nexus: The Missing Data Layer for AI Agents in Industrial Manufacturing

Why industrial product complexity defeats general-purpose AI – and how a harmonised product graph changes the equation.

February 2026

nexwise builds AI-native product data infrastructure for industrial manufacturers. Its core technology, which they rolled out in January 2026, nexus, transforms complex, fragmented product data into a unified foundation that enables AI agents across technical sales, marketing, and service.

Executive summary

Nine out of ten workflows in technical sales, marketing, and service at industrial manufacturers depend on product data. Yet this data remains fragmented across ERP, PIM, CMS, DMS, PDF catalogues, CAD systems, and configuration tools – in formats no AI agent can read.

The consequence is stark: **95% of enterprise generative AI pilots fail** to reach production with measurable business impact, according to the 2025 MIT NANDA study "The GenAI Divide." The core reason is not a lack of model capability. It is a lack of **data readiness**.

This whitepaper introduces nexus, an AI-native product data infrastructure that transforms scattered, heterogeneous product information into a unified, machine-readable product graph. Originating at Technical University Munich, nexwise provides the missing foundational layer that enables AI agents to operate across technical sales, marketing, and service from day one.

We examine the mathematical structure of industrial product complexity, explain why conventional integration approaches fail at scale, and present the architecture of nexus and the harmonised product graph (HPG) that resolves these challenges.

1. The industrial product data challenge

Industrial manufacturers operate in a data environment unlike any other sector. The sheer combinatorial scale of their product portfolios, combined with decades of accumulated data spread across disconnected systems, creates a structural barrier to AI adoption that no amount of model sophistication can overcome on its own. Understanding this challenge in precise terms is the first step toward solving it.

Industrial manufacturers routinely manage product portfolios spanning hundreds of thousands of variants. A single product family of configurable components – where customers select from parameters such as voltage rating, housing material, connection type, protection class, sensing range, and communication protocol – can generate a configuration space of staggering size.

Formally, consider a product described by k independent parameters, where the i-th parameter admits m_i discrete values. The theoretical number of combinations – before constraint logic eliminates invalid ones – is:

$$N = \prod_{i=1..k} m_i$$

For a product family with 12 configurable parameters averaging 8 options each, this yields $8^{12} \approx 6.9 \times 10^{10}$ theoretical combinations – nearly seven billion – from a single product line. Real-world constraint rules (mutual exclusions, dependency chains, regulatory filters) reduce this space, but the valid subset typically remains vast: millions to billions of orderable variants per product family are common in industrial automation and sensor technology.[1]

Across a manufacturer's full portfolio of dozens or hundreds of such product families, the aggregate configuration space grows further still. Leading industrial manufacturers report effective configuration spaces exceeding 10^{20} unique variants – a number that rivals current estimates for the total number of stars in the observable universe (on the order of 10^{24} to 10^{24}).[2]

This is not an academic curiosity. It defines the operational reality that any AI system must navigate to be useful in industrial settings.

1.2 Data fragmentation across systems

The data describing these complex portfolios does not reside in a single system. A typical manufacturer maintains product information across a constellation of specialised tools: ERP for pricing, availability, and order codes; PIM for structured attributes, classifications, and marketing descriptions; CMS and DMS for technical data sheets, application notes, certificates, and manuals;

[1] Tackling Combinatorial Explosion: A Study of Industrial Needs and Practices for Analyzing Highly Configurable Systems, Mukelabai et al. (2018); Entropy Based versus Combinatorial Product Configuration Complexity in Mass Customized Manufacturing, Modrak & Bednar (2016).
[2] The Evolution of Galaxy Number Density at z < 8 and Its Implications, Conselice et al. (2016).

CAD and PLM for dimensional drawings, 3D models, and technical specifications; configuration tools for constraint logic, dependency rules, and valid combinations; and Excel or SharePoint for supplementary data, internal knowledge, and pricing exceptions.

From an information-theoretic perspective, the cost of integrating k heterogeneous data sources grows quadratically without a unifying schema:

$$C(\text{integration}) = O(k^2) \text{ pairwise mappings}$$

Each new source added to the landscape requires mappings to every existing source, creating an ever-expanding web of point-to-point connectors. A manufacturer with 8 product data sources faces up to 28 unique pairwise integrations – each with its own transformation logic, maintenance burden, and failure mode.

