EXPERT SOFT
software development

# AI-Ready Ecommerce Data: A Practical Architecture for Enterprise Platforms

How to prepare enterprise data for reliable AI adoption

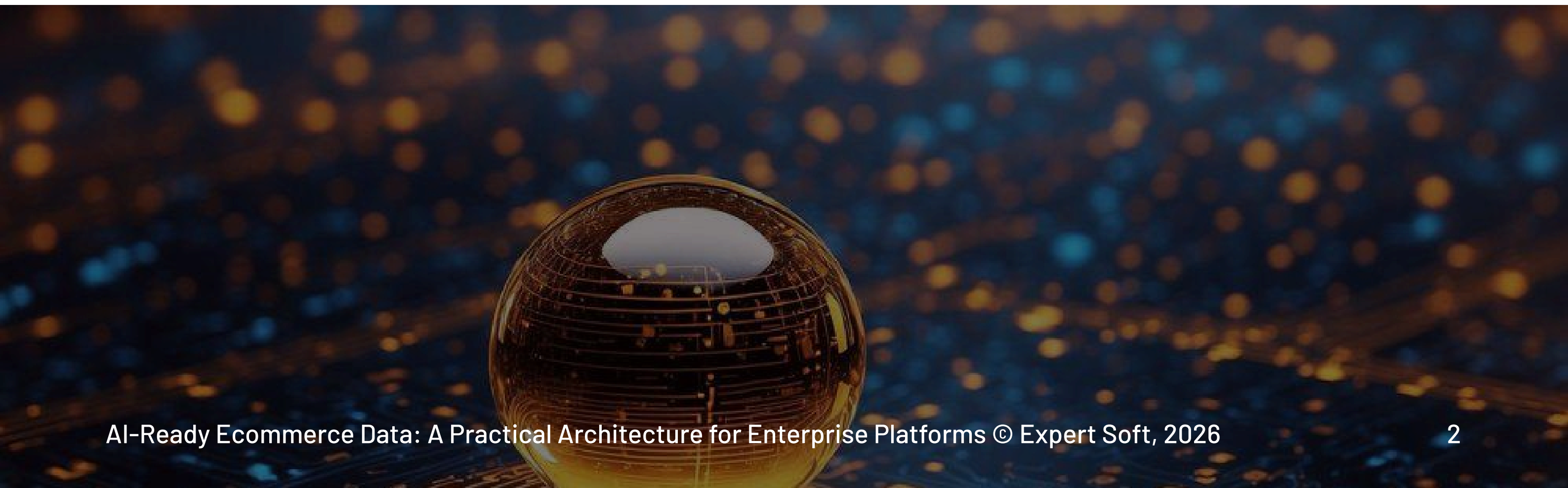
Enterprise ecommerce data rarely lives in one clean system. Product information arrives from PIM platforms, supplier feeds, legacy tools, and internal services, often in different formats and structures. When companies introduce AI capabilities such as search, recommendations, or assistants, they quickly discover that this data was never designed for AI to interpret reliably.

The instinctive response is to standardize everything at the source. In reality, rewriting upstream systems is slow, expensive, and rarely practical for large commerce environments.

This whitepaper explores a more pragmatic path: preparing data for AI without rebuilding the systems that already run the business. It looks at how enterprises can introduce a structured preparation layer that makes operational data usable for AI while keeping source platforms unchanged, using an architectural approach known as Medallion architecture.

# Table of Contents





# Why AI May Feel Risky in Enterprise Ecommerce

When AI enters an ecommerce organization, it is expected to reduce effort, speed up decisions, and make complex operations easier to navigate. In practice, it often does the opposite at first: it increases the perceived risk of acting on data. The pattern is predictable.

In commerce, decisions are only as safe as the ability to verify them. If an answer cannot be traced, repeated, and explained within the pace of daily operations, it does not feel like automation, it feels like uncertainty.

That is why teams may still treat AI as something to experiment with on the side rather than something to rely on in core workflows. Let's start by looking at the recurring symptoms that make AI adoption feel fragile in ecommerce environments, and why they tend to surface once AI touches day-to-day data and operations.

## DISTRIBUTED AND HETEROGENEOUS DATA REALITY

Commerce data is inherently distributed. It is created and maintained across multiple systems, updated at different rhythms, and owned by different teams. That is not a sign of weak discipline, it's how enterprise commerce operates.

