the ten largest companies account for 83.33%, the HHI is approximately 1,281, and the Gini coefficient is approximately 0.434. Taken together, these findings characterize the observed ecosystem as organizationally diverse but product concentrated.

The study also identifies metadata incompleteness as a major limitation and substantive finding. Only 1.39% of product records contain an identifiable startup or company association, while 97.72% of products have OTHER/UNKNOWN pricing classifications. Consequently, the dataset provides substantially stronger evidence regarding the structure of the linked product ecosystem than regarding actual AI pricing strategies. The high proportion of missing or unclassified pricing information means observed pricing patterns should be interpreted as evidence of metadata limitations rather than definitive evidence of the commercial strategies of AI product providers.

The study's central contribution is therefore both methodological and empirical. The findings demonstrate that ecosystem mapping requires careful distinction between the structure of the observed dataset and the structure of the underlying market. Concentration indices, diversity measures, and descriptive distributions can provide meaningful insights, but their interpretation depends fundamentally on the completeness of entity relationships and metadata.

The study further demonstrates the value of combining complementary indicators of ecosystem structure. Concentration measures alone would emphasize the dominance of a relatively small number of organizations, whereas diversity measures alone would emphasize the presence of multiple participants. Their combined interpretation reveals a more nuanced ecosystem structure in which broad organizational participation coexists with unequal product representation.

Several limitations should be acknowledged. The organizational analysis is based on only 42 linked products, preventing the findings from being generalized directly to all 3,027 products or to the global AI market. The pricing analysis is constrained by the overwhelming dominance of the OTHER/UNKNOWN category. In addition, all product records originate from a single source, preventing cross source comparison and limiting external validity.

Future research should therefore prioritize three developments. First, company product entity resolution should be expanded to improve relational coverage across the product population. Second, pricing information should be systematically collected and normalized using a standardized taxonomy that distinguishes among free, freemium, paid, subscription, open source, and other pricing arrangements. Third, future ecosystem datasets should integrate multiple independent sources to enable cross-source validation and reduce dependence on the metadata practices of individual directories.

In conclusion, the study provides a descriptive empirical characterization of the available AI product ecosystem and demonstrates that AI ecosystem structure cannot be assessed independently of data quality. The observed ecosystem combines a relatively diverse set of organizations with substantial concentration in product representation, while incomplete company and pricing metadata impose important limits on inference. These findings highlight the need for more comprehensive entity linkage, standardized metadata, and multi-source ecosystem mapping to support more robust future analyses of the rapidly evolving AI product landscape.

References

- [1] Jha, P. (2026). Mapping the value orientation of artificial intelligence: The AI Orientation Observatory and the Conscious Innovation Index (CII) A global analysis of 66,725 AI applications across 203 countries. *Available at SSRN 6429558*.
- [2] Afshar, F. A., Ayoubi, S., Jafari, A. (2018). The extrapolation of soil great groups using multinomial logistic at regional scale in arid regions of Iran. *Geoderma*, 315, 36–48.
- [3] Amati, M., Tiede, J., Sun, Q. C., Deilami, K., Hurley, J., Fox, A., Dickson, J. (2023). Using machine learning to identify urban forest crown bounding boxes (CBB): Exploring a new method to develop urban forest policy. *Urban Forestry Urban Greening*, 85, 127943.
- [4] Zhang, J., Trautman, D., Liu, Y., Bi, C., Chen, W., Ou, L., & Goebel, R. (2024). Achieving the rewards of smart agriculture. *Agronomy*, 14(3), 452.
- [5] Pandey, P. K., Pandey, P. K., Mahajan, S., Paul, J., Iyer, S. (2024). Digital twin and virtual reality, augmented reality, and mixed reality. In *Digital Twins for Smart Cities and Villages*. Elsevier.
- [6] Pimenow, S., Pimenowa, O., Prus, P., Niklas, A. (2025). The impact of artificial intelligence on the sustainability of regional ecosystems: Current challenges and future prospects. *Sustainability*, 17(11).
- [7] Pimenow, S., Pimenowa, O., Prus, P. (2024). Challenges of artificial intelligence development in the context of energy consumption and impact on climate change. *Energies*, 17(23), 5965.
- [8] He, N., Yin, J., Slater, L. J., Liu, R., Kang, S., Liu, P., Liu, D., Xiong, L. (2024). Global terrestrial drought and its projected socioeconomic implications under different warming targets. *Science of the Total Environment*, 946, 174292.
- [9] Shekhar, H., Sharma, A. (2023). Global food production and distribution analysis using data mining and unsupervised learning. *Recent Advances in Food, Nutrition & Agriculture*, 14(1), 57–70.