1.3 The AI adoption wall

This combination of combinatorial product complexity and systemic data fragmentation creates what we call the AI adoption wall. General-purpose AI models – however capable at natural language understanding or code generation – cannot reason over product attributes scattered across incompatible schemas, configuration constraints buried in proprietary logic engines, technical specifications encoded in PDF tables and CAD metadata, or cross-references between product families that exist only in expert knowledge.

The consequence is predictable: industrial companies invest in AI pilots that work impressively in demos on curated data subsets but fail to scale to production, because they hit the data wall. **The model is not the bottleneck. The data layer is.**

The empirical evidence is stark. The 2025 MIT NANDA study – tracking over 300 publicly disclosed AI initiatives – found that while 80% of organisations are exploring AI tools, only 5% reach production with measurable P&L impact. Between 2024 and 2025, US companies invested an estimated $35–40 billion in generative AI, yet the vast majority saw no measurable return.[3]

The study attributes this systemic failure to what it calls the **"Learning Gap"**: most enterprise AI systems are deployed as stateless wrappers over existing databases. They do not retain feedback, fail to adapt to edge cases, and lack the contextual memory required to learn an organisation's product portfolio over time. Users confirm this: 60% report that tools do not learn from feedback, 55% say too much manual context is required for every interaction, and 45% cannot customise AI to their specific industrial workflows.[4]

When AI is forced to interface with rigid, siloed data structures, it inherits their limitations. It cannot infer that a specific chemical sealant is compatible with a particular pneumatic valve if the underlying

[3] The GenAI Divide: State of AI in Business 2025, MIT Media Lab – NANDA (2025)
[4] The GenAI Divide: State of AI in Business 2025, MIT Media Lab – NANDA (2025)

data architecture does not map that semantic relationship. The AI is only as intelligent as the data topology it is allowed to traverse.

2. Why existing approaches fall short

The industry has not been idle. Manufacturers and their IT partners have invested heavily in data integration, product classification, and – more recently – retrieval-augmented generation. Each of these approaches addresses part of the problem. None addresses the whole of it. The following sections examine why.

2.1 Traditional integration: Point-to-point mapping

The default enterprise approach to data unification – building ETL pipelines between individual systems – buckles under the weight of industrial product complexity. Each pipeline is custom-built for a specific source-target pair, encoding brittle, expensive-to-maintain, and impossible-to-reuse transformation rules across customers.

Worse, these pipelines operate on structured data and ignore the vast unstructured corpus of PDFs, technical documentation, and application notes that contains critical product knowledge. An AI agent tasked with answering a customer's configuration question may need to synthesise information from a data sheet (PDF), a product taxonomy (PIM), a constraint table (configurator), and an application note (CMS) – simultaneously. No ETL pipeline delivers this.

2.2 Static taxonomies and their limits

Product taxonomies – hierarchical classification trees – are foundational to how manufacturers organise their catalogues. Standards like ECLASS or ETIM provide industry-wide classification schemas. Yet taxonomies alone cannot serve as an AI-ready data layer.

First, they impose a single organisational axis on inherently multi-dimensional data. A product that belongs to the "sensors" family may equally need to be found by application (predictive maintenance), by industry (automotive), or by technical specification (IP67, −40 °C to +85 °C). Static hierarchies force the user – or the AI – to navigate one predefined path.

This is not just a data engineering problem; it is a buyer experience problem. Hick's Law tells us that the time required to make a decision increases logarithmically with the number of choices presented. When an industrial buyer arrives at a manufacturer's portal and is confronted with thousands of categories, nested filters, and esoteric technical jargon, cognitive load quickly exceeds acceptable thresholds.[5] The result: **decision paralysis, abandoned sessions, and lost revenue**.

[5] Beyond Taxonomies: User-Centric Navigation for Complex Industrial Product Portfolios in eCommerce, nexwise GmbH (2025)

In reality, an industrial buyer approaches a search with a specific application in mind – "a corrosion-resistant flow meter for a high-pressure saltwater pipeline with wireless telemetry" – which inherently requires searching across multiple axes simultaneously. No static taxonomy can serve this.

Second, taxonomies are static structures in dynamic environments. New products, reclassifications, and evolving application contexts require continuous manual maintenance that few organisations sustain at the needed cadence.