**Field symptom**

AI outputs feel uneven in day-to-day use. The same question can produce different answers depending on which system version of the truth is reached first, and teams quickly notice data gaps, inconsistencies, and even conflicting values across systems that make results hard to rely on.

**Core problem**

AI is consuming data before meaning is stabilized. Without a governed readiness layer that aligns semantics and establishes a consistent view, heterogeneity turns into variability, and variability turns into fragility.

## DATA IS NOT AI-READY

Most commerce organizations are not lacking data. The challenge is that this data was never shaped for consistent machine consumption. Expert teams and target systems often already know how to interpret it correctly within their own context, but AI does not share that context by default. Once AI is asked to read this data as one coherent picture, gaps in semantics and interpretation start to surface.

**Field symptom**

AI struggles with basic consistency. It fills gaps with assumptions, mixes fields that look similar but behave differently.

It also produces  answers that sound reasonable until someone tries to validate them in a concrete workflow. Teams end up double-checking everything, and the perceived value of AI drops fast.

**Core problem**

The issue is representation. Commerce data from various sources often lacks a stable semantic layer — one that defines schemas, field-level meaning, and the rules by which attributes should be interpreted across systems and over time. Without an AI-readable representation that is normalized, unambiguous, and consistent, AI will generalize across ambiguity, turning outputs into risk.

## AI OUTPUT ISN'T DECISION-GRADE

In ecommerce, trust is not built by how fluent an AI sounds. It is built by whether teams can treat an answer as decision-grade under real operating pressure.

**Field symptom**

Different assistants or AI features produce different answers to what feels like the same question. When someone asks where a number came from, the team cannot point to a clear source path quickly, so the output is treated as a suggestion rather than something that can drive actions.

**Core problem**

Trust breaks when answers are not reproducible and traceable. An AI response has to be grounded by design in a consistent, prepared data view and linked back to the inputs that produced it. Without that built-in groundedness, the business has no way to accept the answer as a reliable reference, no matter how helpful it appears.

## NOBODY OWNS THE ANSWER END TO END

AI in commerce rarely lands inside a single team's boundary. It usually sits between domains, data sources, and operational workflows, where ownership has always been shared and therefore unclear.

### Field symptom

When an AI-driven output causes confusion or breaks downstream expectations, the issue turns into a handoff chain. Each team can explain its own system or dataset, but no one can confidently own the final answer, keep it consistent over time, and decide what changes are allowed without breaking other consumers.

### Core problem

AI needs an owned layer of responsibility, not just connected systems. Without clear ownership for the prepared data view and its contracts, the "answer" becomes a byproduct of many moving parts, and stability becomes nobody's job.

## SECURITY AND COMPLIANCE BECOME A BOTTLENECK

In enterprise commerce, AI usually needs access to operational data before it can be useful. That is exactly where governance concerns surface and slow initiatives down.

### Field symptom

Teams struggle to get AI into real workflows. Legal, security, and compliance ask predictable questions: what data is exposed, whether raw fields can leak, how usage is logged, and how access is enforced. When those answers are unclear, approvals drag, scope shrinks, or AI gets confined to low-impact use cases.

### Core problem

The blocker is the lack of a governed readiness boundary. If AI can touch raw or unstable data, and if there is no clear audit trail of what was used and under which rules, the risk profile becomes unacceptable for regulated and enterprise environments.

Taken together, these problems point to the same underlying gap in the commerce stack. AI is being asked to operate on data that is distributed by nature, shaped for human and system-to-system workflows, and governed inconsistently across domains.

The obvious response is to standardize everything at the source before letting AI touch it. In commerce, that is rarely realistic. A more workable path is to keep source systems unchanged and move the effort into how data is prepared and represented for AI use. That is where architecture becomes the lever.

There's more than one way to build a scalable AI core. Explore their trade-offs, strengths, and where each makes the most sense to choose the one that suits you best.

# Medallion Architecture as a Foundation for AI-Ready Data

Instead of trying to reshape every system that produces commerce data, you can change where responsibility sits. The goal is to keep source platforms stable and focus on how their outputs become usable for AI.