- [10] Cavazza, A., Mas, F. D., Campra, M., Brescia, V. (2023). Artificial intelligence and new business models in agriculture: The “ZERO” case study. *Management Decision*, 61(9), 2500–2515.
- [11] Zhou, G., Zhou, L. (2021). Ecological meteorology: Origin, concept, and prospects. *Kexue Tongbao/Chinese Science Bulletin*, 66(2), 210–218.
- [12] Mpakairi, K. S., Dube, T., Dondofema, F., Dalu, T. (2022). Spatio-temporal variation of vegetation heterogeneity in groundwater dependent ecosystems within arid environments. *Ecological Informatics*, 69, 101667.
- [13] Liu, L., Chen, L., Liu, Y., Yang, D., Zhang, X., Lu, N., Ju, W., Jiang, F., Yin, Z., Liu, G., et al. (2022). Satellite remote sensing for global stocktaking: Methods, progress and perspectives. *National Remote Sensing Bulletin*, 26(2), 243–267.
- [14] Zhou, B., Zhu, Y., Yu, H., Xu, J. (2025). Current state, critical review and future prospects of research on tourism ecological security. *Shengtai Xuebao/Acta Ecologica Sinica*, 45(3), 1049–1058.
- [15] Magazzino, C., Gattone, T., Usman, M., Valente, D. (2024). Unleashing the power of innovation and sustainability: Transforming cereal production in the BRICS countries. *Ecological Indicators*, 167, 112618.
- [16] Brail, S. (2020). World cities of ride-hailing. *Urban Geography*, 1–22. <https://doi.org/10.1080/0272363-8.2020.1775030>.
- [17] Howson, K., Ferrari, F., Ustek-Spilda, F., Salem, N., Johnston, H., Katta, S., Heeks, R., Graham, M. (2021). Driving the digital value network: Economic geographies of global platform capitalism. *Global Networks*. <https://doi.org/10.1111/glob.12358>.
- [18] Kenney, M., Zysman, J. (2020). The platform economy: Restructuring the space of capitalist accumulation. *Cambridge Journal of Regions, Economy and Society*, 13(1), 55–76. <https://doi.org/10.1093/cjres/rsaa001>.
- [19] Loonam, J., O'Regan, N. (2022). Global value chains and digital platforms: Implications for strategy. *Strategic Change*, 31(2), 161–177. <https://doi.org/10.1002/jsc.2485>.
- [20] Stallkamp, M., Schotter, A. P. J. (2021). Platforms without borders The international strategies of digital platform firms. *Global Strategy Journal*, 11(1), 58–80. <https://doi.org/10.1002/gsj.1336>.
- [21] Howson, K., Ferrari, F., Ustek-Spilda, F., Salem, N., Johnston, H., Katta, S., Heeks, R., Graham, M. (2021). Driving the digital value network: Economic geographies of global platform capitalism. *Global Networks*. <https://doi.org/10.1111/glob.12358>.
- [22] Bearson, D., Kenney, M., Zysman, J. (2021). Measuring the impacts of labor in the platform economy: New work created, old work reorganized, and value creation reconfigured. *Industrial and Corporate Change*, 30(3), 536–563. <https://doi.org/10.1093/icc/dtaa046>.
- [23] Roundy, P. T., Asllani, A. (2024). Understanding AI innovation contexts: A review and content analysis of artificial intelligence and entrepreneurial ecosystems research. *Industrial Management & Data Systems*, 124(7), 2333–2363.
- [24] Tang, F., Zhang, L. (2025). Impact of data element development on the application of artificial intelligence in enterprises. *Empirical Economics*, 69(6), 3589–3634. <https://doi.org/10.1007/s00181-025-02823>.
- [25] Cavalcanti Pereira, C. D. (2026). The economic structure of the generative AI ecosystem: Capital intensity, capability diffusion, and competitive differentiation across the value chain. *Available at SSRN*.

- [26] Iyer, B., Lee, C. H., Venkatraman, N. (2006). Managing in a “small world ecosystem”: Lessons from the software sector. *California Management Review*, 48(3), 28–47. <https://doi.org/10.2307/41166348>.
- [27] Shipilov, A., Gawer, A. (2020). Integrating research on interorganizational networks and ecosystems. *Academy of Management Annals*, 14(1), 92–121. <https://doi.org/10.5465/annals.2018.0121>.
- [28] Ribeiro, T. F., Nogueira, R., Chimenti, P. (2024). Integrating business ecosystems and social networks: A case study on sustainable transitions. *Sustainability*, 16(24), 11045. <https://doi.org/10.3390/su162411045>.
- [29] Lin, J., Chinthavali, S., Stahl, C. D., Stahl, C., Lee, S., Shankar, M. (2016). Ecosystem discovery: Measuring clean energy innovation ecosystems through knowledge discovery and mapping techniques. *The Electricity Journal*, 29(8), 64–75. <https://doi.org/10.1016/j.tej.2016.09.012>.
- [30] Faber, A., Hernandez-Mendez, A., Rehm, S. V., Matthes, F. (2020). Collaborative modelling and visualization of business ecosystems. *Australasian Journal of Information Systems*, 24.
- [31] Ribeiro, L. C., Silva, V., Chiarini, T. (2025). Mapping the platform economy: A methodology for identifying and locating digital platform companies using NLP techniques. *Quality & Quantity*, 59(5), 3461–3485. <https://doi.org/10.1007/s11135-025-02106>.
- [32] Ribeiro, T. F., Nogueira, R., Chimenti, P., da Fonseca, A. (2026). Mapping business ecosystems using AI: A case study on energy transition. *Journal of Cleaner Production*, 569, 148546.
- [33] Mohammadi, N., Maghsoudi, M., Soghi, M. (2024). Innovation ecosystems in retail: Uncovering technological trends and collaboration networks through patent mining. *IEEE Access*, 12, 186753–186778.
- [34] Hook, D. W. (2026). Market dynamics, governance and open research metadata in the AI era. *arXiv preprint arXiv:2604.19507*.