2.3 RAG on raw documents: Necessary but insufficient

Retrieval-augmented generation (RAG)[6] has emerged as the default pattern for grounding LLMs in enterprise knowledge. The approach – chunking documents, embedding them into vector space, and retrieving relevant passages at query time – works reasonably well for general Q&A over unstructured text.

However, RAG on raw documents fails in industrial product contexts because it lacks the structural understanding that product queries demand. When a customer asks "Which pressure transmitter works with glycerine-filled gauges at 400 bar and has a HART output?", the answer requires understanding that this is a **configuration query** rather than a text retrieval task, cross-referencing product specifications across structured and unstructured sources, validating compatibility constraints that may not appear in any single document, and returning a precise product identifier – not a paragraph of text.

Vector similarity search over document chunks does not provide this. The industrial domain requires a data layer that combines the richness of unstructured knowledge with the precision of structured product semantics.[7]

3. nexus: Architecture of the harmonised product graph

The core architectural insight is the transition from taxonomies to ontologies.[8] A taxonomy describes a strict hierarchy: Hardware > Fasteners > Bolts > Hex Bolts. An ontology goes further by establishing multi-dimensional semantic relationships: Product A "is compatible with" Product B, Component C "requires" Tool D for installation, Material E "is compliant with" Standard F. This interconnected web of relationships provides the contextual foundation that AI reasoning requires.

The architecture follows a clear pipeline: data sources are ingested and harmonised into a unified product graph, which is then made accessible through multiple retrieval mechanisms and agent interfaces.

---

[6] Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, Lewis et al. (2020)
[7] Why AI Data Quality Is Key To AI Success, IBM (2025)
[8] Ontologies for ecommerce personalization – going beyond taxonomies, Earley Information Science (2025)

3.1 Ingestion and Harmonisation

The foundation of nexus is an automated pipeline that connects to the systems where product data lives – websites, SharePoint, PIM exports, cloud storage, CMS, ERP – via standardised connectors. Rather than requiring a manufacturer to restructure their data estate, nexus meets the data where it is.

Structured data processing. Attributes from PIM, ERP, and configuration systems are parsed, cleaned, transformed, structured, and enriched. This is where the depth of industrial product data becomes apparent – and where most generic approaches fail.

Consider a seemingly simple attribute: a sensor's accuracy specification, recorded as "0,1 % FS." A general-purpose system treats this as a text string. nexus resolves it structurally: the value is 0.1, the unit is percent, and "FS" (full scale) is a qualifier that defines what the percentage refers to. The **property resolution engine** determines whether such qualifiers are intrinsic to the property (and should be attached as metadata) or variable across values (and should be captured as a separate dimension). Similarly, a supply specification such as "DC 8...35 V" is decomposed into its constituent parts: supply type (DC), minimum range (8), maximum range (35), and unit (V).

This kind of resolution – multiplied across thousands of properties per product family – is what transforms raw data into machine-readable knowledge. Without it, an AI agent cannot reliably compare products, validate configurations, or answer technical queries with precision.

Unstructured document processing. PDFs, data sheets, and technical documentation are parsed into hierarchical document graphs. Specialised LLM workflows extract structured attributes from tables, images, and natural-language specifications. Vector embeddings capture semantic content for retrieval, while a translation memory preserves the provenance of every transformation.

Traceability. The pipeline maintains a Context Graph – a record of every raw attribute, transformation step, and enrichment decision. This creates full traceability from any graph node back to its source, a requirement in regulated industrial environments.[9]

3.2 The unified product graph (UPG)

The output of the ingestion pipeline is the unified product graph: a machine-readable, queryable representation of a manufacturer's entire product knowledge. It rests on three pillars:

Standardised taxonomy. Products are classified within a structured, ontology-driven hierarchy that goes beyond flat catalogue trees. This provides the skeletal structure that prevents hallucinations – the AI can only navigate and reference products that actually exist within well-defined categories. The taxonomy is guided by but not limited to industry standards like ECLASS and ETIM, enabling nexwise to represent manufacturer-specific product logic alongside standardised classifications.