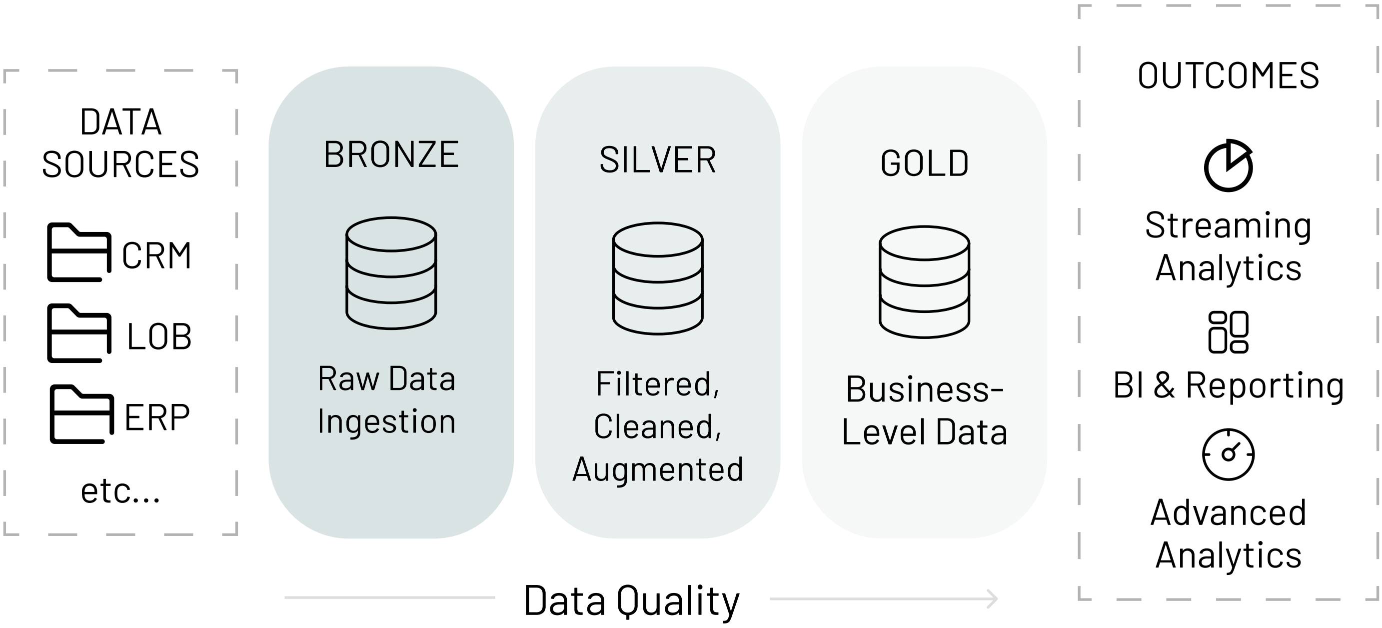
The Medallion architectural approach gives you a clear way to do that: it preserves raw truth, progressively stabilizes meaning, and produces prepared data views that AI and teams can rely on with confidence.

## WHAT MEDALLION ARCHITECTURE IS

Medallion architecture provides a way to make commerce data progressively more usable without changing the systems that produce it. It does this by separating concerns across layers, so raw inputs stay intact, transformation work is explicit, and downstream consumers can rely on prepared, stable representations. This architecture includes several layers.

### Bronze — the ingestion layer

It collects raw ecommerce data from multiple sources, e.g., databases, APIs, event streams, flat files, partner and marketplace feeds, and stores it as a faithful copy of what arrived. Some teams apply light parsing to make the data usable for processing, but the intent stays the same: keep an immutable, high-volume, traceable record of inputs so you can always return to the original payload and preserve repeatability when definitions drift or issues surface.




**Silver — the processing layer**

This is where data is cleaned, structured, normalized, and, where needed, enriched. Fields are aligned to a schema that makes meaning explicit and consistent across sources, and basic quality checks are applied so downstream systems do not have to interpret ambiguity.

In commerce terms, this is where attributes become machine-readable: formats are standardized, naming is consistent, and the same concept stops being represented in five different ways.

**Gold — the consumption layer**

It exposes refined, business-aligned views that teams can use directly for decision-making, automation, and AI-driven use cases. This is where prepared data is shaped into domain-specific products for different consumers, whether that means reporting, analytics, ML models, search, or AI assistants.

In practice, organizations often build multiple Gold layers, each designed around a particular business context or use case.  This is what makes the Gold layer especially valuable: it turns prepared data into stable, reusable representations that can be consumed across functions without forcing every team or system to reinterpret raw operational signals.