[9] From dirty data to AI-ready: Building a unified manufactsuch asuring data ecosystem, RSM US (2025)

Harmonised attributes. Product properties are normalised against the ontology: types are standardised (following ECLASS conventions where applicable), units are resolved, qualifiers are attached, and values are validated. This is the layer that makes attributes genuinely comparable across product families and data sources – the prerequisite for any AI agent to reason over product specifications rather than merely retrieve text.

Learning document graph. Technical documentation – data sheets, application notes, manuals, certificates – is linked to the product graph at the variant level, not merely stored alongside it. Documents are structured into retrievable segments with semantic embeddings, enabling precise and fast references. When an AI agent answers a question, it can cite the specific section of the specific document for the specific product variant – not a generic PDF.

With the ontology as a mediating schema, the integration cost transforms from quadratic to linear:

$$C(\text{integration}) = O(k) \text{ with ontology vs. } O(k^2) \text{ without}$$

Each new data source requires only a single mapping to the ontology, not pairwise mappings to every other source. This architectural property is what makes nexus economically viable for manufacturers with complex, multi-system data landscapes.

3.3 Retrieval and access

nexus provides multiple retrieval mechanisms to serve different query types:

Metadata filtering for structured constraints – "all sensors with IP67 rating and HART output."

Graph-based retrieval for relational queries – "which accessories are compatible with this transmitter in a 4–20 mA loop?"

Semantic search for intent-driven, natural-language queries – "I need something for outdoor level measurement in corrosive environments."

Keyword search for direct lookups – article numbers, product names, specific technical terms.

Queries are processed through intent recognition and routing to the appropriate mechanism (or combination of mechanisms). This multi-modal retrieval goes far beyond generic RAG. The LLM serves as a reasoning and orchestration engine – not as a factual knowledge store – eliminating hallucination risk and ensuring enterprise traceability.

Every query result can return not just answers, but the underlying resources – images, PDFs, CAD files, certification documents – linked through the product graph to the specific variant in question.

3.4 Platform and agent integration

nexus is designed to serve as the product data foundation not only for nexwise's own AI agents, but for a manufacturer's broader technology ecosystem.

nexwise's agent platform sits on top of nexus as an orchestration layer, coordinating data retrieval, reasoning logic, and user-facing interfaces across the use cases described in Section 5. This separation – data infrastructure below, application logic above – is architecturally deliberate: it ensures that the product graph remains a stable, reusable asset regardless of how agent frameworks evolve.

Looking ahead, nexwise is developing a documented nexus API (**REST and MCP**) to open the unified product graph to third-party agent platforms and AI tools – including Microsoft Copilot, Salesforce Agentforce, SAP Joule, workflow automation tools like n8n, and cloud AI platforms. The goal: a manufacturer's investment in product data harmonisation should not be locked into a single application. The same unified graph should be able to power the website's guided selling experience, the sales team's internal assistant, a service chatbot, and a distributor's procurement integration – simultaneously.

> **Interested in shaping the nexus API?** We are looking for design partners to co-develop the API in 2026 – manufacturers who want to build custom AI applications on top of their harmonised product data. If this describes your ambitions, reach out at hello@nexwise.ai.

4. How nexus gets smarter over time

A data layer that is accurate at launch but degrades over time solves nothing. For industrial manufacturers – where product lines evolve, documentation changes quarterly, and customer requirements shift with markets – sustainability is not a feature. It is a prerequisite. nexus is designed to improve continuously through three mechanisms that compound with use.

Accelerating onboarding through learned data patterns. Every attribute transformation, enrichment, and mapping decision is recorded in the Context Graph. When nexus processes a new manufacturer's data, it draws on patterns established across previous deployments. A property resolution solved for one customer's product catalogue – how to decompose "0,1 % FS" or parse "DC 8...35 V" – accelerates setup for the next. For manufacturers, this translates to faster time-to-value: onboarding that begins at weeks, not months, and shrinks further as the system matures.

Improving answer quality through validated Q&A. Technical questions and their validated answers – on features, applications, compatibility, and configurations – are captured and refined over time. Each deployment enriches a growing knowledge base, sharpening retrieval precision. For sales and service teams, this means the system's answers become more accurate and comprehensive with every interaction, not less.

Bridging natural language and product configurations. Conversations, extracted parameters, and resulting configurations are captured as structured pairs. This enables increasingly accurate translation from how customers actually describe their requirements – in natural, often imprecise language – to valid product configurations. For buyers, the experience becomes more intuitive over time; for manufacturers, conversion rates improve as the system learns to interpret the language of their specific market.