This layered model creates a clear boundary between raw inputs and prepared data views. AI systems, analytics, and ML models do not operate on raw ingestion layers but on stabilized representations produced in the upper layers of the architecture. As a result, data preparation becomes reproducible and easier to verify, while the ambiguity that AI would otherwise need to interpret is reduced.

This turns the architecture into a governance mechanism for data consumption rather than just a data processing pipeline. It defines how data becomes trustworthy and usable across analytics, automation, and AI systems.

## RETHINKING MEDALLION FOR AI READINESS

For AI systems, readiness is defined less by cleanliness and more by representation. AI relies on stable semantics, consistent schemas, and clear structural contracts to interpret data correctly. When those elements are missing, models compensate by generalizing patterns in the data, which often means that small ambiguities in fields or formats turn into inconsistent or unreliable outputs.

In ecommerce, AI readiness becomes visible through a small set of architectural principles:

- **Consistent semantics** across systems, so the same attribute or signal carries the same meaning everywhere.
- **Stable schemas and data contracts** that prevent silent structural drift across feeds and pipelines.
- **Validation rules and quality checks** that surface inconsistencies before they propagate into AI outputs.
- **Traceable and repeatable transformations,** making it possible to follow a prepared value back to its source, update Silver or Gold logic when needed, and re-run pipelines without data loss. This is especially important for pilots and for consistent improvements in data quality over time.
- **Predictable change behavior,** where updates to structures or definitions are controlled rather than implicit.

This is why the Silver and Gold layers become the working zone for AI readiness:

**SILVER LAYER**
The point where meaning becomes stable. Here, teams align semantics across sources and introduce minimal structural contracts so the same concept is represented consistently across the ecosystem.

**GOLD LAYER**
The point where prepared data becomes business-ready and directly usable for decisions. Domain-aligned views are created so the data reflects how commerce operations reason about catalog, availability, pricing, and other operational signals.

When this readiness layer exists, many of the fragility signals described earlier begin to fade. Answers become more stable, verification becomes faster, and teams spend less time reconciling different interpretations of the same data. Governance also becomes easier because AI operates on prepared views rather than raw operational feeds.

Importantly, this architecture doesn't require a large transformation program. Readiness can be built incrementally around specific domains, such as catalog attributes, availability, or pricing. Each stabilized domain makes the AI use cases built on top of it more predictable, allowing organizations to improve reliability without redesigning their entire data landscape.

## WHY THIS MATTERS FOR EXPLAINABILITY AND TRUST

In ecommerce operations, AI outputs quickly turn into decisions. Pricing adjustments, assortment visibility, replenishment signals, or operational alerts all rely on data that teams must be able to verify under real working conditions. When an answer cannot be traced back to a clear data path, it stops being useful, even if it sounds convincing.

This is why explainability in enterprise environments is primarily a data architecture question. If the data that feeds AI is not organized in a way that preserves meaning, transformation history, and ownership, no model layer can reliably reconstruct how a result was produced.

A layered preparation approach changes this dynamic. Because data moves through explicit stages of preparation before it becomes available for consumption, the path from raw signal to business-facing value remains visible and controlled. AI outputs can therefore be interpreted against known data views rather than treated as opaque model behavior.

In practice, this shifts AI from something that produces interesting suggestions to something that teams can safely rely on in day-to-day ecommerce operations.

AI becomes sustainable when architecture, platforms, and data responsibility are aligned. Discover what a production-grade enterprise AI architecture includes and how to build it step by step.

# AI Search Readiness as a Practical Application of the Architecture

To make the architecture more concrete, it helps to look at a real implementation. One example is the AI Search Readiness Kit, a solution designed to operationalize the layered readiness approach for commerce catalog data.

Rather than presenting it as a product pitch, we will use it as a practical illustration of how the Medallion model can be implemented in a commerce environment, and how the same architectural logic can translate into predictable AI behavior.

## WHAT AI SEARCH READINESS REQUIRES

AI search on the website is a common domain where organizations try to apply AI. But many teams quickly notice that the behavior of AI systems is often less stable than expected.

Instead of improving relevance, AI search can introduce new types of failure, such as Irrelevant or low-quality search results, even for common queries, inaccurate recommendations or category placement, and unstable results for similar queries.

These behaviors reveal that data readiness for AI search is not simply about cleaning data. It requires preparing the data so that product structure, attributes, and relationships are represented in a way that AI systems can interpret reliably.