Together, these mechanisms ensure that the value of nexus grows with adoption – not just within a single organisation, but across the network of manufacturers it serves.

5. From data layer to AI agents

Infrastructure only matters if it enables applications that deliver measurable business outcomes. nexus is not an end in itself – it is the foundation on which specialised AI agents operate. With the agent platform as the orchestration layer, nexwise deploys these agents across workflows where industrial manufacturers today accept significant inefficiency because their data architecture cannot support anything better.

Product data touches every interaction a manufacturer has – not just internally, but across the full network of people who need to find, understand, configure, and commission products.

Customers searching for the right product on a website, requesting quotes, or specifying configurations for a project. **Employees and internal teams** – from sales engineers and product managers to marketing and application support – who need instant access to accurate, complete product knowledge. **Sales and service partners** – distributors, system integrators, OEM customers – who sell or install the manufacturer's products but lack direct access to the depth of internal expertise. **End users and commissioners** who receive a product and need to install, configure, and operate it correctly for their specific application.

nexus provides the shared data foundation; the agent platform enables purpose-built applications for each group.

5.1 Guided selling – customers & partners

The guided selling agent navigates complex portfolios using conversational AI, enabling product selection for configurable, variant-rich products where traditional filter-and-browse interfaces fail.

Where manufacturers today resort to "Contact us" for complex product inquiries, guided selling provides instant, technically correct guidance – 24/7, at scale.

The economic impact can be expressed directly. The total cost of a B2B transaction is a function of the buyer's search cost, the probability of search failure, and the cost of human assistance required to salvage the interaction:

$$C(transaction) = C(search) + P(fail) \cdot C(assist)$$

In environments governed by static taxonomies, search costs are high (cognitive overload from navigating combinatorial category trees), failure probability is high (buyers cannot find what they need), and the cost of human assistance is extreme (industrial sales engineers are highly compensated specialists). This equation explains why nearly 80% of B2B buyers now prefer a rep-free digital experience[10] for standard purchases, yet digital channels consistently fail to deliver for complex industrial products.

nexus fundamentally alters this equation. By enabling buyers to use natural, application-centric language to locate products, search cost drops toward zero. Because the system dynamically maps intent to the correct product with semantic precision, failure probability plummets. The expensive human resource is reserved for genuinely novel, high-stakes engineering consultations where it is actually needed – enabling organisations to scale revenue without proportionally scaling headcount.

5.2 Product expert – internal teams & partners

The product expert agent provides internal sales teams with real-time access to the full depth of the product knowledge graph. Technical questions that previously required escalation to senior product specialists – on compatibility, configuration options, application suitability, or cross-references between product families – are answered instantly, with context, precision, and source traceability.

The impact goes beyond speed. Sales teams in industrial organisations frequently operate with uneven product knowledge: experienced engineers carry decades of tacit expertise, while newer team members depend on them for complex queries. The product expert agent levels this asymmetry, ensuring consistent knowledge quality across the entire sales organisation. Experienced specialists are freed for complex project business and customer relationships rather than routine internal support.[11]

5.3 Commissioning – end users & field service

Post-sales, the commissioning agent generates individualised setup and installation guidance from the customer's specific product configuration. Rather than generic manuals, service teams receive

[10] Manufacturing Sales: How Leaders Can Drive Growth in a Complex Selling Environment, Tyson Group (2025)
[11] 8 Key Challenges Modern Sales Execs in Manufacturing are Facing, IndustrySelect (2025)

step-by-step instructions tailored to the exact ordered variant – reducing commissioning errors, minimising return visits, and accelerating time-to-operation.

For manufacturers with large field service organisations or channel partner networks, this addresses a persistent challenge: ensuring that the person commissioning the product has access to the same depth of knowledge as the engineer who specified it.

5.4 Beyond these use cases

Guided selling, product expert, and commissioning are nexwise's initial agent deployments – selected because they address the highest-frequency, highest-impact workflows where product data quality directly determines business outcomes. But the unified product graph enables a much broader range of applications.