AI Search readiness typically requires:

- **Clear product and variant structures,** so AI can distinguish base products, variants, and related items instead of treating them as separate records.
- **Normalized attribute vocabularies,** aligning supplier-specific names and formats into a shared catalog language.
- **Harmonized supplier feeds,** ensuring similar products follow consistent structures and classification across partners.
- **Explicit category context,** so attributes are interpreted within the correct domain, not just by text similarity.
- **Prepared catalog views for discovery use cases,** where product information is organized specifically for search, AI assistants, and recommendation engines rather than raw operational feeds.

These requirements make AI search readiness a practical example of the broader architectural principle described earlier. AI systems shouldn't interact with raw supplier feeds directly. They should rely on prepared catalog representations where product structures, attributes, and relationships are already stabilized.

The challenge is implementing a repeatable way to create and maintain these prepared catalog views.

## HOW THE AI SEARCH READINESS KIT IMPLEMENTS IT

The AI Search Readiness Kit is designed to operationalize catalog data readiness without requiring changes to the systems that already manage product data. It operates as a read-only preparation layer, preparing normalized data representations that AI systems can consume while existing PIM platforms and supplier integrations remain unchanged.

At its core, the solution focuses on structuring catalog data so that product meaning becomes explicit and consistent across supplier feeds, product variants, and category contexts.

Importantly, the AI Search Readiness Kit doesn't attempt to "clean the data once." Data readiness is treated as an ongoing process where supplier feeds, product structures, and attribute vocabularies are continuously aligned as new data enters the ecosystem. Because of this, the solution acts less like a one-time data processing pipeline and more like a controlled preparation layer for product information.

The AI Search Readiness Kit supports several key capabilities relevant to AI search readiness:

- **Supplier feed harmonization,** where product data arriving from different partners is aligned to a shared catalog structure and vocabulary rather than remaining in supplier-specific formats.
- **Attribute normalization and enrichment,** ensuring that product attributes follow consistent naming, formatting, and semantic rules.
- **Domain-specific vocabulary alignment,** where abbreviations, units, business terms, and supplier-specific naming are translated into a shared catalog and search language that AI systems can interpret consistently.
- **Variant and product structure alignment,** making relationships between base products, variants, bundles, and compatible items explicit rather than implicit in raw feeds.
- **Category-aware representation,** here attributes and product descriptions are interpreted within their domain context instead of relying only on keyword similarity.

## WHERE THE SOLUTION FITS IN THE MEDALLION ARCHITECTURE

Within a Medallion-style architecture, the AI Search Readiness Kit operates in the layers where meaning and structure are stabilized before data becomes available for consumption. The solution focuses on preparing catalog data in the stages where AI readiness is established:

**Silver layer — data normalization and alignment**
This is where the AI Search Readiness Kit performs most of its work, transforming data received on the Bronze: supplier feeds are harmonized, attribute vocabularies are normalized, and product structures such as variants, bundles, and compatibility relationships are aligned into a consistent representation. At this stage, catalog data meaning becomes stable enough for downstream systems to interpret reliably.

**Gold layer — prepared data views for discovery**
The prepared catalog data is then exposed as domain-aligned views optimized for discovery use cases. Search engines, recommendation systems, and AI assistants consume these views instead of raw feeds, which ensures that product information is interpreted consistently across different AI-driven features.

Seen this way, the AI Search Readiness Kit functions as a practical preparation layer between raw data ingestion and AI consumption. It preserves the integrity of source systems while ensuring that downstream AI capabilities operate on structured, explainable catalog data representations rather than fragmented supplier data.

See how AI search readiness works inside a Medallion architecture. Explore the AI Search Readiness Kit demo.

# Long-Term Capabilities of an AI-Ready Data Architecture

The AI search example illustrates a broader point. AI readiness is not just about preparing data for one use case, such as search or recommendations. It requires an architectural foundation where AI works on prepared, governed data instead of fragmented operational records.

Within a Medallion-style architecture, this preparation layer turns operational inputs into stable representations AI can reliably use. Instead of interpreting inconsistent product feeds and supplier-specific structures directly, AI works with aligned, structured data.