Wherever a workflow depends on accurate, contextual product data, nexus can serve as the foundation. Automated tender and RFQ response: sales teams match technical requirements against the portfolio and generate accurate pricing in minutes rather than days. Quoting and configuration validation: the system checks complex orders against constraint logic before they reach the ERP, catching errors that would otherwise surface downstream. Technical content generation: marketing teams derive product descriptions, comparison tables, and application stories directly from structured data rather than writing them from scratch. Training and onboarding: new employees and partner sales teams ramp up on a complex portfolio quickly, guided by the same knowledge base that powers customer-facing agents.

The agent platform's role is to make each of these workflows deployable without requiring a rebuild of the data layer. Because nexus already provides the harmonised, machine-readable product graph, new agents can be developed and deployed on a foundation that is already production-ready – rather than starting each use case from scratch.

6. Strategic implications: Build vs. buy

For technical leaders evaluating AI strategies, the question is no longer whether to invest in AI but how to invest without repeating the patterns that have led 95% of enterprise AI pilots to stall. The evidence points to a structural distinction between organisations that build from scratch and those that leverage purpose-built, verticalised infrastructure.

Companies that attempted to build specialised AI solutions internally succeeded only **33%** of the time.[12] Organisations that procured specialised, verticalised AI solutions from external vendors succeeded **67%** of the time. The discrepancy reflects the technical difficulty of building stateful, context-aware AI architectures from scratch – particularly the data harmonisation, ontology design, and graph-based retrieval infrastructure that industrial product complexity demands.

[12] The GenAI Divide: State of AI in Business 2025, MIT Media Lab – NANDA (2025)

The underlying dynamic is straightforward. Industrial manufacturers excel at engineering physical products; building and maintaining production-grade AI data infrastructure is a fundamentally different discipline. Attempting to do both in parallel stretches internal resources and typically leads to what the NANDA study calls **"pilot purgatory"** – indefinite evaluation cycles that consume budget without reaching production.

This is compounded by the **"Shadow AI"** phenomenon: while only 40% of companies have official enterprise AI tools, over 90% of employees routinely use personal, consumer-grade AI to complete their work.[13] The gap signals unmet demand – employees recognise the value of AI, but the enterprise tools available to them, constrained by siloed and rigid data architectures, fail to deliver for specialised industrial workflows.

For CDOs and CIOs, the strategic calculus favours specialised external infrastructure that accelerates time-to-value, reduces integration risk, and provides a secure, enterprise-grade alternative to unmanaged consumer tools. For Sales Leaders, it means digital channels that can finally handle the technical complexity that buyers encounter – without defaulting to "Contact us."

7. Conclusion: The data layer as strategic infrastructure

The preceding sections have traced a single thread: from the mathematical complexity of industrial product portfolios, through the failure modes of existing approaches, to an architecture designed to resolve them. What remains is the strategic consequence for manufacturers deciding where to invest next. The industrial sector stands at an inflection point. AI capabilities have surpassed what most manufacturing workflows require. The bottleneck is not model intelligence – it is data readiness.

nexus addresses this bottleneck at its root. By transforming fragmented, heterogeneous product data into a harmonised, semantically rich graph, it provides the foundational layer that AI agents need to function reliably in industrial environments. The architecture reduces integration complexity from quadratic to linear, the ontology-driven approach ensures semantic precision across data sources, and the system improves continuously as it processes more data across deployments. Because the unified product graph is built to be accessible beyond nexwise's own agents – with an open API on the roadmap for 2026 – the investment extends beyond any single application. It becomes the product data infrastructure on which an entire ecosystem of AI-powered workflows can operate.

For manufacturers evaluating AI strategies, the implication is clear: **investing in AI models without investing in the underlying data layer is building on sand**. nexus is the missing infrastructure that enables industrial adoption of AI.

[13] The GenAI Divide: State of AI in Business 2025, MIT Media Lab – NANDA (2025)

About nexwise

nexwise GmbH is a Technical University of Munich spin-off building AI-native product data infrastructure for industrial manufacturers. Its customers include established industry leaders such as Maschinenfabrik Reinhausen – the global market leader in on-load tap-changers for power transformers – as well as manufacturers in automation technology and industrial sensors. The company's core technology, nexus, creates a unified product data layer that enables AI agents for technical sales, marketing, and customer service. nexwise is backed by 42CAP and angel investors with deep domain expertise in industrial technology and AI.

nexwise.ai | hello@nexwise.ai