Once this layered preparation is introduced, the impact extends beyond the original use case. Rather than rebuilding data pipelines for every new initiative, organizations begin to develop shared data capabilities that support multiple AI-driven applications across the commerce stack. Several long-term advantages emerge:

### Reusable AI foundation

Once data structures and attributes are stabilized in prepared layers, the same representations can support search, recommendations, conversational assistants, and analytics without repeated data preparation

### Safer experimentation

AI systems operate on prepared data views rather than operational feeds, allowing teams to test new models or ranking approaches without introducing risk into core commerce platforms.

### Structural explainability

Prepared layers preserve transformation logic and data lineage, making AI outputs traceable to the signals and structures behind them. They also provide a controlled way to govern and improve downstream data quality without affecting source systems.

### Stronger cross-team alignment

Shared prepared data layers act as common contracts between commerce teams, data engineers, and AI specialists. Product attributes, entity definitions, and catalog structures become consistent across systems and use cases.

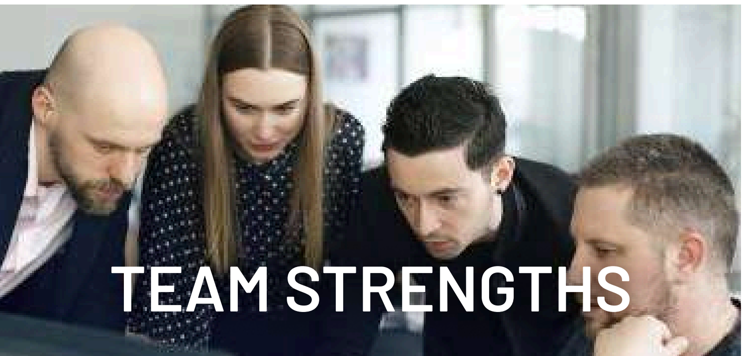
Enterprise AI readiness is closely tied to data architecture. By organizing data through layered preparation, such as the Medallion model, organizations enable AI to operate reliably in everyday ecommerce operations.

# About Expert Soft

Expert Soft is a targeted ecommerce software delivery company, partnering with Fortune 500 companies and global corporations across the US and EU. With SAP Commerce Cloud and Java as our backbone, we know how to ensure scalable and high-performing solutions that can handle 1 mln requests per second, delivering a smooth customer experience.

Developing a payment engine that saved our client about $100 million in operational expenses, ensuring multi-country platform support, adapting solutions for new market entry with tailored enhancements — these are just a few of the challenges our specialists tackle.

We aim to deliver more than a software system. We aim to deliver tailored solutions that maximize profitability within available resources. Our success is driven by:




- All our engineers have a university background
- Specialists excel their skills in our training LABs
- Perfect English skills
- Ready to help 24/7

### CLIENTS

We work with corporations around the world with revenue of over $20 billion and 150K+ employees.

### APPROVALS BY AUDITS

Our ongoing work with corporations is consistently validated through rigorous audits, both by internal teams and Big 4 consulting firms.

### HIGH-LEVEL SECURITY

Approved by assessments from global companies, who are leaders in their respective industries.

### BUDGET EFFICIENCY

By carefully aligning technology investments with your business goals, we ensure optimal value and cost-effectiveness.

### PROFESSIONAL TEAM

No offshore outsourcing and our team's average tenure of 4+ years means you get seasoned problem-solvers, not just coders.

# EXPERT SOFT EXCELS IN

- PAYMENT ENGINE
- MICROSERVICES ARCHITECTURE
- CONTENT MANAGEMENT
- REDESIGN
- E-COMMERCE PLATFORM
- HEADLESS COMMERCE
- MICRO FRONTENDS
- MIGRATION&INTEGRATION

# OUR TECH CORE

### FRONT-END
HTML, CSS, JavaScript (Angular, React, Vue, Next, TypeScript, Jquery), Spartacus

### BACK-END
Java EE, Spring, SAP Commerce (Cloud), Node.JS.

### DEVOPS
Docker, Kubernetes, CI/CD

### UX/UI DESIGN
UX Research, UI Design, Figma, Adobe, Sketch

### QUALITY ASSURANCE
Manual Testing, Test Automation

# TARGETED DOMAINS






# SHARED PATHS, LASTING ECOM VICTORIES

# LET'S TALK SOLUTIONS!

**EKATERINA LAPCHANKA**
Chief Operating Officer
kate.lapsenco@expert-soft.com

+1 585 4997879
+371 25 893 015

**PAVEL TSARYKAU**
CEO & Founder
of Expert Soft

Let's connect

expert-soft.com

LinkedIn